SciPub course, seminar 4, Research Methods & Threats

20231117, NTNU, Trondheim
Robert Feldt

CHALMERS

Which research method(s) to use in which situation?

There is no such thing as a universal way of scientific practice

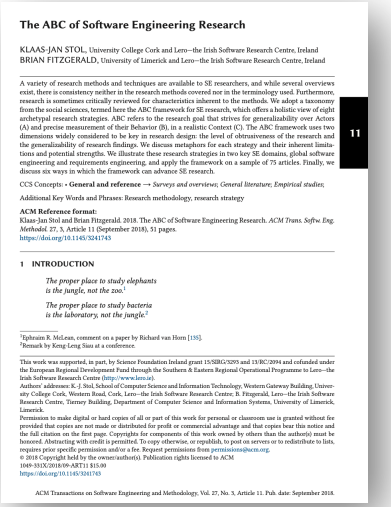
Method selection depends on many non-trivial questions

- What is the purpose of the study?
  Exploratory? Descriptive? Explanatory? Improving?
- What is the nature of the study?
  Inductive? Deductive?
- What is the relation to existing evidence?
  Building a new theory? "Testing" existing theory?
- What is the nature of the questions we ask?
  What-questions? Why-questions?
- What is the nature of the environment?
  Controlled environments? Realistic environments?
- What is the necessary sample?
  Population source?
  Units of analysis?

Criteria for selecting methods

Criteria for environment selection (and sampling)

Not trivial, but possible: checklists



The ABC of Software Engineering Research

KLAAS-JAN STOL, University College Cork and Lero—the Irish Software Research Centre, Ireland
BRIAN FITZGERALD, University of Limerick and Lero—the Irish Software Research Centre, Ireland

A variety of research methods and techniques are available to SE researchers, and while several overviews exist, there is consistency neither in the research methods covered nor in the terminology used. Furthermore, research is sometimes critically reviewed for characteristics inherent to the methods. We adopt a taxonomy from the social sciences, termed here the ABC framework for SE research, which offers a holistic view of eight archetypal research strategies. ABC refers to the research goal that strives for generalizability over Actors (A) and precise measurement of their Behavior (B), in a realistic Context (C). The ABC framework uses two dimensions widely considered to be key in research design: the level of obtrusiveness of the research and the generalizability of research findings. We discuss metaphors for each strategy and their inherent limitations and potential strengths. We illustrate these research strategies in two key SE domains, global software engineering and requirements engineering, and apply the framework on a sample of 75 articles. Finally, we discuss six ways in which the framework can advance SE research.

11

CCS Concepts: • **General and reference** → *Surveys and overviews; General literature; Empirical studies;*

Additional Key Words and Phrases: Research methodology, research strategy

ACM Reference format:
Klaas-Jan Stol and Brian Fitzgerald. 2018. The ABC of Software Engineering Research. *ACM Trans. Softw. Eng. Methodol.* 27, 3, Article 11 (September 2018), 51 pages.
https://doi.org/10.1145/3241743

1 INTRODUCTION

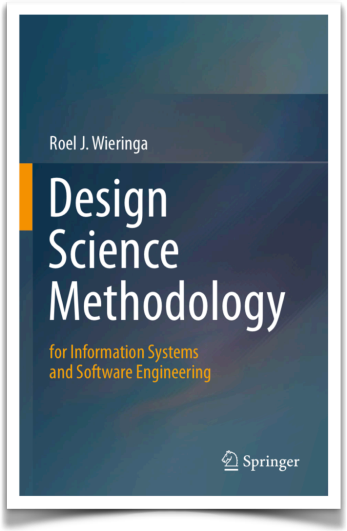
The proper place to study elephants is the jungle, not the zoo.[1]

The proper place to study bacteria is the laboratory, not the jungle.[2]

[1] Ephraim R. McLean, comment on a paper by Richard van Horn [135].
[2] Remark by Keng-Leng Siau at a conference.

