Empirical Methods in Software Research: Experiments with Human Subjects

Prof. Claes Wohlin,
Blekinge Institute of Technology,
Sweden

Professorial Visiting Fellow, University of New South Wales, Sydney, Australia

Today's objective

- A context for empirical methods when doing research and working with industry.
- An in-depth introduction to the experimental process for running experiments with human subjects.

Software Engineering

Our objective:

SERL is focused on working together with its partners to create novel software engineering research solutions to real long-term industrial challenges.

This is our context when doing empirical research, so how do we do it?

A key challenge: conduct empirical work with industry in a win-win situation.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Combine research methods

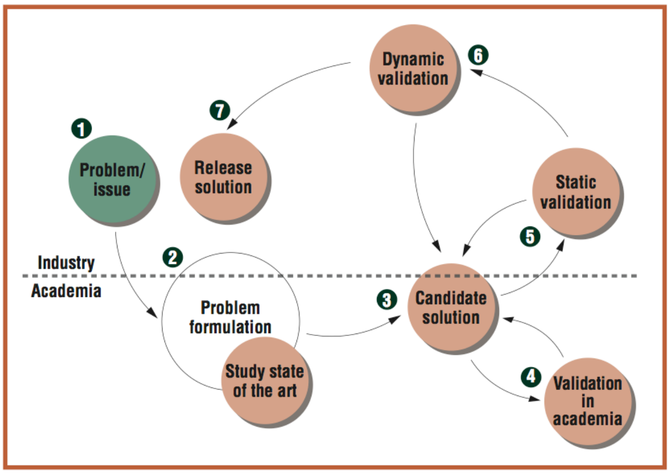
Main methods used:

- Research-in-the-typical: case study
- Research-in-the-large: survey
- Research-in-the-small: experiment

Often mixed designs with different, for example, sub-methods within a case study, which may include interviews and archival analysis.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Our way of doing research: knowledge exchange



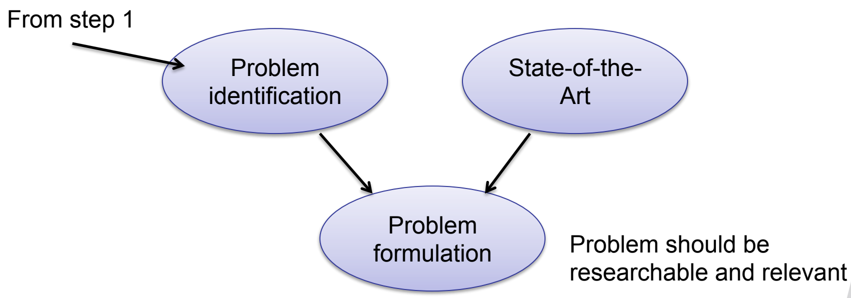
1. Problem/Issue

We often start with:

- Interviews to capture needs
- Perform assessment of processes, products or projects in industry typically both from general descriptions and actual project documentation
- Problems prioritized by industry

2. State-of-the-art and Problem formulation

State-of-the-art is studied often using a systematic review.



www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

3. Candidate solution

A solution to the problem formulation or part of it is proposed based on literature and own inventions in close contact with our collaborative partner.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

4. Academic validation

This is typically done in a controlled experiment with students. The experiment is often done as part of a course and we ensure that we connect a learning objective to the experiment. This may results in improvements of solutions.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

5. Static validation

This refers to validation done in industry, but offline. This is typically done by interviewing a set of people in relation to the solution. Their feedback is taken into account and the solution is potentially revised.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

6. Dynamic validation

This is typically a pilot in a project, part of a project or part of a product. The objective is to evaluate the solution ”live”, but to minimize costs and risks. A case study is used to follow up on the pilot.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

7. Release

Two aspects are important:

- Academically: publications are submitted as we go along in the collaboration.
- Practically: solution should if passing through all steps be released for broader use.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Success stories: generic methods and tailored solutions

- Success 1 – Ericsson
  A method (fault-slip-through) for mapping fault detection to test strategy has been developed and introduced. Savings of about 30% are reported. The method is now being adopted as Ericsson-standard worldwide.
- Success 2 – ABB
  A method for requirements management has been introduced at ABB. A return on investment of about 5-15 times is reported. The method has also been introduced as DanaherMotion and is also now in the process of being implemented at Volvo and Swisscom.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Summary: empirical methods

