Lecture 6

Case studies and experiments Validity

Kari Smolander
kari.smolander@lut.fi

Contents

- **Case study research in software engineering**
- Runeson & Höst: "Guidelines for conducting and reporting case study research in software engineering", 2008.

- **Controlled experiments**

- **Discussion about validity in empirical research**

Case study as a research method

General about case studies

- **The term "case study" appears often in the title of software engineering research papers**
- Very ambitious and well organized studies in the field
- Small toy examples that claim to be case studies
- **The term case study is used in parallel with**
- field study
- observational study
- **The terminology is not always very clear**

What is a case study

- **Empirical method aimed at investigating contemporary phenomena in their context**

- **The use of multiple sources of evidence**

- **The boundary between the phenomenon and its context may be unclear**

- **Information gathering from many entities (people, groups, organizations)**

- **Lack of experimental control.**

Types of case studies

Robson's (2002) classification of research:

- Exploratory - finding out what is happening, seeking new insights and generating ideas and hypotheses for new research
- Descriptive - portraying a situation or phenomenon
- Explanatory - seeking an explanation of a situation or a problem, mostly in the form of a causal relationship
- Improving/emancipatory - trying to improve a certain aspect of the studied phenomenon

Case study is originally used for exploratory purposes and descriptive purposes are also usual

Case studies may be used for explanatory purposes

- This involves testing of created theories in confirmatory studies

Case studies in the software engineering discipline often take an improvement approach, similar to action research

Data in case studies

- **Case studies are mostly based on qualitative data**

- **→ richer and deeper description**

- **A combination of qualitative and quantitative data often provides better understanding of the studied phenomenon ("mixed methods")**

Fixed and flexible processes in research

The research process may be characterized as fixed or flexible (Anastas and MacDonald 1994; Robson 2002)

Fixed design process

- all parameters are defined at the launch of the study

Flexible design process

- key parameters of the study may be changed during the course of the study

Case studies are typically flexible design studies

Experiments and surveys are fixed design studies

Triangulation

Increase the precision of empirical research

Triangulation means taking different angles towards the studied object and thus providing a broader picture

Qualitative data is broad and rich, but not precise

Quantitative data may also require qualitative observation, e.g. to modeling and measurement error evaluation

Types of triangulation (Stake 1995):
- Data (source) triangulation - using more than one data source or collecting the same data at different occasions
- Observer triangulation - using more than one observer in the study
- Methodological triangulation - combining different types of data collection methods, e.g. qualitative and quantitative methods
- Theory triangulation - using alternative theories or viewpoints

Characteristics of case study
(Yin 2003)

Characteristics of a case study (Yin 2003):

- Copes with the technically distinctive situation in which there will be many more variables than data points
- Relies on multiple sources of evidence in a triangulating fashion
- Benefits from the prior development of theoretical propositions to guide data collection and analysis
  - *(Note the difference between Yin and Grounded Theory)*

A case study will never provide conclusions with statistical significance

But: many different kinds of evidence, figures, statements, documents, are linked together to support a strong and relevant conclusion.

More characteristics (Perry & al. 2005; Runeson & Höst 2008)

Perry et al. has similar criteria for a case study (Perry et al. 2005). A case study :

- Has research questions set out from the beginning of the study
- Data is collected in a planned and consistent manner
- Inferences are made from the data to answer the research question
- Explores a phenomenon, or produces an explanation, description, or causal analysis of it
- Threats to validity are addressed in a systematic way."

Key characteristics of a case study (Runeson & Höst 2008):

- 1) it is of flexible type, coping with the complex and dynamic characteristics of real world phenomena, like software engineering
- 2) its conclusions are based on a clear chain of evidence, whether qualitative or quantitative, collected from multiple sources
- 3) it adds to existing knowledge by being based on previously established theory, if such exist, or by building theory

Case study process (Runeson & Höst 2008)

1. Case study design: objectives are defined and the case study is planned.

2. Preparation for data collection: procedures and protocols for data collection are defined.

3. Collecting evidence: execution with data collection on the studied case.

4. Analysis of collected data

5. Reporting

Case study design

- **A plan for a case study includes (Robson 2002):**
- Objective - what to achieve?
- The case - what is studied?
- Theory - frame of reference
- Research questions - what to know?
- Methods - how to collect data?
- Selection strategy - where to seek data?

