The "What" and "How" of Case Study Rigor: Three Strategies Based on Published Work

Michael Gibbert[1] and Winfried Ruigrok[2]

Abstract

To provide evidence-based strategies for ensuring rigor of case studies, the authors examine what rigor types authors report and how they report them by content analyzing all case studies published 1995–2000 in 10 management journals. Comparing practices in articles addressing rigor extensively and less extensively, the authors reveal three strategies for insuring rigor. First, very few case study authors explicitly label the rigor criteria in terms of the concepts commonly used in the positivist tradition (construct, internal, and external validity, as well as reliability). Despite this, papers addressing rigor extensively do report concrete research actions taken to ensure methodological rigor. Second, papers addressing rigor extensively prioritized rigor types: more, and more detailed, strategies were reported for ensuring internal and construct validity than for external validity. Third, emergent strategies used in the field were reported, such as setbacks and serendipities, that necessitated changes to the originally planned research procedures. Authors focus squarely on the concrete research actions taken, carefully relaying them to the reader so that the reader may appreciate the logic and purpose of trade-off decisions in the context of the specific case study.

Introduction

Weick recently observed that richness—one of the key strengths of case studies— "may lie in the eye of the beholder" (Weick, 2007). Similarly, rigor—often seen as one of the key weaknesses of case studies—often seems to lie in the eye of the beholder and may even involve "persuading" readers and reviewers (Siggelkow, 2007) of the "credibility" of methodological procedures (e.g., Silverman, 2005, 2006). To illustrate, the rigor of quantitative research is subject to standardized procedures and can be assessed in the published manuscript, thanks to a high degree of codification in the reporting conventions (e.g., Gephart, 2004; Scandura & Williams, 2000). By contrast, the procedures for assessing the rigor of qualitative work are much less standardized (Amis & Silk, 2008;

[1] Management Department, Bocconi University, Milan, Italy
[2] Research Institute for International Management, University of St. Gallen, St. Gallen, Switzerland

Corresponding Author:
Michael Gibbert, Management Department, Bocconi University, Via Roentgen 1, 20136 Milan, Italy.
Email: michael.gibbert@uni-bocconi.it

Pratt, 2008). Authors, reviewers, and readers therefore do not have ready access to codified ways of reporting and assessing how rigor was ensured (Eisenhardt & Graebner, 2007).

One of the main reasons for this lack of understanding of *what* to report (e.g., which criteria to use to ensure the rigor of qualitative work), and *how* to report it (e.g., how to prioritize among these criteria and how to best apply them) are decades of unresolved debates between two broad camps, positivists and interpretivists (e.g., Daft & Lewin, 1990; March, Sproull, & Tamuz, 1991; Miles, 1979; Pratt, 2008). Moreover, the different camps seem to be in disagreement on just how to categorize different key texts in the methodology literature at any given point in time, let alone over time. For example, Yin's (1994) text, although commonly viewed as subscribing to a positivist position, explicitly includes data analysis strategies that come from (supposedly) more interpretive persuasions, such as grounded theory (Glaser & Strauss, 1967). Both key concepts underlying grounded theory—that of ''constant comparison,'' in which data collection and analysis are an iterative process, and that of ''theoretical sampling,'' in which data collection decisions are progressional and subject to the theory in construction—are widely considered invaluable to the determination of quality in qualitative research (Fendt & Sachs, 2008, p. 431). However, as Amis and Silk (2008, p. 476) recently pointed out, while originally conceived as a largely positivist approach, overtime grounded theory has been modified to fit more interpretive and constructivist approaches (e.g., Corbin & Strauss, 1988).

This lack of understanding as to what makes "high quality" or methodologically rigorous research (e.g., Easterby-Smith, Golden-Biddle, & Locke, 2008; Gibbert, Ruigrok, & Wicki, 2008) is unfortunate, because papers building theory from cases are frequently considered the "most interesting" (Bartunek, Rynes, & Ireland, 2006; Eisenhardt & Graebner, 2007) and are also among the most impactful papers in the academic community (Eisenhardt, 1989). Furthermore, case studies have provided the management field with some of its most ground-breaking insights (e.g., Burgelman, 1983; Chandler, 1962; Penrose, 1960; Peters & Waterman, 1982; Pettigrew, 1973; Prahalad & Hamel, 1990). However, as Scandura and Williams remind us, "without rigor, relevance in management research cannot be claimed" (Scandura & Williams, 2000, p. 1263).

The purpose of this article is to contribute to the ongoing debate about criteria for ensuring the rigor of published case studies by (a) shedding more light on *what* rigor strategies get reported in published case studies, and (b) revealing *how* authors can best report these actions so as to ensure the rigor of their published work. We take an evidence-based approach to assessing rigor, that is, we compare practices reported in published articles that address rigor extensively and less extensively. This approach is different from other recent work in at least two ways. First, our approach is based on published work, rather than on self-reports of authors that published as well as submitted (but did not publish) manuscripts (e.g. Pratt, 2008). Second, as will be discussed below, whereas the two broad epistemological camps are in disagreement as to what specific labels to use when it comes to ensuring rigor (so-called *primary reports*, e.g. of construct validity, internal validity, generalizability, or external validity, and reliability), they do seem to be in at least partial agreement when it comes to the concrete research actions that are necessary to ensure rigor (so-called *secondary reports*, i.e., reporting of concrete research actions). We therefore focus on both primary and secondary reports of validity and reliability strategies. The sample is all case studies published[1] between 1995 and 2000 in the following journals: *Academy of Management Journal, Administrative Science Quarterly, California Management Review, Journal of International Business Studies, Journal of Management, Journal of Management Studies, Long Range Planning, Management Science, Organization Science, Organization Studies, Sloan Management Review*, and *Strategic Management Journal*. In line with previous research on methodological rigor, we focused on high-ranking general management journals of United States and European origin, because publication in these journals is generally seen as essential for achieving tenure at most business schools (e.g., Boyd, Gove, & Hitt, 2005; Bergh, Perry, & Hanke, 2006; Judge, Cable, Colbert, & Rynes, 2007; Piekkari, Welch, & Paavilainen, in press).

We begin by discussing how case study researchers can ensure rigor in terms of primary and secondary reports relating to the broad criteria construct validity, internal validity, external validity, and reliability. Comparing practices in articles that address rigor extensively and less extensively, we suggest three strategies used by case study authors. The first strategy shows *what* authors report to ensure rigor: authors addressing rigor extensively focus on concrete research actions rather than abstract criteria that are suggested by methodologists of various persuasions. The second and third strategies show *how* authors report these concrete actions. Specifically, the second strategy suggests that authors addressing rigor extensively carefully prioritize the more fundamental types of rigor (internal and construct validity) over others (generalizability, in particular, as well as reliability). Third, authors addressing rigor extensively report the emergent strategies they used in the field for ensuring validity. For example, setbacks and serendipities that necessitated changes to the originally planned research procedures are problematized, focusing squarely on the concrete research actions taken, carefully relaying them to the reader so that he or she may appreciate the logic and purpose of trade-off decisions in the context of the specific case study. The final section discusses implications for existing work on rigor in qualitative research and details implications for authors and editors.

Ensuring Rigor in Case Study Research: Validity and Reliability

Case studies are defined as "research situations where the number of variables of interest far outstrips the number of datapoints" (Yin, 1994, p. 13). Data in the case study method are collected by multiple means and may consist of qualitative research techniques such as interviews, document analysis, various modes of observation, including ethnographical and anthropological strategies as well as the use of quantitative data. While case studies may, and often do, use quantitative data, a key difference with other research methods is that case studies do not attempt to control the context (e.g., Benbasat, 1984; Benbasat, Goldstein, & Mead, 1987; Yin, 1994). As such, case studies enable a researcher to study contemporary phenomena in a real-life setting, where boundaries between context and phenomenon tend to be blurred (Stake, 1995; Yin, 1994).

It has been suggested that the most influential model used to ensure the rigor of case study research adheres to what is commonly called the "natural science model" (e.g. Eisenhardt & Graebner, 2007; Piekkari et al., in press). According to this model, natural science with its positivist world view is the ideal that social science should try to emulate. As will be discussed below, the natural science model groups a number of research actions under four criteria: construct validity, internal validity, external validity, and reliability (Behling, 1980; Campbell & Stanley, 1963; Cook & Campbell, 1976, 1979). As also discussed below, the concrete research actions underlying these criteria have been adapted for use in case studies by Yin (1994), as well as other scholars from both interpretivist as well as positivist traditions (e.g., Campbell, 1975; Denzin & Lincoln, 1994; Eisenhardt, 1989; Kidder & Judd, 1986; Kirk & Miller, 1986; Silverman, 2005, 2006; Stake, 1995).

Construct Validity

The construct validity of a procedure refers to the extent to which a study investigates what it claims to investigate, that is, to the extent to which a procedure leads to an accurate observation of reality (Denzin & Lincoln, 1994). One of the main challenges for case study researchers is to develop a well-considered set of actions, rather than using "subjective" judgments (Yin, 1994, p. 41). Construct validity and the notion of "objective" knowledge it presupposes represents the one criterion where interpretivists and positivists find it difficult to develop common ground. For instance, Silverman's influential interpretivist text explicitly rejects construct validity from his list of criteria to ensure "credible" research (Silverman, 2005, 2006). Silverman argues that models underlying qualitative research are "typically not compatible with the assumption that an objective reality can

be obtained from different ways of looking at it [since] many of the models that underlie qualitative research are simply not compatible with the assumption that 'true' fixes on 'reality' can be obtained separately from particular ways of looking at it" (Silverman, 2005, p. 212).

The positivist literature, however, provides concrete research actions that need to be considered to ensure construct validity. Two main strategies have been suggested. First, researchers have sought to triangulate, that is, adopt different angles from which to look at the same phenomenon, by using different data collection strategies and different data sources (Denzin & Lincoln, 1994; Jick, 1979; Pettigrew, 1990; Stake, 1995; Yin, 1994). Thus, an important strategy to ensure the construct validity of a case study is the triangulation[2] of different sources of data, such as interview data, archival sources, and participatory or direct observation. Practically speaking, authors may report that they themselves or their assistants conducted the interviews. This is in contrast to interview data taken from already existing interviews (e.g., interviews with a company's CEO, published in *Fortune Magazine*) or other archival sources (such as company reports). These archival data may, nevertheless, be used to triangulate and corroborate interview data. Additionally, participant and direct observation may be used for triangulation. These data come from taking part in, or observing, for example, meetings and workshops in the researched organization. Another recommendation is to have transcripts and drafts of the evolving case study reviewed by peers (i.e., academics other than the authors of the case study). This is considered a different operation from sharing transcripts and drafts with key informants (i.e., members of the researched organization) for consistency and accuracy.