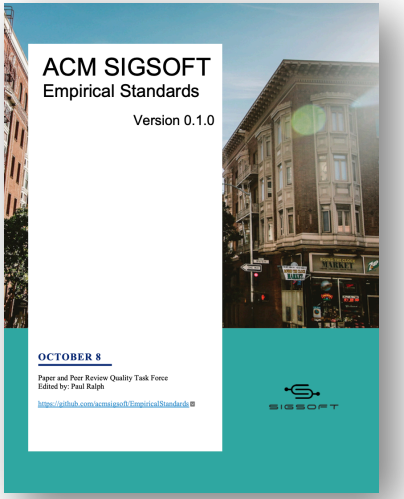
This work was supported, in part, by Science Foundation Ireland grant 15/SIRG/3293 and 13/RC/2094 and cofunded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero—the Irish Software Research Centre (http://www.lero.ie).
Authors' addresses: K.-J. Stol, School of Computer Science and Information Technology, Western Gateway Building, University College Cork, Western Road, Cork, Lero—the Irish Software Research Centre; B. Fitzgerald, Lero—the Irish Software Research Centre, Tierney Building, Department of Computer Science and Information Systems, University of Limerick, Limerick.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
1049-331X/2018/09-ART11 $15.00
https://doi.org/10.1145/3241743

Good starting point on positioning of research methods

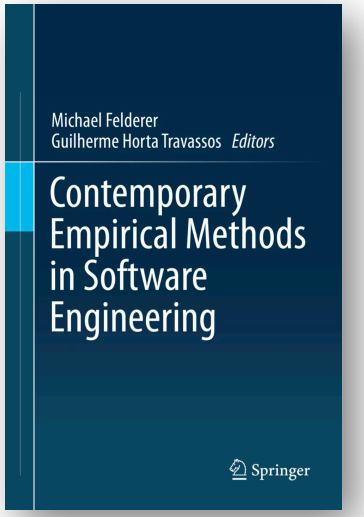


More advanced

- Chapter 16 + Appendix
- http://bit.ly/checklists-design_science



Overview of quality criteria as per recommendation by research community:



Deep dives into selected research methods

Basic types of methods

IN REVIEW An Introduction to Research Methods

Goal Met	Research Methods	Advantages/Disadvantages
Description	Observational method	Descriptive methods allow description of behavior(s)
	Case study method	Descriptive methods do not support reliable predictions
	Survey Method	Descriptive methods do not support cause-and-effect explanations
Prediction	Correlational method	Predictive methods allow description of behavior(s)
	Quasi-experimental method	Predictive methods support reliable predictions of one variable based on another
		Predictive methods do not support cause-and-effect explanations
Explanation	Experimental method	Allows description of behavior(s)
		Supports reliable predictions of one variable based on another
		Supports cause-and-effect explanations

[Jackson2010]

The ABC of Software Engineering Research

KLAAS-JAN STOL, University College Cork and Lero—the Irish Software Research Centre, Ireland
BRIAN FITZGERALD, University of Limerick and Lero—the Irish Software Research Centre, Ireland

Table 2. Examples of Actors, Behavior, and Context in Software Engineering Research

Actors	Managers, software engineers, users, software systems, software development artifacts incl. defects, tools, techniques, prototypes
Behavior	System behavior (e.g., reliability, performance, and other quality attributes), software engineers' behavior and antecedents such as productivity, motivation, and intention
Context	Industrial settings, organizations, software projects, development teams, software laboratory, classroom, meeting rooms

Table 3. Comparison of Knowledge-Seeking versus Solution-Seeking Research in Software Engineering Research

	Knowledge-Seeking Research	Solution-Seeking Research
Goal	To generate or propose scientific claims and to evaluate and validate those claims. This may also include the development of instruments or other artifacts with the specific purpose of supporting or enabling these knowledge-seeking activities, e.g., the construction of an instrument that facilitates data gathering or analysis.	To design or develop new or improve existing solutions that can help to overcome or ameliorate challenges, bottlenecks, and other problems in the development of software systems and supporting processes.
Focus of research	A phenomenon within software engineering, or a characteristic thereof, that is not (or not sufficiently) well understood or known.	A specific software engineering challenge, obstacle, or problem, and the design or creation of a solution.
Outcome	Empirical findings, descriptions, insights generated by simulations, theoretical or conceptual frameworks, or hypotheses.	Artifacts that include algorithms, tools, notations (incl. languages), models, mechanisms, and techniques.
Example	Research question: How does Extreme Programming work? Sharp et al. conducted an ethnographic study to capture how developers employ XP development practices [180].	Research question: how to select a manageable subset of the input data faster in order to automatically find performance bottlenecks? Luo et al. proposed FOREPOST: an adaptive, feedback-directed learning testing system that helps to find performance bottlenecks [124]. (Note that the *evaluation* of FOREPOST is a knowledge-seeking study.)