Preparation for data collection

- **The case study protocol defines the field procedures for data collection**

- **The protocol is a continuously changed document that is updated when the plans for the case study are changed**

- **Includes also consideration of ethical factors such as**

- Informed consent

- Review board approval

- Confidentiality

- Handling of sensitive results

- Feedback

- **Subjects and organizations must explicitly agree to participate in the case study, i.e. give informed consent**

Collecting evidence

Data collection techniques can be divided into three levels (Lethbridge et al. 2005):

First degree

- Direct methods = the researcher is in direct contact with the subjects and collect data in real time. This includes interviews, focus groups and observations

Second degree

- Indirect methods = the researcher directly collects raw data without actually interacting with the subjects during the data collection
  - *for example remote observation through video*

Third degree

- Independent analysis of work artifacts = already available and sometimes compiled data is used
  - *archival data and documents, such as requirements specifications and failure reports or organizational databases*

Analysis of collected data

What is analysis in a case study?

Different approaches require different approaches

- Positivist – interpretive – critical
- It is possible to report a Grounded Theory or ethnography as a case study

Robson (2002) mentions the following approaches:

- Immersion approaches
  - *Very low level of structure, inductive, reliant on intuition and interpretive skills of the researcher*
- Editing approaches
  - *Few a priori codes. Most codes are defined inductively based on findings of the researcher during the analysis*
- Template approaches
  - *More formal a priori set of codes based on research questions*
- Quasi-statistical approaches
  - *Formalized approaches that include, for example, calculation of frequencies of codes, words and phrases*
  - *Epistemologically a quite problematic approach*

Reporting case studies

- **Robson's (2002) set of characteristics for a case study report:**
- tell what the study was about
- communicate a clear sense of the studied case
- provide a "history of the inquiry" so the reader can see what was done, by whom and how
- provide basic data in focused form, so the reader can make sure that the conclusions are reasonable
- articulate the researcher's conclusions and set them into a context they affect.

Experiments

What are experiments (Wohlin & Höst 2001)

- **In an experiment the researcher has an active control over the study**
- participants carry out the tasks that they are assigned to
- **In a typical case study the researcher is more of an observer**
- **An experiment can be planned and designed to ensure high validity**
- **Drawback: it is not possible to control a real life situation**
- **Experiments are done in labs and in other limited environments, such as in classrooms**
- **Experiments are often conducted for comparison (of techniques, methods, working procedures, etc.)**

Experiments are fixed designs

- **An experiment attempts to draw conclusions that are valid for a large population**

- **A sample of the entire population is used in the experiment as in most statistical methods**

- **A random sample from the population is ideal, but this is often impossible**

- **Randomization is an important principle**

- The decision of a treatment that every participant should be subject to is random

Variables in experiments

Typical variables of experiments include:

- **Independent variables**
  - *Describe the treatments in the experiment. For example in a tool comparison, the choice of tool is an independent variable*
- **Dependent variables**
  - *Studied to investigate whether they are influenced by the independent variables*
  - *The objective of the experiment is to determine if and how much the dependent variables are affected by the independent variables.*
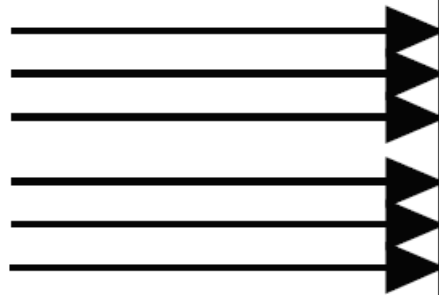
The independent and dependent variables are formulated to cover one or several hypotheses

Confounding factors are variables that may affect the dependent variables without the knowledge of the researcher

An objective of the design is to minimize the effect of confounding factors

Variables in experiments
(Wohlin & Höst 2001)

Fig. 1. Variables in an experiment.

Process of an experiment

Three key parts are (Wohlin & Höst 2001):

- **Commit participants**
  - *Participants must be committed to the tasks*
  - *Consideration of sensitive material (→ difficulties in getting commitment)*
- **Prepare instrumentation**
  - *All the material for the experiment must be prepared*
    - Instructions to the participants
    - Forms that should be used by the participants during the tests, etc.
  - *Data collection instruments (automatic, form based)*
  - *Analysis instruments for data analysis (statistics etc.)*
- **Execution**
  - *The part of the experiment where the participants, subject to their treatment, carry out the task that they are assigned to*
  - *Random allocation of treatments to participants*
  - *The participants use the prepared instrumentation to receive instructions and to record data that can be used later in analysis.*

Analyzing experi- mentation data

- The objective to decide whether there is an effect of the value of the independent variable(s) on the value of the dependent variable(s)