Use different methods, we typically use:

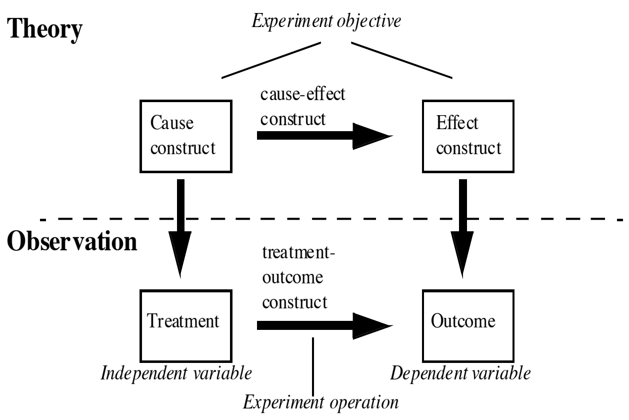
- Interviews and archival analysis in the start
- Systematic reviews to capture literature
- Experiments for first validation ←
- Interviews for first industry evaluation
- Case study for pilot evaluation
- Case study for release of solution

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Experiment

- Experiments are carefully planned and fully controlled. An experiment should be replicable, i.e. somebody else should be able to repeat it.
  - This type of method will be used to illustrate empirical studies with human subjects.

Experiment principle



www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Independent and dependent

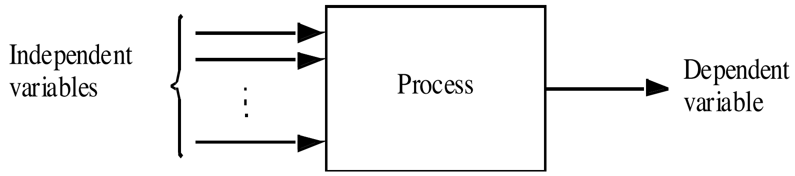
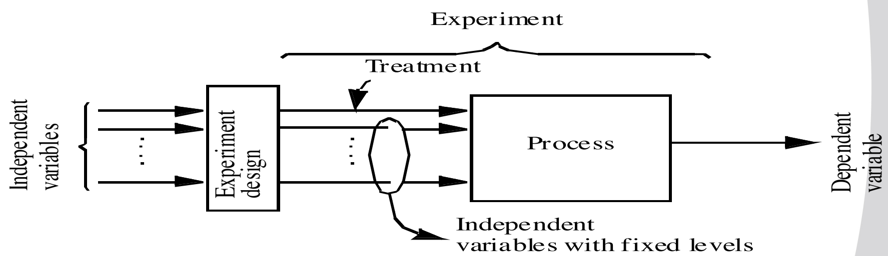
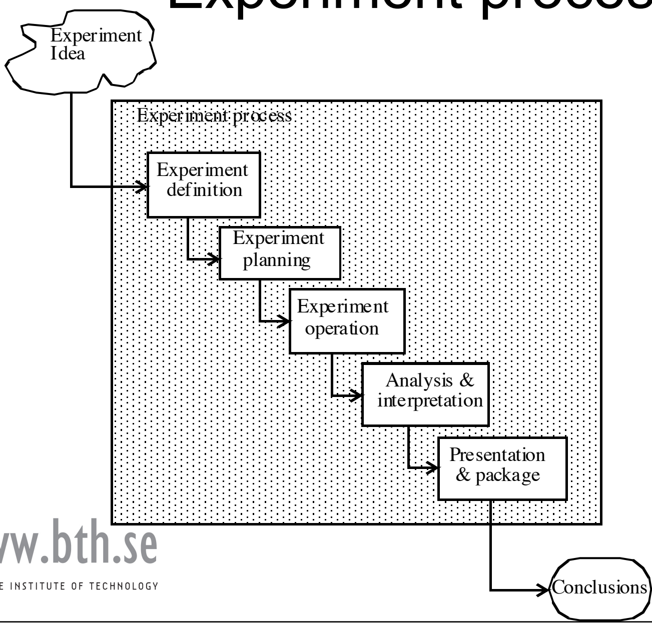


Illustration of experiment



Experiment process



Experiment definition

- The goal template is:

  Analyse <*Object(s) of study*>
  for the purpose of <*Purpose*>
  with respect to their <*Quality focus*>
  from the point of view of the <*Perspective*>
  in the context of <*Context*>.