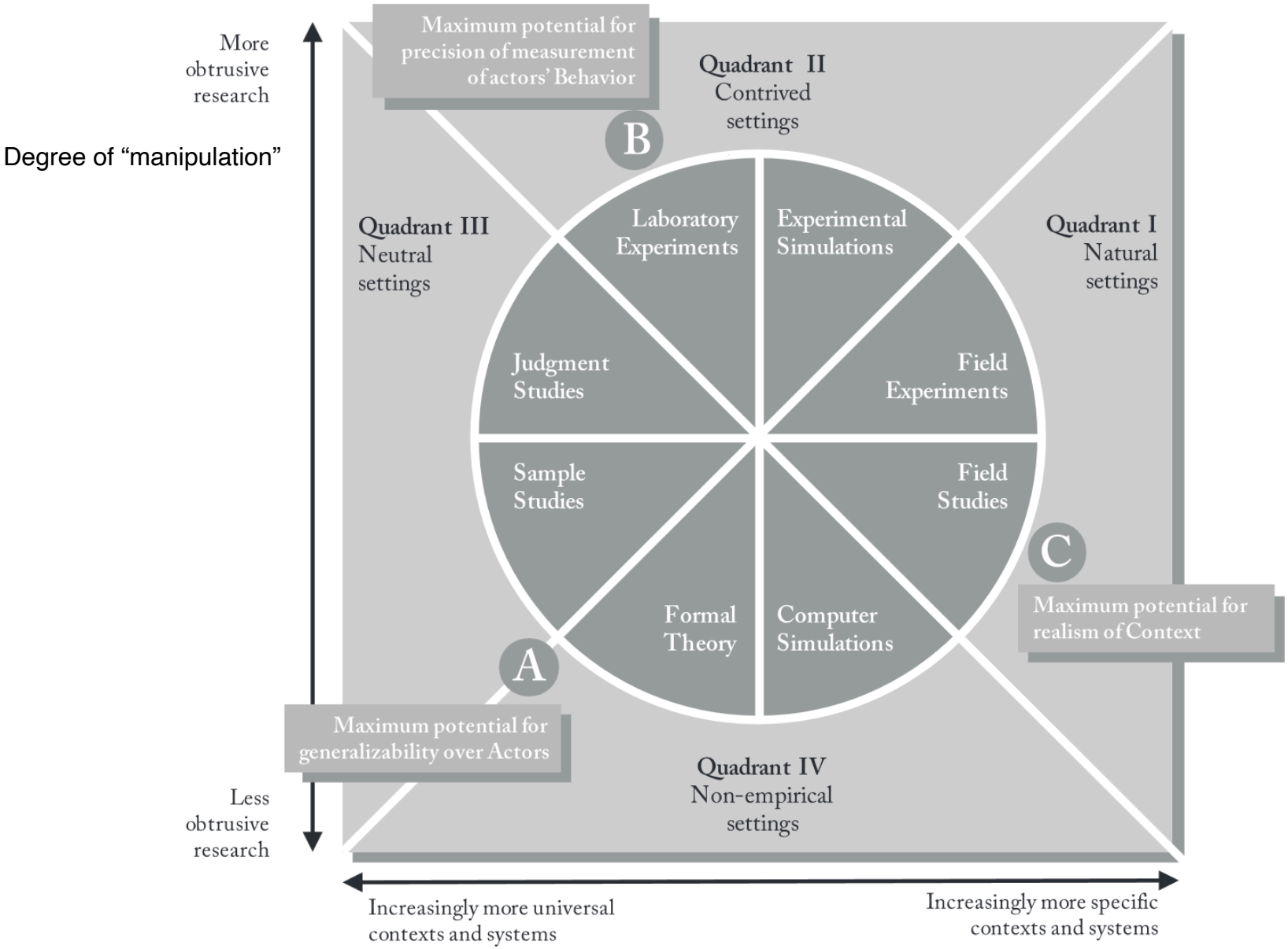


Fig. 1. The ABC framework: eight research strategies as categories of research methods for software engineering (adapted from Runkel and McGrath [170]). Quadrants I to IV represent different research settings.