- Hypothesis testing

- The null hypothesis H0 denotes that there is no effect of the independent variable on the dependent variable

- The objective of the hypothesis test is to reject this hypothesis with a known significance

- As discussed in quantitative research lecture

Strenghts and weaknesses of experiments

- **Strengths**
- Easy to show causal relationships
  - *Independent variables (cause) → dependent variable (effect)*
- Simple experiments can be quite easily replicated

- **Weaknesses**
- Artificial situations
- Controls human behaviour
- Reduces social situation to a minimum
- Requires definition of variables before the experiment
  - *Are these really the important variables? How do we know that?*

Validity of empirical studies

Validity of empirical studies

Many ways to classify aspects of validity and threats to validity in the literature

Some criteria apply to positivistic case studies (Yin 2003, Runeson & Höst 2008), experiments and other quantitative studies

- Realist ontology – reality exists independent of observation
- Objective epistemology – truth is independent of the observer and it can eventually be found through reason

Others apply to interpretive and critical studies

- Relativist ontology – reality is constructed intersubjectively
- Transactional or subjectivist epistemology – we cannot separate ourselves from what we know

Positivist criteria of validity

Typically four aspects of the validity,

- Construct validity
- Internal validity
- External validity
- Reliability

E.g. Runeson & Höst 2008, Wohlin & Höst 2001

These criteria are required by many software engineering journals (even when they do not fully align with the research approach of the study)

The following is based on Runeson & Höst 2008

Construct validity

- To what extent the operational measures that are studied really represent what the researcher have in mind and what is investigated

- If, for example, the constructs discussed in the interview questions are not interpreted in the same way by the researcher and the interviewed persons, there is a threat to the construct validity

Internal validity

- **Internal validity is of concern when causal relations are examined**

- **the extent to which a causal conclusion based on a study is warranted**

- **When one factor affects an investigated factor there is a risk that the investigated factor is also affected by a third factor**

- **If the researcher is not aware of the third factor and its effect, there is a threat to the internal validity**

External validity

- **To what extent it is possible to generalize the findings**

- **To what extent the findings are of interest to other people outside the investigated case**

- **For case studies, the intention is to enable analytical generalization**

- the results can be extended to cases with common characteristics and hence for which the findings are relevant

- Defining a theory as the objective in case studies

Reliability

- To what extent the data and the analysis are dependent on the specific researchers

- Hypothetically, if another researcher later on conducted the same study, the result should be the same

- Clarity of data collection and analysis methods important

Criteria for interpretive studies
(Lincoln & Cuba, 1985)

Lincoln and Guba's Evaluative Criteria of trustworthiness include

- **Credibility** - confidence in the 'truth' of the findings
- **Transferability** - showing that the findings have applicability in other contexts
- **Dependability** - showing that the findings are consistent and could be repeated
- **Confirmability** - a degree of neutrality or the extent to which the findings of a study are shaped by the respondents and not researcher bias, motivation, or interest.

Techniques for establishing credibility

Prolonged Engagement
- Spending sufficient time in the field to learn or understand the culture, social setting, or phenomenon of interest

Persistent Observation
- To identify those characteristics and elements in the situation that are most relevant and focusing on them in detail and deepness

Triangulation
- using multiple data sources, methods, etc. in an investigation

Peer debriefing
- exposing oneself to peers in analytical sessions for the purpose of exploring aspects of the inquiry

Negative case analysis
- searching for and discussing elements of the data that do not support or appear to contradict to those that are emerging from data analysis

Referential adequacy
- identifying a portion of data to be archived, but not analyzed – used later for testing the validity

Member-checking
- data, analytic categories, interpretations and conclusions are tested with members of those groups from whom the data were originally obtained

Techniques for establishing transferability

Thick description

- By describing a phenomenon in sufficient detail one can begin to evaluate the extent to which the conclusions drawn are transferable to other times, settings, situations, and people

Techniques for establishing dependability

Inquiry audit

- A researcher not involved in the research process examine both the process and product of the research study
- The purpose is to evaluate the accuracy and evaluate whether or not the findings, interpretations and conclusions are supported by the data

Also confirmability audit for establishing confirmability

Techniques for establishing confirmability

Audit trail

- a transparent description of the research steps taken from the start of a research project to the development and reporting of findings

Triangulation

- See earlier: using multiple data sources, methods, etc. in an investigation

Reflexivity

- an attitude of attending systematically to the context of knowledge construction, especially to the effect of the researcher, at every step of the research process

Thank you!

Questions? Comments?