Second, researchers have been encouraged to establish a clear chain of evidence to allow the reader to reconstruct how the researcher went from the initial research questions to final conclusions (Yin, 1994, p. 102). In practice, whether a clear chain of evidence is provided depends on careful explication of the data collection procedures, including a reflection on the planned versus actual process, as well as on a discussion of data analysis procedures. A case study's indication of data collection circumstances, such as organizational access, time frame, and interviewee selection approach should be reported. Authors are also encouraged to be explicit about how the planned data collection differed from the actual process (e.g., in terms of difficulties, how this impacted results, and how such difficulties were contained). Discussion of data analysis procedures includes references to qualitative and quantitative data analysis procedures, such as thick descriptions (Geertz, 2003), software packages for analyzing qualitative data, or statistical analysis procedures.

Internal Validity

Internal validity is also called "logical validity" (e.g. by Cook & Campbell, 1979; Yin, 1994) and refers to the presence of causal relationships between variables and results. Whereas construct validity is relevant mainly during the data collection phase, internal validity applies also to the data analysis phase, even though many decisions regarding internal validity are made in the design phase (Yin, 1994, p. 105). Here, the issue is whether the researcher manages to construct a plausible causal argument that is cogent enough to defend the research conclusions.

The concrete research actions that can or should be taken to ensure internal validity, in particular, provide common ground for authors of positivist and interpretivist persuasions. For example, Silverman's influential interpretivist text suggests that "the two central concepts in any discussion of the credibility of scientific research are 'validity' and 'reliability'" (Silverman, 2006, p. 225). While Silverman does not refer to internal validity explicitly, he argues that the single main challenge for qualitative researchers wishing to ensure validity is to convince themselves (and their audience) that their findings are genuinely based on critical investigation of all their data and do not depend on a few well-chosen examples, a problem he describes as "anecdotalism" (Silverman, 2005, p. 211). In line with the positivist tradition, all of Silverman's strategies for avoiding anecdotalism are based on Popper's (1959) critical realism, which demands that researchers seek to

systematically falsify initial hunches about the relations between phenomena in their data. Thus, only if researchers cannot falsify (or refute) the existence and direction of a certain relationship, should that relationship be considered valid. Researchers using quantitative methods typically attempt to satisfy Popper's demand for attempts at falsification or refutation by carefully excluding spurious correlations (e.g., by introducing multivariate analyses). Silverman suggests that researchers using qualitative methods can address the refutation principle with the help of the constant comparative method, comprehensive data treatment, and deviant-case analysis. First, the constant comparative method (Glaser & Strauss, 1967) suggests that the qualitative researcher should always attempt to find another case through which to test out a provisional hypothesis. For example, it could only be claimed with confidence that beginning medical students tended to be idealists if several cohorts of first-year students all shared this perspective (Silverman, 2005, p. 213). The second principle, comprehensive data treatment, demands that "in qualitative research all cases of data are incorporated in the analysis" (Silverman, 2005, p. 215). Silverman points out that such comprehensiveness goes beyond what is normally demanded in quantitative methods. For instance, in survey research, it is usually sufficient to arrive at significant, nonspurious correlations. By contrast, in qualitative research, working with smaller data sets open to repeated inspection, researchers should "not be satisfied until [their] generalization is able to apply to every single gobbet of data you have collected" (Silverman, 2005, p. 215). Third, deviant-case analysis suggests that researchers incorporate into the analysis also cases that do not fit with the theoretical framework or developing hypotheses. As such, deviant-case analysis can be a direct result of comprehensive data treatment.[3]

In summary, three strategies in particular have been proposed to ensure internal validity. First, case study researchers should formulate a clear research framework, which demonstrates that variable x leads to outcome y, and that y was not caused spuriously by a third variable z. Practically speaking, according to Yin, one way to ensure internal validity is to assess whether the research framework for a given case study was explicitly derived from the literature: the issue here is whether authors provide diagrams or verbal descriptions of the relationships between variables and outcomes. Second, through pattern matching, researchers should compare empirically observed patterns with either predicted ones or patterns established in previous studies and in different contexts (Denzin & Lincoln, 1994; Eisenhardt, 1989). Here, authors are encouraged to compare and discuss relationships between their own data and previous research. As a third method, theory triangulation enables a researcher to verify findings by adopting multiple perspectives (Yin, 1994). In this case, authors are encouraged to report different theoretical lenses and bodies of literature used, either as research frameworks to guide data gathering and analysis or as means to interpret findings.

External Validity

"External validity," or "generalizability" is grounded in the intuitive belief that theories must be shown to account for phenomena not only in the setting in which they are studied but also in other settings. Neither single nor multiple case studies allow for statistical generalization, that is, inferring conclusions about a population (Lee, 2003, p. 222; Numagami, 1998, p. 3; Yin, 1994, p. 31). This does not mean, however, that case study authors should give up on generalizability. The key is the differentiation between statistical generalization and analytical generalization. Whereas statistical generalization refers to the generalization from observation to a population, analytical generalization denotes a process that refers to the generalization from empirical observations to theory, rather than a population (e.g., Yin, 1994). In her widely cited paper, Eisenhardt (1989) argued that case studies can be a starting point for theory development and suggests a cross-case analysis involving 4 to 10 case studies that may provide a sound basis for analytical generalization. Instead of conducting and analyzing multiple case studies of different organizations, researchers may also conduct different case studies within one organization (a nested approach, e.g., Yin, 1994). The rationale for the

selection of a case study (e.g., to what extent it was a representative case for an organization in a given industry or for an industry) should also be reported. Last, but not least, researchers should provide a clear rationale for the case study selection, and ample details on the case study context (e.g., competitive dynamics, financial data, business cycle) to allow the reader to appreciate the researchers' sampling choices (Cook & Campbell, 1979, p. 83).

Reliability

"Reliability" refers to the absence of random error, enabling subsequent researchers to arrive at the same insights if they conducted the study along the same steps again (Denzin & Lincoln, 1994). Similarly, Silverman defines reliability as "the degree of consistency with which instances are assigned to the same category by different observers or different occasions" (Silverman, 2005, p. 210). Silverman points out that in many qualitative studies, "the reader has to depend on the researcher's depiction of what was going on" (Silverman, 2005, p. 221). Silverman suggests that reliability can be ensured by a principle he calls "low-inference descriptors." The author explains that although no piece of research can be free from the underlying assumptions that guide it, detailed data presentations that make minimal inferences are always preferable to researchers' presentations of their own (high-inference) summaries of their data. With regard to interview data, in particular, Silverman suggests a number of means for increasing reliability, including tape recording all face-to-face interviews, carefully transcribing these tapes, as much as possible using fixed-choice answers and/or inter-rater reliability checks on the coding of answers to open-ended questions, as well as presenting long extracts of data in the research report.

Thus, key words here are transparency and replication. Transparency can be enhanced through strategies such as careful documentation and clarification of the research procedures, for example, by producing a case study protocol—a report that specifies how the entire case study has been conducted. Authors are also encouraged to make reference to a case study database, in which data such as interview transcripts, preliminary conclusions, and the narratives collected during the study are organized in such a way as to facilitate retrieval for later investigators (Yin, 1994), that is, to facilitate the replication of the case study (e.g., Leonard-Barton, 1990).

In summary, the methodology literature has proposed a number of research actions or strategies to codify validity and reliability concerns. Below, we investigate the level of rigor displayed by case studies published in leading management journals in terms of these criteria. We then draw on the most rigorous to tease out best practices in ensuring rigor, pointing also to concrete examples of articles and exemplar quotes within these articles.

Method

Sampling

Previous studies on validity and reliability in management research selected journals based on their exposure and credibility (e.g., Mitchell, 1985; Scandura & Williams, 2000; Schriesheim, Powers, Scandura, Gardiner, & Lankau, 1993). Earlier work was primarily longitudinal, centering around three journals either comparing two 3-year periods (Scandura & Williams, 2000) or examining a single one over a 5-year period (e.g., Mitchell, 1985). As noted above, we focused on high-ranking general management journals of United States and European origin as publishing in these journals is essential for achieving tenure in most business schools. Specifically, we selected case studies that were published over a 6-year period (1995–2000) in 12 journals: *Academy of Management Journal, Administrative Science Quarterly, California Management Review, Journal of International Business Studies, Journal of Management, Journal of Management Studies, Long*

Range Planning, Management Science, Organization Science, Organization Studies, Sloan Management Review, and *Strategic Management Journal*.

To select these journals, we used the Tahai and Meyer ranking that is based on the citations articles receive (Tahai & Meyer, 1999). Tahai and Meyer ranked the 12 journals among the most influential in strategy and management. The Tahai and Meyer ranking was the most comprehensive at the time of the study and also the most appropriate for an analysis of methodological rigor and its relationship with impact, because it is based on the notion that the "most-often cited journals contain the ideas that are most closely scrutinized, evaluated, and extended" (Tahai & Meyer, 1999, p. 280). Furthermore, since we were interested in case studies in management journals, we set aside related journals that had an explicit focus on psychology, finance, economics, entrepreneurship, and so on (even though many of these can be found on other influential lists, such as the Financial Times list).[4] Although there may be disagreement about which journals to include on a list such as this one, most overlap with other lists of the most influential management journals (e.g. Bergh et al., 2006; Boyd et al., 2005).

To identify case studies within these journals, we decided to be as inclusive as possible from the outset. A research assistant carried out a key word search on the "EBSCOhost Electronic Journals Service" database for the 12 journals over the 6-year period, using the search term "case," in title, abstract, and full text of each article. Of the articles thus identified, we included only those that were consistent with the definition of case studies provided earlier (cf., Benbasat, 1984; Benbasat et al., 1987; Yin, 1994), that is, that reported results based on primary fieldwork in one or more organizations, in which no experimental controls or manipulation were involved, and which used multiple sources of data. In addition to this, due to our focus on business management (rather than administration at large), we concentrated on for-profit organizations, and therefore excluded case studies in the public and nonprofit sector, such as those on kibbutzim, schools, and universities. We found no case studies that fitted these criteria in the *Journal of Management* and *Management Science*. In the remaining 10 journals, 159 case studies fitted these criteria.