Study Type	Metaphor and Setting	Purpose	Typical Methods & Data	Inherent Limitations
Field Study	*Jungle*: Natural setting that exists before the researcher enters it. Minimal intrusion of the setting so as not to disturb realism, only to facilitate data collection.	Facilitates study of phenomena and actors and their behavior in natural contexts. Exploratory, to understand *what's going on, how things work,* or to generate hypotheses.	Case study, ethnography, observational study; qualitative data incl. interviews, field notes, archival documents, may include quantitative data.	• No statistical generalizability • No control over events • Low precision of measurement
Field Experiment	*Nature reserve*: Natural, pre-existing setting (in vivo), but some level of intrusion due to the deliberate manipulation of aspects of the setting; study affected by confounding factors.	To investigate, evaluate, or compare techniques, practices, processes, or approaches within a real-world and pre-existing setting.	Evaluative case study, quasi-experiment, Action Research; studies may use either quantitative data or qualitative data.	• No statistical generalizability • Precision of measurement affected by confounding contextual factors
Experimental Simulation	*Greenhouse, flight simulator*: Contrived setting (in virtuo) created specifically for a study to represent a concrete type of setting. Environment is created by the researcher to study behavior of actors.	To study behavior of participants or systems in a controlled setting that resembles a real-world, concrete class of settings as closely as possible.	Simulation games, management games, instrumented multiplayer games; quantitative or qualitative data, depending on the simulation instrument.	• Generalizability reduced as setting is designed to mirror a specific type of setting • Realism reduced due to artificial setting
Laboratory Experiment	*Cleanroom, test tube*: Contrived setting (in vitro) created specifically for a study, with high degree of control of all measured variables.	To study with a high degree of precision relationships between variables, or comparisons between techniques; may allow establishment of causality between variables.	Randomized controlled experiments and quasi experiments, comparative evaluations with benchmark studies; usually quantitative data exclusively.	• Abstract or unrealistic context due to highly artificial setting • Typically scope of problem reduced to study the "essence," optimizing internal validity at cost of external validity