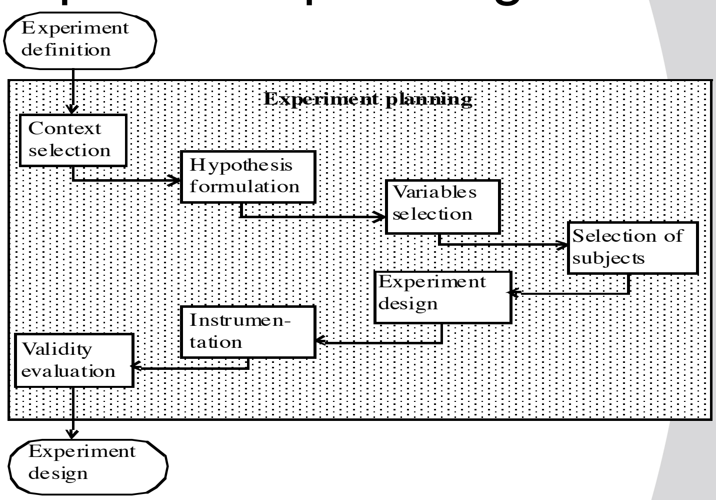
An example definition

Analyse *the PBR and checklist techniques*
for the purpose of *evaluation*
with respect to *effectiveness and efficiency*
from the point of view of *the researcher*
in the context of *students reading requirements documents*.

PBR – Perspective-based reading

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Experiment planning



Planning phase overview

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Steps in planning 1(4)

- Context:
  - Off-line vs. On-line
  - Students vs. Professionals
  - Toy vs. Real problems
- Hypothesis formulation:
  - Null hypothesis (no real underlying trend or pattern) and alternative hypothesis. The objective is to reject the null hypothesis with as high significance as possible.

Steps in planning 2(4)

- Variables:
  - Independent (input)
  - Dependent (output)
- Subjects
  - Sampling strategy, sampling from population
- Design principles
  - Randomization
  - Blocking (e.g. on experience)
  - Balancing (same number of subjects in groups)

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Steps in planning 3(4)

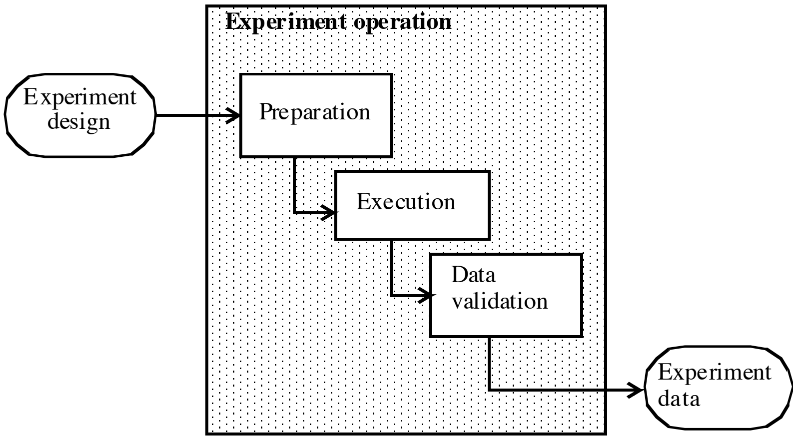
- Design types

  A large number of standard designs exist, and we should select an appropriate design type depending on treatments and number of subjects and of course the objective (hypothesis) of the experiment.

Steps in planning 4(4)

- Instrumentation
  - Objects
  - Guidelines
  - Measurement instruments
- Validity evaluation
  - Conclusion validity: treatment to outcome (observation), ability to draw correct conclusions
  - Internal validity: treatment causes outcome, threat to causal relationship
  - Construct validity: relationship between theory and observation
  - External validity: generalization