Quantitative Data Analysis: Coding of Rigor Actions

The first step in the data analysis was a quantitative assessment of the rigor dimensions. Specifically, the case studies' methodological rigor was analyzed by the 24 coding dimensions in Table 1. As discussed above, the 24 coding dimensions were adapted to the case study method by Yin (1994) as well as others (e.g., Campbell, 1975; Eisenhardt, 1989; Kidder & Judd, 1986; Kirk & Miller, 1986; Stake, 1995) and involve the four criteria that are commonly used to codify the rigor of field research: construct validity, internal validity, external validity, and reliability (Campbell & Stanley, 1963; Cook & Campbell, 1976, 1979). In line with Yin, we distinguished between the explicit mentioning of the four types of validity and reliability (primary reports) and the concrete research actions or strategies reported to ensure each validity and reliability type (secondary reports). Both were coded dichotomously, that is, either a paper "scored" on a given report, and a "1" was entered into the spreadsheet or a paper did not score, in which case a "0" was entered.

Primary reports on internal validity, construct validity, external validity, and reliability were recorded according to the terminology provided by Cook and Campbell (1979). Internal validity and its synonym, "logical validity" were code 1. Construct validity and its synonym, "concept validity" were code 2. External validity and its synonym "generalizability," as well as the verb "to generalize" were code 3. Likewise, reliability was coded when the term or its synonym "replicability," or the verb "to replicate" were mentioned explicitly; this was code 4. When a given article explicitly mentioned a given report more than once, only one code was assigned.

Secondary reports were descriptions of research actions or strategies that were performed to ensure a case study's validity and reliability (Table 1). Secondary reports were recorded according to the definitions and categorizations provided by Yin (1994) and involved concrete actions in the

Table 1. Primary and Secondary Reports on Validity and Reliability in Case Study Research

Primary reports	
Action/research procedure	Coding rules in this study
1. Construct validity	To be mentioned explicitly
2. Internal validity	To be mentioned explicitly (or synonym "logical validity")
3. External validity	To be mentioned explicitly (or synonym such as "generalizability," "generalize")
4. Reliability	To be mentioned explicitly (or synonym such as "replicability," "replicate")
Secondary reports	
Construct validity: research actions	Coding rules in this study
Data triangulation	
5. Archival data	Internal reports, minutes, or archives; annual reports; press; or other secondary articles
6. Interview data	Original interviews carried out by researchers
7. Participatory observation derived data	Participatory observation by researchers
8. Direct observation derived data	Direct observation by researchers
9. Review of transcripts and draft by peers	Peers are academics not coauthoring the paper
10. Review of transcripts and draft by key informants	Key informants are or have been working at organization investigated
Clear chain of evidence	
11. Indication of data collection circumstances	Explanation how access to data has been achieved
12. Check for circumstances of data collection vs. actual procedure	Reflection of how actual course of research affected data collection process
13. Explanation of data analysis	Clarification of data analysis procedure
Internal validity: research actions	Coding rules in this study
14. Research framework explicitly derived from literature	Diagram or explicit description of causal relationships between variables and outcomes
15. Pattern matching	Matching patterns identified to those reported by other authors
16. Theory triangulation	Different theoretical lenses and bodies of literature used, either as research framework, or as means to interpret findings
External validity: research actions	Coding rules in this study
Cross case analysis	
17. Multiple case studies	Case studies of different organizations
18. Nested approach	Different case studies within large organization
19. Rationale for case study selection	Explanation why this case study was appropriate in view of research question
20. Details on case study context	Explanation of, for example, industry context, business cycle, product/market combinations combinations, financial data
Reliability: research actions	Coding rules in this study
21. Case study protocol	Report of how entire case study was conducted
22. Case study database	Database with all available documents, transcribed interviews, case study notes, etc.
23. Organization mentioned by own name	Name to be mentioned explicitly (as opposed to anonymous)
Alternative Reports	
24. Alternative actions and reports (whether primary or secondary reports)	

categories internal validity, construct validity, external validity, and reliability. According to Yin, one way to ensure the first of these, internal validity, is to assess whether the research framework for a given case study was explicitly derived from the literature: here, we looked for diagrams or

verbal descriptions of the relationships between variables and outcomes; this was code 5. Pattern matching represents another approach to ensure internal validity—here, we recorded when authors compared and discussed relationships between their own data and previous research; this was code 6. A final research strategy in the internal validity category is theory triangulation; we recorded instances where authors reported different theoretical lenses and bodies of literature used, either as research frameworks to guide data gathering and analysis or as means to interpret findings (code 7).

An important strategy to ensure the construct validity of a case study is the triangulation of different sources of data, such as interview data, archival sources, and participatory or direct observation. We recorded interview data (code 8), when it was reported that the authors or their assistants conducted the interviews themselves. When already existing interviews (e.g., interviews with a company's CEO, published in a magazine or newspaper) or other archival sources (such as company reports) were used, we assigned a different code (Code 9). When researchers participated in meetings and workshops in the researched organization, we recorded this as Code 10; direct observation was recorded as Code 11. In line with Yin's recommendations, authors' mentioning that transcripts and drafts of the evolving case study were reviewed by peers (i.e., academics who were not authors of the case study), which were also recorded (Code 12). A different code (Code 13) was assigned when authors shared transcripts and drafts with key informants (i.e., members of the researched organization) for consistency and accuracy.

Whether a clear chain of evidence is provided depends on an explication of the data collection procedures, including a reflection on the planned versus actual process, as well as on a discussion of data analysis procedures. A case study's indication of data collection circumstances, such as organizational access, time frame, and interviewee selection approach was recorded (Code 14). When authors were explicit about how the planned data collection differed from the actual process (e.g., in terms of difficulties, how this impacted on results, and how such difficulties were contained); this was recorded as Code 15. Discussion of data analysis procedures (Code 16) included references to qualitative and quantitative data analysis procedures, such as thick descriptions, software packages for analyzing qualitative data, or statistical analysis procedures.

A key factor to extend a case study's external validity is whether a cross-case analysis was conducted. A case study that involved several organizations and that reported an analysis across the insights gathered in each of these organizations was coded (Code 17). A case study that involved several subcases "embedded" in the overall context of one organization was assigned a different code (Code 18). The rationale for the selection of a case study (e.g., to what extent it was a representative case for an organization in a given industry or for an industry) was also recorded (Code 19). Another code (Code 20) was assigned when details were recorded regarding the specific context of the researched case, such as competitive dynamics, financial data, business cycle, and so on.

Finally, research strategies or actions taken to ensure a case study's reliability were recorded. In this category, we looked for descriptions of a case study protocol, which would give a detailed report on how the case study was conducted (Code 21). When reference was made to a case study database, in which data such as interview transcripts, preliminary conclusions, and drafts were kept, we assigned Code 22. Another reliability strategy was whether the organization was mentioned by its own name (as opposed to anonymized), making it possible for other researchers to contact that organization; this was Code 23. Furthermore, we also looked for alternative research frameworks that explicitly stated relevant actions and strategies but which were *not* consistent with the four positivist quality criteria construct validity, internal validity, generalizability, and reliability as proposed by Yin (e.g., interpretivist or critical approaches), this was Code 24 (see Table 1).

Each article was examined by three coders, and a coding sheet was compiled for all 24 codes. Thus, we produced 3×159 (the number of raters times the number of papers) coding sheets. We read and coded the full article and not just the methods section (as performed by Scandura & Williams, 2000) because it appeared that some authors presented methodological considerations in the discussion or limitations sections.

Table 2. Frequencies per Journal of *Primary Reports* (Explicit Mentioning of Terms)

Journal	Case study papers 1995–2000	Construct validity	Internal validity	External validity	Reliability
Academy of Management Journal (AMJ)	7	1	0	1	2
Administrative Science Quarterly (ASQ)	2	1	0	1	1
California Management Review (CMR)	11	0	0	1	1
Journal of International Business Studies (JIBS)	3	0	1	2	1
Journal of Management Studies (JMS)	35	0	0	11	3
Long Range Planning (LRP)	35	0	1	2	0
Organization Science (Osc)	12	0	1	4	4
Organization Studies (Ost)	18	0	0	8	2
Sloan Management Review (SMR)	23	0	0	0	0
Strategic Management Journal (SMJ)	13	0	2	1	2
Total	159	2	5	31	16

To assess the inter-rater consistency of the coding process, the three coders physically met to discuss the 24 codes for each paper. This "multiple-rater" approach has been used previously in methodology surveys (Bullock & Tubbs, 1990; Gerstner & Day, 1997; Larsson, 1993; Scandura & Williams, 2000). In discussing codes, we followed a consensus-coding approach, in which multiple coders meet to discuss discrepancies in coding to agree on the most correct one (as recommended by Bullock & Tubbs, 1987, pp. 202-203). Where initial codings differed, the three coders discussed the appropriate final coding and referred back to the actual paper until there was consensus. All preconsensus and consensus codes were subsequently fed into a spreadsheet. Overall preconsensus coding inter-rater reliability measured as average pairwise percent agreement (APPA) of coding across raters (Larsson, 1993) was 0.97. Inter-rater reliability measured by Fleiss' κ, which provides a more robust measure than APPA because it takes into account the agreement occurring by chance, was 0.93 (Fleiss, 1971).

Results of the Quantitative Analysis

Table 2 provides the results for primary reports, and Table 3 for secondary reports.

Table 2 shows that of the 159 papers that were coded, only two provided primary reports on construct validity (both counts were in 2000) and only five on internal validity (two counts in 1996, and one count in 1998, 1999, and 2000 each). Primary reports on external validity and reliability occurred slightly more often. Notably, no primary reports for alternative rigor strategies were found.

