CS-E5005
Research Methods in Software and Service Engineering

Introduction

Kari Smolander
kari.smolander@aalto.fi
About the course

This course is about scientific research
Most examples and applications are taken from software engineering, but most topics can be directly transferred and used to any other scientific field

In the core is the method of science

• How can we do research so that the results are valid for science?
• This means that we must understand what is science
• “How” means that we understand the method and the practice of science
• This also includes very difficult problems in the philosophy of science
  - These will be mostly just mentioned
• “Practice of science” requires the ability to plan research and to produce valid results
  - The core objective of the course is to understand the essence of this
**From the study guide**

<table>
<thead>
<tr>
<th>Level of the Course</th>
<th>The course is only for students who have completed their Bachelor’s Degree.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching Period</td>
<td>I-II (Autumn)</td>
</tr>
<tr>
<td>Workload</td>
<td>Lectures 14 h, individual studies 16 h, assignment 100 h.</td>
</tr>
<tr>
<td>Learning Outcomes</td>
<td>Students are able to understand scientific research and evaluate the validity and reliability of research results. They are able to plan an empirical study, collect and analyze data, and report the results. Students improve their skills to carry out Master’s Thesis.</td>
</tr>
<tr>
<td>Content</td>
<td>The goal is to introduce the participants to scientific research methods, approaches and processes used in the field of software engineering. In addition, the course provides students practice on formulating research questions and planning empirical studies.</td>
</tr>
<tr>
<td>Assessment Methods</td>
<td>Assignments and classroom activity. Exam.</td>
</tr>
<tr>
<td>and Criteria</td>
<td></td>
</tr>
<tr>
<td>Study Material</td>
<td>To be announced later.</td>
</tr>
<tr>
<td>Substitutes for</td>
<td>Replaces the former course T-76.5050 Methods for Software Engineering</td>
</tr>
<tr>
<td>Courses</td>
<td>and Business Research.</td>
</tr>
</tbody>
</table>
Course web

All communication, material, assignments, etc. through MyCourses

• https://mycourses.aalto.fi/course/view.php?id=13106
Doing the course

Components

• Lectures 7x2 h – 1 point each
• Course work
  - Research planning – min 5, max 10 points
  - Research execution (a small research task)
  - Research report and seminar presentation – min 10, max 20 points
• Learning diary – 7 x 3 points = max 21
• Max points from assignments and classroom activity 58, pass 29

• Reading package for the exam
  - Needed also in the course work
  - Exam evaluated separately

• Final grade is the average of these two
Lectures

All lectures in T4 Wed 14-16

1. Introduction (Kari) 14.9.
2. Research methods and process (Sari) 21.9.
3. Data collection (Sari) 28.9.
4. Qualitative research (Kari) 5.10.
5. Quantitative research (Sari) 12.10.
6. Case studies (Sari) 19.10.
7. Action research and Design science (Kari) 2.11.

Course work presentations start 9.11.
Course work

• The course assignment is a small scientific study that you perform and report
• The assignment is done in pairs
• Collect information from people, get somewhat generalizable and interesting results
• The minimum number of research participants
  - Qualitative methods – 5 (both students participate in all cases)
  - Quantitative methods – 10 participants
Implementing and reporting the course work

1. Select your topic
   • Must be related to software engineering or business

2. Select an initial research question
   • The used method needs to be suitable to answer to the research question
   • The method needs to be one of those presented during the course

3. Write a research plan before the DL (5 October)

4. Perform the study
   • Reserve enough calendar time to find people and agree dates with them

5. Write the final report before the DL (8 November)

6. Present your results
   • 10 minutes presentation including 5-7 slides
Learning diary

• 2-3 questions to answer about each week’s lectures
• May include reading the articles in the reading package
• SHORT answers are required
  - Answer length will be restricted to 300 words each
• The goal is to facilitate understanding of the core of each lecture
Reading package

• A set of articles about scientific methods
• Required for exam, but useful also for course work

• The final set of articles (~5-6) will be available in MyCourses
1st Lecture: What is science? What is software engineering as science?

- Science and scientific research
- Scientific community
- Some important terms of science
- What kind of research is software engineering?
What is science and scientific research?

Narrow meaning of science
- The intellectual and practical activity encompassing the **systematic** study of the **structure** and **behaviour** of the **physical** and **natural** world through **observation** and **experiment**
- Produces testable explanations and predictions about the universe.

Broad meaning of science
- A **systematically organized body of knowledge** on a particular subject
- Reliable and teachable knowledge about a topic

What does scientific research mean in software engineering?
- Is software engineering a study of physical and natural world?
  - Make two examples of for and against.
Background to previous: Software engineering

IEEE 610.2 Standard Glossary of Computer Applications Terminology

• The application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software

Two views on software engineering

• Software as a technical artifact
  - The structural qualities of software
  - The qualities of software development tools

• Software development as an industrial activity
  - Development processes
  - How human organizations work when developing software
  - Practical usability and applicability of tools and development principles
How science works?