Experiment operation

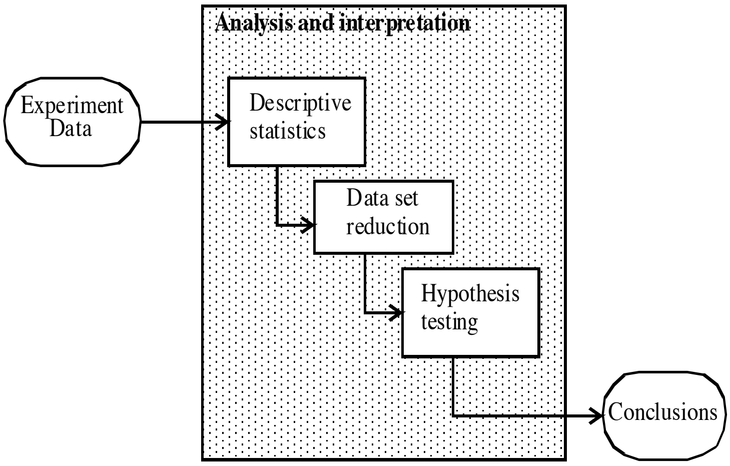


Steps in operation

- Preparation
  - Commit participants
  - Instrumentation (availability)
- Execution
  - Data collection
  - Experimental environment
- Data validation (general check)

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Analysis and interpretation



Steps in analysis 1(2)

- Descriptive statistics
  - Scale types (nominal, ordinal, interval and ratio)
  - Measures of central tendency, dispersion and dependency
  - Graphical visualization
- Data set reduction
  - Outliers

Steps in analysis 2(2)

- Hypothesis testing
  - Parametric tests (assumes a specific distribution, usually a normal distribution)
  - Non-parametric tests (no assumption on distributions)

The different types of tests are related to the standard design types. The intention is to be able to reject the null hypothesis with a statistical significance.

Interpretation

The statistical analysis forms the basis for interpretation.

The interpretation is the foundation for decision-making based on engineering principles.

Packaging

Report outline:

- Introduction
- Problem statement
- Experiment planning
- Experiment operation
- Data analysis
- Interpretation of results
- Discussion and conclusions
- Appendix

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Additional concerns

- Triangulation
- Replication
- Lab packages
- Meta-analysis

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Simplistic example: experiment

- Problem: We want to evaluate reading techniques for inspections.
- We have two competing methods. State a null hypothesis, for example method A and method B show no difference in defect detection for requirements specifications.
- Let people inspect a requirements specification with a known number of defects.
- Use a statistical method to evaluate the hypothesis. Could the null hypothesis be rejected?
- Determine which method is the best.
- Decide whether or not to start using the method.

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Some conclusions

- There is a lack of validated results in the field,
- Empirical studies is needed in software engineering to evaluate and validate development process activities,
- Empirical studies mean that the human dimension in software development can be included in the analysis.

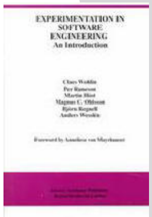
www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

Selection of publications for illustration

- Wohlin et al., ”Experimentation in Software Engineering – An Introduction”, Kluwer Academic Publishers, 1999.
- Höst, Wohlin and Thelin, ”Experimental Context Classification: Incentives and Experience of Subjects”, Proceedings ICSE '05, pp 470-478, 2005.
- T. Gorschek, P. Garre, S. Larsson and C. Wohlin, "Technology and Knowledge Transfer in Practice - Requirements Engineering Process Improvement Implementation", IEEE Software, Issue November/ December, pp. 88-95, 2006.

Examples of successful use of knowledge exchange with industry:

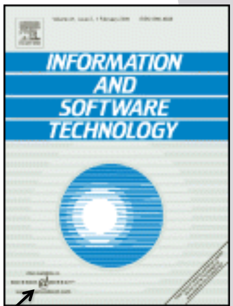
- T. Gorschek and C. Wohlin, "Requirements Abstraction Model", Requirements Engineering Journal, Vol. 11, No. 1, pp. 79-101, 2006.
- L-O. Damm, L. Lundberg and C. Wohlin, "A Model for Software Rework Reduction through a Combination of Anomaly Metrics", Journal of Systems and Software, Vol. 81, No. 11, pp. 1968-1982, 2008.



www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY

A selection of references

- Basili, Selby and Hutchens, ”Experimentation in Software Engineering”, IEEE Transactions on Software Engineering, Vol. SE-12, No. 7, pp. 733-743, 1986.
- Juristo and Moreno, ”Basics of Software Engineering Experimentation”, Kluwer Academic Publishers, 2001.
- Fenton and Pfleeger, ”Software Metrics. A Rigorous and Practical Approach” (2nd edition), International Thomson Computer Press, 1998.
- Yin, ”Case Study Research. Design and Methods” (4th edition), Sage, 2009.



Systematic reviews are welcome
Barbara Kitchenham is special editor
for systematic reviews

www.bth.se
BLEKINGE INSTITUTE OF TECHNOLOGY