Table 3 shows the frequencies of secondary reports (i.e., the concrete research actions taken to ensure validity and reliability, as discussed above and in Table 1). As can be seen from Table 3, there are relatively more secondary reports than primary reports. How to interpret this discrepancy between primary reports and secondary reports? One way of looking at this gap is to say that authors underutilize positivist criteria for ensuring rigor (as illustrated by the virtual absence of primary reports). This interpretation suggests that management researchers employing case studies score even lower than management scholars using other methodologies. For example, Scandura and Williams (2000) in their analysis of validity and reliability across methodologies also found that validity and reliability strategies were underutilized. In particular, they found a significant decrease in the reports on internal, external, and construct validity between the two time periods they studied (1985–87 and 1995–97). At the same time, Scandura and Williams found that overall, validity and reliability reports decreased, they also found that up to 85.6% of articles published in the *Academy of*

Table 3. Frequencies per Journal of *Secondary Reports* (Concrete Reporting of Research Actions)

Report[8]/Journal	AMJ	ASQ	CMR	JIBS	JMS	LRP	OSC	OST	SMR	SMJ
Construct validity										
Data triangulation										
5. Archival data	1	5	6	1	28	16	11	14	13	13
6. Interview data	6	2	8	3	32	17	9	16	11	13
7. Participatory observation derived data	0	1	4	0	6	8	0	3	0	1
8. Direct observation derived data	1	1	1	0	11	9	2	6	1	4
9. Review of transcripts and draft by peers	0	1	2	1	0	1	3	2	3	0
10. Review of transcripts and draft by key informants	0	0	1	2	4	5	3	4	1	0
Clear chain of evidence										
11. Indication of data collection circumstances	5	2	6	3	21	6	7	17	7	4
12. Circumstances of data collection vs. actual procedure	2	1	0	0	3	0	0	1	5	0
13. Explanation of data analysis	3	2	2	1	7	1	4	7	5	0
Internal validity										
14. Research framework explicitly derived from literature	3	0	2	3	12	16	6	7	7	2
15. Pattern matching	3	2	0	0	8	0	2	1	4	0
16. Theory triangulation	1	4	0	0	7	2	1	5	3	0
External validity										
17. Multiple case studies	3	0	4	2	14	9	4	11	5	3
18. Nested approach	2	0	1	0	4	1	3	2	4	3
19. Rationale for case study selection	6	2	4	3	20	6	8	16	10	2
20. Details on case study context	0	2	7	2	27	15	9	16	10	15
Reliability										
21. Case study protocol	0	1	0	0	0	0	1	1	1	0
22. Case study database	2	1	3	1	13	5	5	7	6	3
23. Organization mentioned by own name	2	1	8	2	19	27	8	7	9	17
Papers based on case studies published per journal	7	2	11	3	35	35	12	18	23	13

Management Journal, Administrative Science Quarterly, and *Journal of Management* reported internal validity strategies and up to 40% reported construct validity strategies (Scandura & Williams, 2000, pp. 1259-1263).

Another way of looking at the discrepancy between primary reports and secondary reports is to assume[5] that authors seem to stay away from explicitly subscribing to a particular ideology (as illustrated by the scarcity of explicit reports of positivist criteria such as internal validity, construct validity, generalizability, or reliability, and by the absence of explicit reports of research actions other than positivist ones), while actually performing and reporting the concrete research actions that should be taken to ensure rigor. Qualitative analysis was needed, therefore, to reveal exactly what authors report in the secondary reports and how they report them.

Qualitative Data Analysis: Comparing Practices in Articles That Address Rigor Extensively and Less Extensively

The second step in the data analysis was a qualitative content analysis. The aim was to look beyond the descriptive statistics in the quantitative analysis and to reveal *what* rigor strategies get reported and *how*. Given the big variance in the utilization of rigor strategies across papers (Table 3), a

method was needed that allowed for a comparison of practices in articles that address rigor extensively and less extensively. The most commonly used technique for contrasting cases based on dichotomous variables is to look for cases that are similar in all respects, except the variables of interest (Gerring, 2007). Specifically, for exploratory research such as the current study, one looks for cases that differ on the outcome of theoretical interest (the degree of rigor, in our case) but are similar on various factors that might have contributed to that outcome (the reported rigor strategies, in our case, see Gerring, 2007, p. 131). Thus, if more of the concrete research actions addressing the individual types of validity are reported and explicitly named (primary and secondary reports), the more extensively rigor is addressed in a given case study (Yin, 1994, p. 33). Consequently, to investigate in greater depth the "best practices" on the level of individual papers, we focused our content analysis on those articles that addressed rigor very extensively (namely those that reported most primary and secondary reports) and contrasted them with the papers that addressed rigor less extensively.

To identify the papers that addressed rigor extensively, and to contrast them with the group that addressed rigor less extensively, we made a first cutoff at those case studies that reported at least 50% of the primary and secondary reports in Table 1. Of these 19 papers, we selected those papers that reported at least 2 of 3 of the secondary reports associated with providing a clear chain of evidence (construct validity), 2 of 3 reports associated with internal validity, 3 of 4 for generalizability, and 2 of 3 reports for reliability. These "top 10" papers addressing rigor most extensively were, in descending order, Bacharach, Bamberger, & Mc Kinney, 2000; Birkinshaw, Bresman, & Håkanson, 2000; Bresman, Birkinshaw, & Nobel, 1999; Buchel, 2000; Doz, 1996; Larsson & Finkelstein, 1999; Laurila, 1997; Maznevski & Chudoba, 2000; Miller, 1997; and Woiceshyn, 2000). The average number of reports in the top 10 most rigorous papers was 13. By contrast, the average number of reports in the less rigorous papers was six.

In line with other recent research investigating the rigor of qualitative research quantitatively, as well as qualitatively (e.g., Pratt, 2008), the actual statements in the papers (the secondary reports) were used as data that constitute themes related to best practice actions. As such, qualitative analysis of rigor strategies adhered to common principles for qualitative pattern matching (Locke, 1996; Miles & Huberman, 1984). That is, we performed a qualitative analysis to uncover the fundamental strategies or best practice actions involved in creating rigor. Specifically, we read (and re-read) the textual database (the papers addressing rigor either extensively or less extensively) to discover or label variables, categories, concepts, and properties and their interrelationships. We analyzed in an iterative fashion, traveling back and forth between the data and the emerging strategies (Corbin & Strauss, 1988). First, we began by open coding the data to better understand how the authors used specific forms of rigor both individually and together. This part of the qualitative analysis was concerned with identifying, naming, categorizing, and describing phenomena found in the text. Essentially, each line, sentence, paragraph, and so on, in a paper was read in search of the answer to the repeated question "What is this about? What is being referenced here?" Common statements were used to form provisional categories and first-order codes. As illustrated in Table 4, two first, rough, categories were formed: *what* rigor strategies get reported and *how* they get reported. These categories were then detailed, becoming more specific themes. This involved a move from open to axial coding, that is, the mirroring of categories in either of the two groups ("papers that address rigor extensively" and "papers that address rigor less extensively") with emerging themes. For example, as regards "what" rigor strategies papers report, Tables 2 and 3 suggest that relatively few papers, overall, focus on internal and construct validity. As Table 4 illustrates, qualitative content analysis then revealed that papers addressing rigor extensively actually put more emphasis on internal and construct validity, especially relative to external validity. By contrast, papers that address rigor less extensively focused mainly on external validity (and not much else). As we continued to read papers addressing rigor extensively, contrasting them with papers addressing rigor less extensively, it

Table 4. Themes Arising From Comparison of How Authors Address Rigor

Categories, concepts	Emerging themes: rigor addressed		Typical examples of papers that address rigor	
	Extensively	Less extensively	Extensively	Less extensively
What rigor strategies get reported	Quality criteria in rigorous papers are not "named", yet they are addressed.	Quality criteria (external validity, in particular) are often named, but it is not explained how they were addressed.	We analyzed the data using the constant comparative method [...] careful to ensure that the particular categories and constructs identified were not based on instances related by a non-representative informant or on non-representative events described by representative informant [we looked] purposively for counter-evidence (i.e., that the categories identified were, for example, not as comprehensive [...] as we inferred). (Bacharach, Bamberger, & McKinney (2000, pp. 711-713) [I]n an effort to triangulate or corroborate our hunches about particular constructs and patterns, we paid particular attention to [counter-evidence] When such counter-evidence was found, we revised the coding scheme and, where relevant, recoded interview transcripts. (Bacharach et al., 2000, p. 713)	Our version of the story [the case study] relies on various ROLM [the researched organization] marketing and strategy documents currently in the possession of one of us. (without further mention of data collection and analysis strategies used, Lane & Maxfield, 1996, p. 218) The scores which do not confirm the theoretically anticipated outcomes (as presented in Table 1) are shown in bold. (no discussion of how that counter-evidence was used to refine the initial framework, see van Hoek, Voss, & Commandeur, 1999, p. 513)

| *How rigor strategies get reported* | Rigorous papers say little (if anything) about external validity; they seem to "belabor" internal and construct validity, though | Papers that address rigor less extensively say little (if anything) about internal and construct validity; of all the validity types, external validity receives most attention | We specifically sought to interview managers with a lot of firsthand experience of interaction with the partner, in particular managers who had been 'exchanged' between partners. (the focus of the paper was on the evolution of cooperation between partner firms, see Doz, 1996, p. 58)

The interview tapes were transcribed, and I then coded the transcripts to categorize the data and to facilitate analysis. The codes were developed both before and during the field work, based on the ideas that emerged from the interview notes, as well as the literature on innovation, technology adoption, and organizational learning. (Woiceshyn, 2000, p. 1102)

The constant comparison method starts with labeling raw data into basic units of analysis. [...] Incidents are compared and grouped into the same or different units of analysis. Once this process was completed, units were grouped around themes, thus categorizing the phenomenon. The categorized phenomenon is again given a name to pull together a number of subcategories through the criteria of similarity. (Buechel, 2000, p. 643) | The primary sources included 32 semistructured individual interviews with 25 different people at all levels and areas of the organization, as well as with several external consultants and researchers. (Lovas & Goshal, 2000, p. 879)

This paper presents a case study elaborated from direct observations, interviews with actors, and analysis of written documents. (no further explication of data collection or analysis is given, Derong & Faure, 1995, p. 30)

The linking theme of the research framework was the relationship between organizational systems and corporate performance. (no further explication of how the data were analyzed, to prove or disprove this relationship Mueller & Dyerson, 1999, p. 236). There are obvious limits to what can be inferred from analysis of a single historical case. But, the richness of detail available in this extended case should enable us to seek insights into some of the general propositions about adaptation to technological change. (Rosenbloom, 2000, p. 1100) |

(continued)

Table 4 (continued)