Intersubjective pattern recognition or intersubjective verifiability
• Concepts that can be accurately communicated between individuals
• Basis of empirical and scientific investigation

Scientific method
• Explanation of events of nature in a reproducible way
• Hypothesis = proposed explanation of a phenomenon
• Hypothesis is tested by experimentation or observation
• If a hypothesis survives, it can be adopted into a scientific theory
• Theory = logically reasoned, self-consistent model or framework for describing the behavior of natural phenomena

The scientific method is the basis of natural sciences, but not necessarily of others (we will discuss this later)
Mathematics is very essential to other sciences, especially the natural ones.
Statistical methods are very important to most sciences.
Computation and computational sciences enable the advancement of many other sciences.
Is mathematics a science? Are pure computational sciences really sciences?
- They do not study natural world
- No experimentation or observation are required

A difference between formal sciences and empirical sciences, such as natural and social sciences.
Applied research vs. basic research

- Systematic inquiry involving the practical application of science
- Driven by utility and solves practical problems using theories, knowledge, methods and techniques of science

- Basic research is driven by curiosity
  - Searches knowledge
  - Have led to unexpected and unimaginable technological advancements
Scientific community

• The group of all interacting scientists
• Many sub-communities in many levels on particular fields
• Also: interdisciplinary communities
• Who decides what research is valid and of high quality?
  - Answer: the scientific community
  - There is a hierarchy, but also independency and peer review
• Peer review
  - Most research writings are selected for publication by peer review
  - Typically 2-4 competent reviewers evaluate the manuscript
• Important results provoke intensive discussion and counter-publications
• In some fields (but not in all!) the results must be repeated by independent researchers before they are generally accepted
Some important scientific terms

• Deduction and induction
• The problem of induction
• Falsifiability
• Accumulation of knowledge and scientific revolutions
• Theory
• Validity
Deduction and induction

Two basic forms of reasoning
• Deductive reasoning
• Inductive reasoning

Deduction
- Observation
- Confirmation

Induction
- Observation
- Pattern
- Tentative hypothesis
- Theory

Theory
- Hypothesis
- Observation

Aalto University
School of Science
Research methods
2016
Deduction and induction (2)

Deductive reasoning

All Frenchmen like red wine
Pierre is a Frenchman

⇒ Therefore, Pierre likes red wine

Inductive reasoning

The first five eggs in the box were rotten
All the eggs have the same best-before date stamped on them

⇒ Therefore, the sixth egg will be rotten too

(Okasha, 2002)
The problem of induction

The famous example
• “All swans are white”
  - *This was confirmed true by observing individual swans = induction*
  - *Until a black swan species was found from Southern Australia*

Induction cannot prove something 100% true
• This problem applies to most research!
  - “*Genetically modified maize is safe for humans*”
    - Tested with a large number of humans and none of them suffered any problems
  - *Most statistical research use a sample of a population*
    - Sometimes it is possible to use the whole population
  - *All qualitative research (more later) is based on induction*

Karl Popper’s (1934) solution: all scientific theories must be falsifiable
• A single experiment or observation can falsify a theory
Falsifiability

• Karl Popper (1934), “Logik der Forschung”
• Also: refutability
• Empirical theories are characterized by falsifiability
  - It must be possible to prove a scientific theory false

• An asymmetry between verifiability and falsifiability
  - One contradictory observation is enough to falsify a universal statement (“all swans are white”)
  - How many observations are needed to verify a universal statement?
Accumulation of knowledge

Ideal: science constructs a coherent and consistent theory on which knowledge is accumulated over time. The knowledge and theory approaches perfection over time.

Two problems

• Scientific revolutions
• It is not always possible to fully agree on a coherent and consistent theory (e.g. in human sciences)
  - The observation and results require interpretation of human action

Thomas Kuhn: “The Structure of Scientific Revolutions” (1962)

• “Normal science” accumulates knowledge (normal = dominant paradigm)
• Anomalies cause a crisis and a scientific revolution, where the accumulated knowledge is not necessarily valid any more
• A paradigm shift and a new paradigm follows
  - For example from Ptolemaic to Copernican cosmology
  - From aether theory to electromagnetic radiation
What is theory?