Categories, concepts	Emerging themes: rigor addressed		Typical examples of papers that address rigor	
	Extensively	Less extensively	Extensively	Less extensively
	Rigorous papers are surprisingly "honest" about their work in the field, what went wrong and how it was fixed; the research strategy as planned often differed from the one as implemented	Methods sections in papers that address rigor less extensively read as if the field research was all "plain sailing"; the research strategy as planned does not seem to differ much from the one as implemented	Although data collection and analysis are presented in two sections here, because they represent different conceptual stages of the research process, chronologically the two activities overlapped [...]. For example, early interviews and observations guided our development and presentation of the communication logs and questionnaires, and our analyses of these earlier meetings guided what we looked for in later meetings and interviews. (Maznevski & Chudoba (2000, p. 479) We also observed from other, real-time, process studies [...] that it was extremely difficult to be sure not to influence ongoing processes and still maintain a legitimate presence in the field insofar as managers would quickly be tempted to seek advice from the researcher and ask the researcher to intervene in the process, as a quid pro quo for allowing further observations. (Doz, 1996, p. 58)	Interviews were audio recorded, transcribed verbatim, and content analyzed. Data analysis and interpretation focused on the specific interview questions as well as on unanticipated issues that emerged from the data itself. (methods section stops here, no further mention if these unanticipated issues influenced data analysis, and if they prompted further data collection, see Andrews & Delahaye, 2000, p. 799) Forty-six CCT employees were interviewed for this study, with interviews lasting from forty-five to ninety minutes. Repeat interviews were conducted with ten people. Some of the interviews were exploratory in nature, but most were semi-structured. The interviewees were at various managerial levels and from a broad span of functions. (Levy, 2000, pp. 349-350)

became evident that the category "how" rigor strategies get reported needed to be split into two different subcategories. Specifically, it appeared, first, that papers seemed to prioritize the validity and reliability types: papers addressing rigor extensively reported mostly strategies related to internal and construct validity, whereas papers addressing rigor less extensively focused mainly on external validity (at the expense of the other validity types). Second, papers addressing rigor less extensively often read like an idealized account of what happened in the field: the research strategy could be implemented largely as planned. By contrast, papers addressing rigor extensively actually thematized unexpected findings, problems in the field, and how these problems were resolved. Contrasts such as these between the two categories of papers were used when revisiting the data to see whether and how well or badly they met the emerging strategies (Glaser & Strauss, 1967; Locke, 2001). Thus, as we moved from open to axial coding, we divided the original category of how rigor strategies get reported into two specific strategies, as detailed in Table 4. Overall, this travelling back and forth between the data (the papers) and the themes therefore yielded three fundamental strategies for ensuring rigor.

Three Strategies for Ensuring Case Study Rigor Based on Published Work

Below, we report the result of the qualitative content analysis in the form of three strategies for ensuring the rigor in published case studies. Following earlier work on rigor in qualitative research (e.g., Pratt, 2008), we also illustrate each of them with various quotes taken from articles that display a certain strategy most clearly and completely. Given the constraints of a journal-length article, the selection and length of the quotes are necessarily limited. To select the articles to quote from, we decided to follow van Maanen's approach in that we "consider each article [. . .] to be something of an exemplar, and thus worthy of our field's highest, reward, imitation" (van Maanen, 1998, p. xxiv).[6]

Talk the Walk: Report Concrete Research Actions Rather Than Abstract Criteria

As pointed out in the introduction, epistemological debates around *what* to report when it comes to ensuring rigor (i.e., which criteria to use to ensure the rigor of qualitative work) have been around for at least three decades (e.g., Daft & Lewin, 1990; March et al., 1991; Miles, 1979; Pratt, 2008). To capture these dynamics, Pratt (2008) offered the metaphor of fitting oval pegs (qualitative criteria for rigor) into round holes (positivist, quantitative criteria for rigor). The author also suggested that case study researchers should ensure rigor by "focusing on the process of fitting," rather than on "making oval pegs seem rounder, or by making round holes bigger or more oval friendly" (Pratt, 2008, pp. 497-504). Case studies in our sample that addressed rigor extensively fall into line with that recommendation. They focus on the concrete research actions that were taken and carefully walk readers through their methodological choices and decisions (rather than calling pegs oval or holes round, to borrow Pratt's terminology).

The analogy we offer to capture these dynamics, "talk the walk," aims at pointing to the fact that authors of best practice case studies in our sample "talk" readers through the various milestones on this "walk." Thus, our analysis suggests that a first tactic for ensuring rigor in case studies is to focus less on the coherence between affirmations about rigor criteria from various persuasions and subsequent actions to implement them (walk the talk) and instead to focus more on transparency, that is, relaying to the reader the concrete research actions taken, so that he or she may appreciate the logic and purpose of these actions in the context of the specific case study at hand (talk the walk). In line with Pratt's observations, papers in our sample that addressed rigor extensively took a very conversational but detailed approach in walking the reader through the major parts of their methods,

including case selection, data collection, and data analysis (Pratt, 2008, p. 503). Specifically, authors provided concrete descriptions of how categories and codes were formed iteratively (rather than providing only the superficial "nod" to Glaser and Strauss' 1967 work, see Pratt, 2008, p. 503, without actually talking the reader through the application of this framework in the specific context). For example, Bacharach et al. (2000) take great care to relay to the reader the concrete actions that were taken when applying the constant comparative method and do so in a very "hands-on" way:

> We analyzed the data using the constant comparative method [. . .] we began to identify theoretical categories and make comparisons across categories well before the formal process of data analysis began. [. . .] In generating our coding scheme, we were careful to ensure that the particular categories and constructs identified were not based on instances related by a non-representative informant or on non-representative events described by representative informants (Miles & Huberman, 1984, p. 231). [. . .] We were also careful to avoid the "holistic bias," the tendency of researchers to see more meaningful patterns than exist in reality by looking purposively for counter-evidence (i.e., that the categories identified were, for example, not as comprehensive or as orthogonal as we inferred) (Miles & Huberman, 1984, p. 231). (Bacharach et al. (2000, pp. 711-713)

As part of the "talk the walk" strategy, papers addressing rigor extensively carefully explained when data collection was stopped and when analysis began; authors reported that their material was organized in a case study database (enabling potential future investigators to reaccess the database), thereby enhancing reliability. Bacharach et al. (2000) provide an example of such a conversational, yet to-the-point and detailed approach to ensuring construct validity and reliability:

> In the majority of cases, [interviewees] spontaneously pulled out and referred to written case files as the basis for their descriptions. When this did not happen spontaneously, in an effort to minimize any retrospective bias, we asked them to do so, reminding them to protect recipients' anonymity and confidentiality. [. . .] We further attempted to reduce the risk of informants imposing some sort of reconstructive rationality on their actions by adopting several of the suggestions proposed by Miles and Huberman (1984, p. 233), such as (1) spending several days with each informant, to enhance the level of trust and openness (2) being careful not to lead the informant in his or her responses; and (3) attempting to share as little of our knowledge or as few of our hunches with informants as possible, so as to avoid giving any implicit request for confirmation [authors then give extensive information how this happened for all three points]. (Bacharach et al., 2000, pp. 711-712)

Similarly, Doz (1996) talks the reader through the selection process for interviewees in a nontechnical, yet cogent way that enables the reader to appreciate the rationale for selecting interviewees, given the specific context and aims of the case study:

> The selection process for the interviewees was simple: when the initial agreement to collaborate in the research was granted—separately by each partner firm, the initial discussions leading to access also identified the key participants in the collaboration process. We specifically sought to interview managers with a lot of firsthand experience of interaction with the partner [the object of study was the development of cooperation in strategic alliances], in particular managers who had been 'exchanged' between partners (at least in terms of where they were located) and managers concerned with the partnership but with little direct experience. (Doz, 1996, p. 58)

In summary, to reveal the substantive aspect of *what* gets reported, we offer the analogy "talk the walk": best practice case studies do not simply use well-worn labels and rhetoric suggested by methodologists as a means for affirming the coherence between their claims and actions. Instead, best practice case studies focus squarely on the concrete research actions taken, carefully relaying them

to the reader so that he or she may appreciate the logic and purpose of these actions in the context of the specific case study. As Table 4 suggests, this practice of papers addressing rigor extensively is in contrast to papers addressing rigor less extensively, which seem to be far less detailed, more "codified" and technical when it comes to reporting validity and reliability strategies. Basically, what differentiates the two categories of papers is the level of detail with which the *application* of a specific rigor criterion to the studied context is relayed.

Priority Ordering of Validity Types: Internal and Construct Validity Over External Validity

The first strategy, "talk the walk" has so far demonstrated *what* authors reported, by focusing on the secondary reports for ensuring validity and reliability types individually. The second strategy deals with the interrelationships of the validity types, or *how* authors report the concrete research actions. Importantly, according to Cook and Campbell (1979), individual validity types are not independent but build on each other. No single piece of research can satisfy all criteria (Cook & Campbell, 1979):

> Some ways of increasing one [type of] validity will probably decrease another kind. Means have to be developed for avoiding all unnecessary trade-offs between one kind of validity and another and to minimize the loss entailed by the necessary trade offs [...] For investigators with theoretical interests, our estimate is that *the types of validity, in order of importance, are probably internal, construct, statistical conclusion, and external validity*. [...] [J]eopardizing internal validity for the sake of increasing external validity usually entails a minimal gain for a considerable loss. (pp. 83-84; own emphasis)

As discussed earlier, several authors that are commonly thought of as interpretivist actually are in line with Cook and Campbell's (positivist) argument that the avoidance of spurious relationships (i.e., internal validity), in particular, should take precedence when ensuring rigor, and that rigor does *not* involve simply "ticking off" certain criteria individually. Thus, authors that are commonly thought of as positivist (such as Yin) as well as authors that are considered interpretivist (such as Silverman) are largely in agreement as to the concrete research actions that are necessary to ensure rigor as well as the prioritization among these research actions (internal and construct over external validity).

Empirically, the relative emphases case study authors put on individual validity and reliability types are therefore particularly noteworthy. At first glance (Tables 2 and 3), case-study authors in our sample seem to do the exact opposite of what should be done: as can be seen from Tables 2 and 3, over the period investigated, case-study authors have generally been more concerned with addressing external validity than construct validity, or reliability, let alone internal validity. Tables 2 and 3 show that part of this effect comes from case studies published in three European journals (*Journal of Management Studies, Long Range Planning*, and *Organization Studies*), although external validity scores are higher for case studies published in the other journals as well.

Case study authors' preoccupation with external validity may be understandable, given pressures that require them to acknowledge the limitations involved in "learning from samples of one or fewer" (March et al., 1991) and the heavy emphasis put on external validity in Eisenhardt's (1989) highly influential paper. However, our results suggest that too many researchers have sacrificed other types of validity for the sake of external validity. As Cook and Campbell (1979) as well as Silverman (2006) suggest, that may be too high a price to pay. As such, there are sound theoretical arguments, suggesting that case study authors wishing to address rigor extensively should prioritize internal and construct validity over external validity. So, how can the apparent difference between theoretical admonitions to prioritize internal and construct validity over external validity be reconciled with the criteria in use that seem to do the exact opposite?

A closer look at Table 3 reveals that certain journals (particularly the *Academy of Management Journal, Administrative Science Quarterly*, and the *Strategic Management Journal*) show more than twice as many internal and construct validity counts than do all other seven journals taken together. External validity counts provide a different picture. Here, authors in *California Management Review, Journal of International Business Studies, Journal of Management Studies, Long Range Planning, Organization Science, Organization Studies*, and *Sloan Management Review* seem to pay significantly more attention to external validity (and not much else) than do authors in *Academy of Management Journal, Administrative Science Quarterly*, or the *Strategic Management Journal*.

The proportion of reported concrete research actions (secondary reports, Table 3) of internal and construct over external validity reports is particularly illuminating. As discussed above, relatively more emphasis should be put on construct and internal validity than on external validity. Specifically, Yin (1994) suggests four strategies that can be used to address external validity and a total of 12 strategies for construct and internal validity (Table 1). This points to a relationship of construct and internal validity strategies to external validity actions of 3:1. *Academy of Management Journal, Administrative Science Quarterly*, or the *Strategic Management Journal* exhibit a relationship of 1:1, whereas the other seven journals exhibit a strong imbalance toward external validity. Thus, while these practices are clearly not perfect they at least balance attention between validity types (Gibbert et al., 2008).

Furthermore, the content analysis suggests that not only did best practice authors report more research actions taken to ensure internal validity and construct validity they also reported them in a different way. Specifically, internal and construct validity was treated with what might be called an "offensive" strategy. That is, authors were actively and astutely arguing their case, and almost "selling" the rationale of taking certain decisions in the particular context of the study, for example, for the selection criteria for cases, why the selected cases are comparable, and how this helps ensure causality. The article by Doz (1996) provides a good illustration:

> We selected these two cases expecting maximum differences in processes, testing the robustness of our framework. Continuation and/or restructuring or discontinuation of the alliance was taken as the dependent variable, to result from a particular array of nonequivalent independent variables characterizing the process of collaboration. (Doz, 1996, p. 59)

As illustrated also by the other quotes in Table 4, papers that address rigor extensively such as the one by Doz above clearly pinpoint the expected relationships between variables and select cases that best match the research questions at hand. In this way, the reader can appreciate the logic and purpose of the case selection procedure, *given* the specific context of the study. In other words, the reader accepts the method for pinpointing causal relationships, even if the paper does not explicitly mention the term internal validity. Miller (1997) provides another excellent illustration of how the selection criteria for case studies can be convincingly communicated to the reader:

> A particularly graphic instance of what needs elucidating and explaining is provided by two cases which happen to be almost identical in all important respects: the same decision, implemented at about the same time, in two organizations situated a matter of a few miles apart, and operating in the same industry. They provide an intriguing comparison, illustrating how the same decision can bring either success or failure, and begin to suggest some factors which may be important in the management of successful implementation. (Miller, 1997, p. 579)

Overall, when it comes to internal validity, papers addressing rigor extensively used an "offensive" strategy: they made an effort to not just relay the rationale for case selection and analysis to the reader but actually to convince the reader of its appropriateness for a given research question. By

contrast, discussions of external validity were not only more limited, but authors also seemed to adopt what might be called a "defensive" strategy. That is, rather than actively arguing why a certain study scores high on external validity, authors tended to simply acknowledge that generalizability is limited. Laurila (1997) provides an illustrative example:

> It is acknowledged that a study concentrating on one company in a specific business context is always limited with respect to statistical generalizability. The feasibility of the case study approach is thus based mainly on the opportunities it creates for observing and describing a complicated research phenomenon in a way that allows analytical generalizations [. . .] Thus the most important factors justifying a one-case case study are the selection of a representative or revelatory case (Yin, 1981) and the use of appropriate research methodology. (Laurila, 1997, pp. 222-223)

As a further manifestation of the "defensive" reporting strategy for external validity, authors often problematized generalizability in their methods section and then again in the concluding sections, when discussing limitations of the study.[7] For example, Birkinshaw et al. (2000) write in their methods section:

> The question of generalizability should be raised here. We were aware that by studying Swedish companies there would potentially be generalizability problems, because the process appears to be highly dependent on the cultural dispositions of the acquiring company's management. Furthermore, the acquisitions here were all highly international, all focused in large part on R&D operations, and all were undertaken by companies with considerable acquisition experience, so that there are several dimensions that potentially limit the generalizability. (Birkinshaw et al., 2000, pp. 402-403)

Birkinshaw et al. (2000) then raise the topic of generalizability again briefly in the limitations section:

> In terms of case study design, we are careful to acknowledge that this research cannot readily be generalized beyond the specific constraints we set, i.e. Swedish multinationals acquiring foreign companies with significant R&D operations. Our intention, [however], was to put forward a number of propositions and conceptual arguments that are not, to our knowledge, specific to the Swedish context. (Birkinshaw et al., 2000, p. 242).

In contrast to the papers that seemed very concerned with rigor, and which prioritized internal and construct validity over external validity both in tone as well as in substance, papers addressing rigor less extensively did the exact opposite. Rather than focusing on the more fundamental validity types, they focused on what Cook and Campbell (1979) consider the least important validity type: generalizability. This emphasis on generalizability as one of the main criteria for low N qualitative research may be understandable, given decades of debates about if and how we can generalize from samples of $N = 1$ (e.g., March et al., 1991) but is clearly counterproductive when it leads authors to de-emphasize the criteria ensuring that we actually study what we believe we study (construct validity) and that we have made a conscientious effort to avoid drawing conclusions from spurious relationships (internal validity).

In summary, best practice authors fell into line with the notion that a logical prerequisite for external validity is a case study's internal and construct validity. That is, case study authors addressing rigor extensively demonstrate that they not only are aware of the four validity and reliability criteria, they also demonstrate that they are aware of the relationships among them (Campbell & Stanley, 1963; Cook & Campbell, 1976, 1979). Specifically, our results suggest that these authors upheld the standards for rigor in case study research that require case studies first to prioritize internal validity and construct validity over external validity. Thus, the second strategy shows that rigor is about getting the priorities right not doing "last things first."

Necessity Is the Mother of Rigor: Creatively Use Setbacks and Make Best Use of Existing Resources

As a third strategy, papers addressing rigor extensively laid open not only how the individual validity and reliability types were addressed (talk the walk), and within a specific order (prioritization of internal and construct validity over external validity), but also described some of the challenges and problems that were encountered in doing so. Many methodologists have observed that qualitative research in general and case study research in particular is difficult to plan and execute (e.g., Eisenhardt, 1989; Denzin & Lincoln, 1994; Yin, 1994). Therefore, some decisions that were taken a priori need to be reconsidered in the field, for example, because selected companies, teams, or individuals no longer are available for data collection, or because methodological frameworks do not apply neatly to the studied context. An interesting insight from comparing the papers addressing rigor extensively with the papers addressing rigor less extensively is that papers in the former category actually report these issues, along with the means they used to deal with them, rather than *not* reporting them and coming up with a "neat" methodology section (as did case studies addressing rigor less extensively). For example, as illustrated in Table 4, papers that are less concerned with rigor often reported something like: "Interviews were audio recorded, transcribed verbatim, and content analyzed. Data analysis and interpretation focused on the specific interview questions as well as on unanticipated issues that emerged from the data itself" (Andrews & Delahaye, 2000, p. 799). The methods section of that particular article stops here, that is, it stops where things start to get interesting for the reader. Despite this, no further mention is made of why these issues were "unanticipated," if they influenced data analysis, and if they prompted further data collection.

By contrast, papers addressing rigor extensively did address unexpected findings and problematized how this influenced data analysis and possibly further data collection. Consistent with the notion that internal and construct validity should be prioritized over external validity, the third strategy of creatively using setbacks to increase rigor was most evident in these two criteria. First, as regards construct validity, best practice authors problematized trade-offs and compromises in ensuring rigor in an often disarmingly honest and open way. Specifically, they acknowledge the difficulties certain methodological practices bring in reality (e.g., Laurila, 1997):

> This paper is based on archival and oral history evidence despite the difficulties involved in this kind of retrospective interpretation. In brief, the study consists of a retrospective analysis of the development of a paper industry firm and its paper mill through numerous written documents and in-depth interviews of the key actors. (Laurila, 1997, p. 224)

Moreover, authors not only report these issues but often make "virtue of necessity" by actually trying to take advantage of serendipities and emerging problems. Thus, in addition to acknowledging issues, authors also explain how certain trade-offs and compromises could *still* have some value in the context of the research project. Doz' (1996) quote in Table 4 provides a good example:

> The three cases offered the *additional benefit* of being multipoint partnerships, involving different technologies and products, allowing comparison of process variable differences between the 'cases within the case' while holding overall corporate and geographic contexts constant. The fact that some of the individual projects, within the same alliance, were significantly more successful than others allowed significant analysis of project-level conditions, as distinct from partnership or corporate-level conditions. (Doz, 1996, pp. 57-59, emphasis added)

Second, as regards internal validity, typically, the data analysis strategy and concrete operationalization were explicitly derived from the literature (e.g., Glaser & Strauss, 1967; Miles & Huberman,

1984), a key requirement in Yin's framework (Yin, 1994). At the same time, papers addressing rigor extensively went much beyond simply citing a specific method and instead reported how it was applied in the concrete context studied, what some of the difficulties were, and even how authors took advantage of serendipities and difficulties that make the studied context particularly suitable for the application of certain rigor criteria, thereby greatly enhancing the "credibility" of the research (Silverman, 2005, 2006):