Theory is abstract and generalized thinking

- Can be normative/prescriptive or not
- Can be a body of knowledge
- A well-substantiated explanation of some part of the world, based on facts

In contrast with “practice” (praxis – doing)

Different fields require different kind of theories

- Theories of physics
  - A mathematical framework that is capable of producing predictions of physical systems
- Theories of society
  - Often conceptual generalizations of empirical observations
Table 2. A Taxonomy of Theory Types in Information Systems Research

<table>
<thead>
<tr>
<th>Theory Type</th>
<th>Distinguishing Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Analysis</td>
<td>Says what is. The theory does not extend beyond analysis and description. No causal relationships among phenomena are specified and no predictions are made.</td>
</tr>
<tr>
<td>II. Explanation</td>
<td>Says what is, how, why, when, and where. The theory provides explanations but does not aim to predict with any precision. There are no testable propositions.</td>
</tr>
<tr>
<td>III. Prediction</td>
<td>Says what is and what will be. The theory provides predictions and has testable propositions but does not have well-developed justificatory causal explanations.</td>
</tr>
<tr>
<td>IV. Explanation and prediction (EP)</td>
<td>Says what is, how, why, when, where, and what will be. Provides predictions and has both testable propositions and causal explanations.</td>
</tr>
<tr>
<td>V. Design and action</td>
<td>Says how to do something. The theory gives explicit prescriptions (e.g., methods, techniques, principles of form and function) for constructing an artifact.</td>
</tr>
</tbody>
</table>
Validity of scientific reasoning

How to evaluate the validity of a proposition, inference or conclusion in research?

An extremely difficult philosophical question that includes
- Epistemology – theory of knowledge
- Ontology – theory of reality

Each field and research methodology may have differences in the criteria for validity

Perhaps we can discuss this later.
Positivistic method

The method of natural sciences

• The purpose of science is explanation, finding general laws and causal connections
• Science is value-free and based on objective, measurable facts
• The researcher is an independent observer
• Studies and measurements can be repeated any time and by anyone with the same results
• Research results increase the amount of scientific knowledge
• The accumulated scientific method forms an internally harmonious system (no conflicts)
Is positivistic method the criterion of science?

In natural sciences only the research that fulfils the criteria in the previous slide is acceptable
• The method of natural sciences is the typical view of scientific method

The method of human sciences
• The positivistic method does not work in social sciences where the understanding and interpretation of human activity is in the center
  - How to measure unambiguously attitudes, social norms, beliefs, perceptions, fears, motives, etc.?
  - The researchers cannot be completely neutral, independent, and value-free
  - The presence of a researcher may have an effect on the studied activity
  - It is often not possible to repeat the study with unchanged conditions

Is software engineering a natural science?
Software engineering

IEEE 610.2 Standard Glossary of Computer Applications Terminology

• The application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software

Two views on software engineering

• Software as a technical artifact
  - The structural qualities of software
  - The qualities of software development tools

• Software development as an industrial activity
  - Development processes
  - How human organizations work when developing software
  - Practical usability and applicability of tools and development principles
Software development as human activity

Software is a technical artifact (although there is much human knowledge encoded in it) but software development is a human activity

• How to study human activities?
• How to measure and evaluate human activities?

It is (relatively) easy to measure qualities and properties of software, but more difficult to measure and evaluate software development
Software development as human activity

What affects on human activities?

• Genetics
• Attitudes
• Social norms
• Perceived behavioral control, experiences on difficulties and rewards
• Faith, beliefs, religion, philosophy
• Instincts, fear, habits
• Etc.

How to measure these unambiguously in software development?

• Is it possible?
An example

- **Agile manifesto** (2001)
- This is **not** about software technology
- This is **about** human activity/behaviour

---

**Manifesto for Agile Software Development**

We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

- **Individuals and interactions** over processes and tools
- **Working software** over comprehensive documentation
- **Customer collaboration** over contract negotiation
- **Responding to change** over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

Kent Beck  
Mike Beedle  
Arie van Bennekum  
Alistair Cockburn  
Ward Cunningham  
Martin Fowler  
  
James Grenning  
Jim Highsmith  
Andrew Hunt  
Ron Jeffries  
Jon Kern  
Brian Marick  
  
Robert C. Martin  
Steve Mellor  
Ken Schwaber  
Jeff Sutherland  
Dave Thomas
Does software engineering study humans?

An example: a researcher wishes to study how agile methods can be taken into use in a very large software organization.

Software engineering research must often evaluate the actions of individuals and human organizations:

- It may be essential to understand the motives, values, expectations and objectives of individuals and organizations.
- Software engineering must take also humans as the research subjects.
Questions?
Comments?
Lectures

All lectures in T4 Wed 14-16
1. Introduction (Kari) 14.9.
2. Research methods and process (Sari) 21.9.
3. Data collection (Sari) 28.9.
4. Qualitative research (Kari) 5.10.
5. Quantitative research (Sari) 12.10.
6. Case studies (Sari) 19.10.
7. Action research and Design science (Kari) 2.11.

Course work presentations start 9.11.