> To make access to data easier, a clear trade-off was made in favor of the retrospective analysis of relatively recent partnerships over the real time process study of unfolding partnerships. We took such an historical perspective also to control for differences in the collaborative experience of the companies: in each case the collaboration projects we studied were the first major strategic partnership in which the partners, or at least the involved subunits engaged into. [sic.] We also observed from other, real-time, process studies [. . .] that it was extremely difficult to be sure not to influence ongoing processes and still maintain a legitimate presence in the field insofar as managers would quickly be tempted to seek advice from the researcher and ask the researcher to intervene in the process, as a quid pro quo for allowing further observations. (Doz, 1996, p. 58)

A closely related strategy was to "make use of what is at hand" in the best possible way. Specifically, papers addressing rigor extensively seem to have looked at the resources they have at hand (data, research assistants, data collection circumstances), asking themselves how to make the most of these resources in terms of increasing rigor. For instance, many business schools employ several case writers to write case studies for teaching purposes. The additional benefit from a rigor perspective of having several case studies on similar topics (but written by different case writers) is illustrated in the following quote:

> Most of the data collection on AT&T/OIivetti and GE/SNECMA was done by research associates who had not been involved with the initial framework development, and were not deeply familiar with the framework. [O]ne of the researchers not directly involved in the field research at any of the six companies use[d] all of the detailed transcripts of interviews to test the framework. This allowed someone new to both the data and the framework to run an independent test of the pattern matches between the various cases. (Doz, 1996, p. 59)

In summary, papers that seemed less concerned with rigor often had overly "neat" methods sections where the research strategy was implemented almost to the letter as it was planned. On the other hand, papers addressing rigor extensively, while by no means having "messy" methodology sections, nevertheless pointed to problems encountered on the way and used these emerging issues to guide further data collection and analysis. Often, the purpose was not to solve these problems but instead to take advantage of them as datapoints in their own right.

Conclusion and Implications for Authors and Editors

The purpose of this article is to contribute to the ongoing debate about criteria for ensuring the rigor of published case studies by shedding more light on *what* rigor actions get reported in published case studies and revealing *how* authors can best report these actions so as to ensure the rigor of their published work. Quantitative and qualitative analysis of 159 case studies published in leading journals suggests three strategies for ensuring rigor. Rather than subscribing a priori to a particular research tradition, we take an evidence-based approach that is anchored in published work: we compare papers that address rigor extensively with papers that address rigor less extensively. Our approach contributes to the emerging work on rigor, which is based on empirical evidence (published articles), rather than theoretical considerations (e.g., Piekkari et al., in press; Pratt, 2008). At the same time,

our approach is in contrast to other research on rigor in qualitative research, which, while based on research practice, relies on self-reports of authors based on their published as well as unpublished work (e.g., Pratt, 2008). As such, the three strategies we identified contribute to the methodology literature as follows. First, authors that are relatively more concerned with rigor seem to stay away from subscribing explicitly to a particular research tradition. Empirically, this was evident from the significant gap between primary reports (the explicit mentioning of specific validity and reliability types) and secondary reports (the reporting of the concrete research actions taken to ensure validity and reliability). As discussed in the theoretical part, authors from positivist as well as interpretivist persuasions seem to largely agree on the secondary reports that should make it into the final manuscript, if not on the primary reports. Thus, the results of the present research point to some common ground between the two broad camps of positivists and interpretivists. Given that methodologists have been in theoretical disagreement as to which criteria to use, the observed convergence of actual research actions from an empirical perspective points to some much-needed common ground in reporting conventions-in-use (e.g., Guba & Lincoln, 2005, 1989; Pratt, 2008; van Maanen, 1998).

Second, the papers that addressed rigor extensively do not simply "tick off" validity and reliability criteria individually but instead consider their hierarchical relationship (Cook & Campbell, 1979). Specifically, internal and construct validity are given more attention than external validity. Although the three types of validity build on each other, case study authors are exhorted in the literature to pay special attention to this criterion. Eisenhardt's (1989) influential paper argued how (multiple) case studies could be used to build theories and enhance external validity. March, Sproull, and Tamuz have provided guidelines for "learning from samples of one or fewer" (March et al., 1991, p. 1). Thus, the finding that papers addressing rigor extensively prioritize internal and construct validity over external validity provides a much-needed argument to counterbalance the inappropriate bias of case study theorists and practitioners toward external validity.

Third, many methodologists have cautioned case study authors that actual research procedures in the field are very likely to differ significantly from the planned strategy (Campbell, 1975; Eisenhardt, 1989; Yin, 1994). At the same time, so far, there has not been much guidance as to what exactly authors should do when the research strategy "as planned" differs from the research strategy "as implemented" (to borrow Henry Mintzberg's terminology). Authors that are particularly concerned with rigor seem to make use of their constraints, such as setbacks in the field, or changes to the procedures that were necessitated due to unexpected resource and data-access constraints, and more generally take advantage of the resources at hand, actively problematizing the "emergent" research strategy. Taken together, these strategies make not only for more interesting reading but increases also the credibility, transparency, and persuasiveness of the research procedures (Siggelkow, 2007; Silverman, 2005, 2006).

In summary, revealing what rigor criteria are used in published case studies and how is of interest to authors, as well as reviewers and editors of case studies. *Authors* may appreciate an insight into editorial practices of high-ranking journals; understanding what rigor actions get reported in published articles, and how, may improve the chance of acceptance. Authors should therefore focus on reporting the concrete research actions taken (rather than abstract validity and reliability types), carefully discussing how they were adapted and used in the studied context. Golden-Biddle and Locke (1997) used the notion of showing versus telling: the importance of balance between providing raw data for your readers (showing) and explaining your data (telling). Our results suggest that papers addressing rigor extensively strive for this balance also when it comes to the methods section. Specifically, an overly "neat" storyline (the research strategy worked out exactly as planned) may disguise or downplay the "real" story that may have been much more emergent. Including the considerations that led to this emergent strategy, and therefore showing, rather than just telling, may improve rigor of the research procedures and credibility of the results.

Based on our results, *reviewers and editors* may understand better how authors address the issue of rigor in different journals, what the benchmarks are, and how to help authors increase the rigor of their work. Specifically, reviewers and editors are encouraged to look beyond the "usual suspects" of case study rigor (e.g., generalizability) and to consider the prioritization of validity and reliability types as discussed above. At the same time, reviewers and editors are cautioned to not overburden authors. Because no given piece of research is likely to be able to satisfy all validity and reliability criteria (Cook & Campbell, 1979), authors need to take (and be encouraged to report) trade-off decisions in an informed way, namely by prioritizing internal and construct validity over generalizability. In other words, an author wishing to address rigor extensively may emphasize internal and construct validity at the expense of generalizability but not vice versa.

Limitations and Implications for Future Research

A number of limitations suggest avenues for further work. First, the results presented here may not be characteristic of the 10 journals surveyed (especially given the short time period studied). In particular, we did not test the effects of changes in editorial staff (editors-in-chief or associate editors) of the journals during the time period we investigated. We also focused squarely on for-profit business organizations, so studies focusing on nonprofit or public sector organizations are very much in order.

Second, it would also be useful to study longer time periods to test for longitudinal effects and to find more case studies. Perhaps a comparison between earlier and later time periods would have indicated an increase in rigor over time. We acknowledge that one time period may not be sufficient to establish trends and encourage the use of two time periods for comparative purposes. Despite these limitations, we hope that we have suggested a conceptual template, data coding, and analysis approach that may be useful for further reviews of case study research.

Third, this study relied almost entirely on what researchers (and implicitly also reviewers and journal editors) *reported* on the research procedures. For instance, if a paper does not provide any primary or secondary reports on "reliability," this does not necessarily mean that the researcher did not consider this issue. It is possible that editors requested detailed information on the research process from the authors but eventually decided not to publish such information. We think, however, that editors should encourage, rather than discourage, case study authors to make use of the conceptual template discussed in this article, that is, to consider and report validity and reliability actions in the right order.

Notes

1. The rationale for choosing a delayed-yet-recent time period (1995–2000) was initially to also assess the effect of rigor on impact (citation counts). We tested for the antecedents of papers' citation scores in terms of case study rigor as well as journal impact scores. For this purpose, we collected the individual journal impact factors 2003–2007 from ISI. However, we find that the paper impact scores are entirely explained by the journal impact factors.

2. There are two main views on triangulation. One view, which figures in the model of validity used in the present article, is that triangulation should lead to convergence which demonstrates validity. Another view, perhaps more consistent with the interpretive perspective, is that different sources of data may reveal different insights since different data sources capture different information. The validation approach used in this article relies on the first view. We would like to thank the Associate Editor for pointing out this contingency.

3. Silverman discusses deviant-case analysis by reference to an interesting study of reports by relatives about family members who had died alone. In that study, most relatives reported that a family member dying alone was an unwelcome event and that they wished to have been present at the death, had they been able to. Seale used this evidence to argue that such accounts demonstrated a relative's feeling of moral adequacy.

However, in a small minority of cases people said that they had not wanted to be present at such a death. Rather than treating these examples as insignificant, Seale examined them in greater detail and found that in all cases, respondents actually gave legitimizations for their position. Seale concluded that, rather than disconfirming the moral adequacy hypothesis, they actually further strengthened it. The reason, according to Seale was that by offering legitimizations, the people in the deviant cases actually showed that they themselves perceived their orientation as deviant from the social norm. Thus, the analysis of deviant cases can strengthen the internal validity of research (Silverman, 2005, pp. 215-219).

4. The journals in our sample are ranked nos 2, 6, 24, 11, 7, 23, 14, 13, 8, 26, 36, and 1, respectively (Tahai & Meyer, 1999, p. 291). These were the 12 most impactful general management journals, after the more specialized journals were left out. For example, *Journal of Applied Psychology* ranked as no 3, *Organizational Behavior and Human Decision Processes* ranked as no 4, and *Academy of Management Review* ranked as no 5. Thus, neither of the journals that were left out were management journals and/or reported empirical studies such as case studies.

5. To check this assumption, we interviewed at least one of the authors of all articles in the top rigor bracket. Interviews revealed that none of the reviewers demanded methods sections be changed on ideological grounds (e.g., that positivist approaches for ensuring rigor be used). In many cases, however, methods sections changed in volume.

6. For space constraints, we report only these ''exemplars.'' However, full tables detailing secondary reports into the four validity and reliability categories are available from the communicating author.

7. Generalizability was the only rigor criterion that was addressed in the concluding sections of an article. All the other rigor criteria were dealt with only in the methods section.

8. Numbering consistent with Table 1.

Acknowledgments

Declaration of Conflicting Interests

Funding

References

Amis, J. M., & Silk, M. L. (2008). The philosophy and politics in qualitative organizational research. *Organizational Research Methods*, *11*, 456-480.

Andrews, K. M., & Delahaye, B. L. (2000). Influences on knowledge processes in organizational learning: The psychological filter. *Journal of Management Studies*, *37*, 796-810.

Bacharach, S., Bamberger, P., & McKinney, V. (2000). Boundary management tactics and logics of action: The case of peer-support providers. *Administrative Science Quarterly*, *45*, 704-736.

Bartunek, J. M., Rynes, S. L., & Ireland, R. D. (2006). Editors' forum: What makes management research interesting, and why does it matter? *The Academy of Management Journal*, *49*, 9-15.

Behling, O. (1980). The case for the natural science model for research in organizational behavior and organization theory. *Academy of Management Review*, *5*, 483-490.

Benbasat, I. (1984). An analysis of research methodologies. In F. Warren McFarlan (Ed.), *The Information systems research challenge* (pp. 47-85). Boston, MA: Harvard Business School Press.

Benbasat, I., Goldstein, D., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*, *11*, 369-386.

Bergh, D. G., Perry, J., & Hanke, R. (2006). Some predictors of SMJ article impact. *Strategic Management Journal*, *27*, 81-100.

Birkinshaw, J., Bresman, H., & Håkanson, L. (2000). Managing the post-acquisition integration process: How the human integration and task integration processes interact to foster value integration. *Journal of Management Studies*, *37*, 396-425.

Boyd, B. K., Gove, S., & Hitt, M. A. (2005). Construct measurement in strategic management research: Illusion or reality? *Strategic Management Journal*, *26*, 239-258.

Bresman, H., Birkinshaw, J., & Nobel, R. (1999). Knowledge transfer in international acquisitions. *Journal of International Business Studies*, *30*, 439-462.

Buchel, B. (2000). Framework of joint venture development: Theory-building through quantitative research. *Journal of Management Studies*, *37*, 55-83.

Bullock, R. J., & Tubbs, M. E. (1987). The case meta-analysis for OD. *Research in Organizational Change and Development*, *1*, 171-228.

Bullock, R. J., & Tubbs, M. E. (1990). A case meta-analysis of gainsharing programs as organization development interventions. *Journal of Applied Psychology*, *26*, 383-404.

Burgelman, R. A. (1983). A model of the interaction of strategic behavior, corporate context and the concept of strategy. *Academy of Management Review*, *8*, 61-70.

Campbell, D. T. (1975). Degrees of freedom and the case study. *Comparative Political Studies*, *8*, 178-193.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi experimental designs for research.* Boston, MA: Houghton Mifflin.

Chandler, A. D. (1962). *Strategy and structure.* Cambridge, MA: Harvard University Press.

Cook, T. D., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in fields settings. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223-326). Chicago: Rand McNally.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimental design: Design and analysis issues for field settings.* Skokie, IL: Rand McNally.

Corbin, J. M., & Strauss, A. (1988). *Unending care and work.* San Francisco: Jossey-Bass.

Daft, R. L., & Lewin, A. Y. (1990). Can organization studies begin to break out of the normal science straight-jacket? An editorial essay. *Organization Science*, *1*, 1-9.

Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research.* Thousand Oaks & London: Sage.

Derong, C., & Faure, G. O. (1995). When Chinese companies negotiate with their government. *Organization Studies*, *16*, 27-54.

Doz, Y. (1996). The evolution of cooperation in strategic alliances: initial conditions or learning processes? *Strategic Management Journal*, *17*, 55-82.

Easterby-Smith, M., Golden-Biddle, K., & Locke, K. (2008). Working with pluralism: Determining quality in qualitative research. *Organizational Research Methods*, *11*, 419-429.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*, 532-550.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*, 25-32.

Fendt, J., & Sachs, W. (2008). Grounded theory method in management research: Users' perspectives. *Organizational Research Methods*, *11*, 430-455.

Fleiss, J. L. (1971). Measuring nominal scale agreement among multiple raters. *Psychological Bulletin*, *76*, 378-382.

Geertz, C. (2003). *The interpretation of cultures: Selected essays.* New York: Basic Books.

Gerring, J. (2007). *Case study research: Principles and Methods*. Cambridge University Press. Boston.

Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader-member exchange theory. *Journal of Applied Psychology*, *82*, 827-844.

Gephart, R. P. (2004). Qualitative research and the academy of management journal. *Academy of Management Journal*, *47*, 454-462.

Gibbert, M., Ruigrok, W., & Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal*, *29*, 1465-1474.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory.* Chicago, IL: Aldine Press.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park: Sage.

Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 191-215). Thousand Oaks: Sage.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, *24*, 602-611.

Judge, D. A., Cable, D. A., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited? Article, author, or journal? *Academy of Management Journal*, *50*, 491-506.

Kidder, L., & Judd, C. M. (1986). *Research methods in social relations* (5th ed.). New York: Holt, Rinehart & Winston.

Kirk, J., & Miller, M. L. (1986). *Reliability and validity in qualitative research.* London: Sage.

Lane, D., & Maxfield, R. (1996). Strategy under complexity: Fostering generative relationships. *Long Range Planning*, *29*, 215-231.

Larsson, R. (1993). Case survey methodology: Quantitative analysis of patterns across case studies. *Academy of Management Journal*, *36*, 1515-1546.

Larsson, R., & Finkelstein, S. (1999). Integrating strategic, organizational and human resource perspectives on mergers and acquisitions: A case of synergy realization. *Organization Science*, *10*, 1-26.

Laurila, J. (1997). The thin line between advanced and conventional new technology: A case study on paper industry management. *Journal of Management Studies*, *34*, 219-239.

Lee, A. S. (2003). Generalizing generalizability in information systems research. *Information Systems Research*, *14*, 221-243.

Leonard-Barton, D. (1990). A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multi sites. *Organization Science*, *1*, 248-266.

Levy, D. L. (2000). International sourcing and supply chain stability. *Journal of International Business Studies*, *26*, 343-360.

Locke, K. (1996). A funny thing happened! The management of consumer emotions in service encounters. *Organization Science*, *7*, 40-59.

Lovas, B., & Goshal, S. (2000). Strategy as guided evolution. *Strategic Management Journal*, *25*, 801-822.

March, J. G., Sproull, L. S., & Tamuz, M. (1991). Learning from samples of one or fewer. *Organization Science*, *2*, 1-13.

Maznevski, M., & Chudoba, K. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science*, *11*, 473-492.

Miles, M. B. (1979). Qualitative data as an attractive nuisance: The problem of analysis. *Administrative Science Quarterly*, *24*, 590-601.

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Newbury Park, CA: Sage.

Miller, S. (1997). Implementing strategic decisions: Four key success factors. *Organization Studies*, *18*, 577-602.

Mitchell, T. R. (1985). An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Journal*, *10*, 192-205.

Mueller, F., & Dyerson, R. (1999). Expert humans or expert organizations? *Organization Studies*, *20*, 225-256.

Numagami, T. (1998). The infeasibility of invariant laws in management studies: A reflective dialogue in defense of case studies. *Organization Science*, *9*, 2-15.

Penrose, E. T. (1960). The growth of the firm. A case study: The Hercules powder company. *Business History Review*, *34*, 1-23.

Peters, T. J., & Waterman, R. H. (1982). *In search of excellence: Lessons from America's best-run companies.* New York: Harper & Row.

Pettigrew, A. (1973). *The politics of organizational decision making.* London: Tavistock.

Pettigrew, A. (1990). Longitudinal field research on change: Theory and practice. *Organization Science*, *1*, 267-292.

Piekkari, R., Welch, C., & Paavilainen, E. (in press). The case study as a disciplinary convention: Evidence from international business journals. *Organizational Research Methods*.

Popper, K. (1959). *The logic of scientific discovery.* Tuebingen: Hutchinson.

Prahalad, C., & Hamel, G. (1990). The core competence of the corporation. *Harvard Business Review*, *May-June*, 79-91.

Pratt, M. (2008). Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier American journals. *Organizational Research Methods*, *11*, 481-509.

Pratt, M. (2000). The good, the bad, and the ambivalent: Managing identification among Amway distributors. *Administrative Science Quarterly*, *45*, 456-493.

Rosenbloom, R. S. (2000). Leadership, capabilities, and technological change: The transformation of NCR in the electronic era. *Strategic Management Journal*, *21*, 1083-1103.

Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, *43*, 1248-1264.

Schriesheim, C. A., Powers, K. J., Scandura, T. A., Gardiner, C. C., & Lankau, M. J. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management*, *19*, 385-417.

Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal*, *50*, 20-24.

Silverman, D. (2005). *Doing qualitative research.* London: Sage.

Silverman, D. (2006). *Interpreting qualitative data.* London: Sage.

Stake, R. E. (1995). *The art of case study research.* Thousand Oaks: Sage.

Tahai, A., & Meyer, M. J. (1999). A revealed preference study of management journals' direct influences. *Strategic Management Journal*, *20*, 279-296.

Van Hoek, R. I., Voss, B., & Commandeur, H. R. (1999). Restructuring European supply chains by using postponement. *Long Range Planning*, *32*, 505-518.

Van Maanen, J. (1998). Different strokes: Qualitative research in the administrative science quarterly from 1956 to 1996. In J. Van Maanen (Ed.), *Qualitative studies of organizations* (pp. ix-xxxii). London: Sage.

Weick, K. (2007). The generative properties of richness. *Academy of Management Journal*, *50*, 14-19.

Woiceshyn, J. (2000). Technology adoption: Organizational learning in oil firms. *Organization Studies*, *21*, 1095-1118.

Yin, R. K. (1981). The case study crisis: Some Answers. *Administrative Science Quarterly*, *26*, 58-65.

Yin, R. K. (1994). *Case study research: Design and methods.* London: Sage.

Bios

Michael Gibbert is an associate professor of Management at Bocconi University, Milan. His case-based and other work on innovation under resource constraints appeared in *Long Range Planning, Journal of Consumer Psychology, Journal of Management Inquiry, Journal of Product Innovation Management, Research Policy, Strategic Management Journal*, and the *Wall Street Journal*.

Winfried Ruigrok is a professor of International Management at St. Gallen University, Switzerland. He is Academic Director, Master of Business Administration, Director, Executive School, and Director of Research Institute for International Management at St. Gallen University.