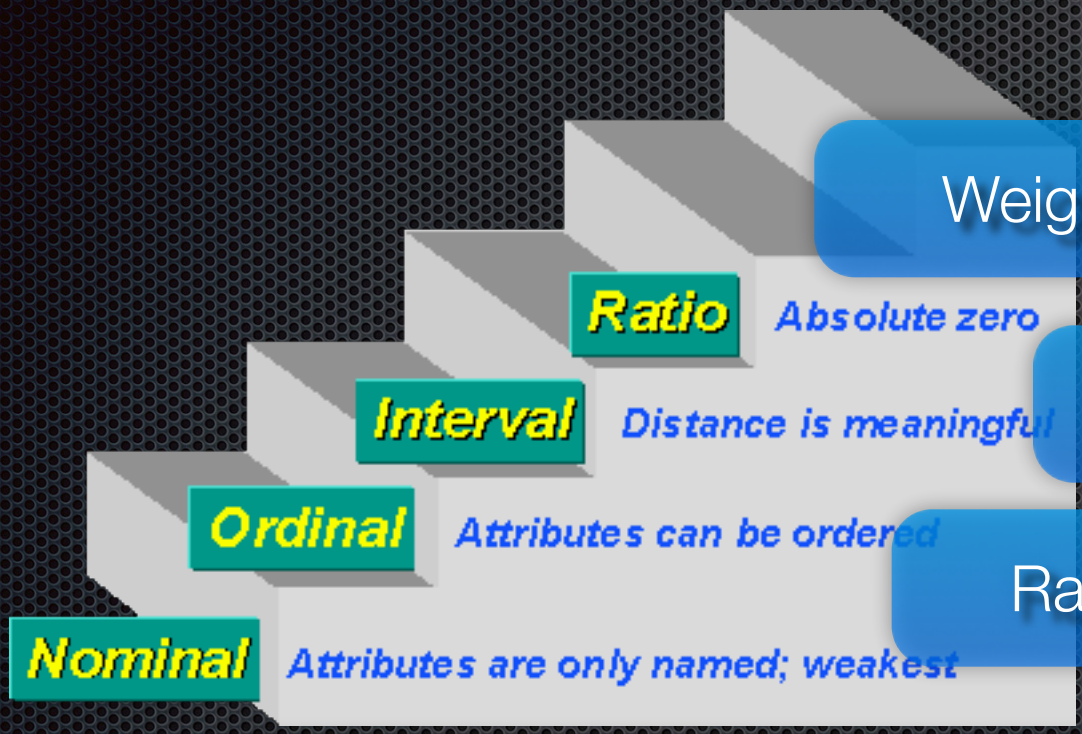
Study Type	Metaphor and Setting	Purpose	Typical Methods & Data	Inherent Limitations
Judgment Study	*Courtroom*: Neutral setting; may be actively designed to nullify the context, so that "responses" are in relation to some stimulus (question or instructions), independent of setting.	To elicit information from subjects for purposes of evaluation or study of an object. To seek generalizability of responses to stimuli, not generalizability to a population.	Delphi studies, interview studies, focus group, evaluation studies; use of qualitative and/or quantitative data.	• Responses not related to any specific or realistic context • Less generalizability than sample studies due to lack of *representative* sampling • Less control and precision of measurement than a lab. exp.
Sample Studies	*Referendum*: Neutral setting. Limited level of precision of measurement; no variables are manipulated. The researcher must deal with whatever data is collected.	To study the distribution of a particular characteristic in a population (of people or systems), or the correlation between two or more characteristics in a population. Information is sought of the subjects.	Software repository mining, questionnaires, interviews; analysis includes correlational methods, e.g., regression. Typically, quantitative data (e.g., Likert scales) but can include qualitative data.	• Reductionist—depth of and number of data points per participant limited • Data collection not "interactive": no option to clarify questions; repository data comes *as is*, no opportunity to *manipulate* variables, only to *correlate* them
Formal Theory	*Jigsaw puzzle*: Nonempirical setting; typically a research office or library.	To develop a conceptualization, framework, or theory on a topic. Focus is on formulating relations among concepts, or explanations that hold for a wide range of contexts.	Conceptual reasoning, concept development, development of propositions and/or hypotheses; framework development.	• Low on realism: does not consider a specific context but rather abstract concepts • No manipulation of variables or measurement (no empirical information is gathered)
Computer Simulation	*Forecasting system*: Nonempirical setting (in silico); no recording of observations in the real world. There are no actors (people, real-world systems) or real-world behavior: everything is specified in the simulation.	To model a particular system or phenomenon that facilitates evaluation of a large number of complex scenarios that are captured in the preprogrammed model.	Development of software programs that contain symbolic representations of all variables a researcher considers important; usually these variables are derived and calibrated based on prior empirical studies.	• No manipulation of variables or precision of measurement (no empirical data is gathered) • Results will be as good as the accuracy of the model representing the simulated system • Low generalizability as it attempts to model a specific class of real-world systems

Table 7. Examples of Different Research Strategies Used in Requirements Engineering Research

Strategy	Authors	Objective	Study Setting	Study Procedure	Findings
Field Study	Damian and Zowghi [37]	To investigate the impact of distributed stakeholders on RE activities in GSD.	*Natural:* first author spent 7 months on-site at an organization.	Document study, observation, interviews.	Four major problems and 8 specific challenges related to requirement engineering activities in distributed development.
Field Experiment	Lauesen and Vinter [111]	To identify a cost-effective way to avoid requirement defects.	*Natural:* company staff and researcher collaborated on-site, using real products to evaluate new approaches.	Action research, data include defect reports, time spent, usability issues, timeliness of the project, product sales.	Several conclusions, incl.: scenarios and early usability testing are beneficial techniques; classifying defects according to the source of the defect was not helpful.
Experimental Simulation	Lerch et al. [116]	To identify the computer support needs of automation staff in a large organization.	*Contrived:* simulation environment with experimental stimuli that were previously collected.	3 distinct groups of 3 people each to gauge how level of expertise and circumstances affect behavior. Participants received training.	Insights into different information needs and search strategies and decision-making strategies depending on users' expertise. Insights on performance with feedback/feedforward.
Laboratory Experiment	Porter et al. [159]	To investigate the hypothesis that scenario-based inspections are more effective than ad hoc inspections.	*Contrived:* classroom laboratory exercise with graduate students doing a course in formal methods.	Measurement of effect of detection methods (ad hoc, checklist, scenario) on 4 dependent variables incl. fault detection rate.	4 key findings, incl.: scenario-based method leads to higher fault detection rate than other methods; scenario reviewers were more effective at detecting faults the scenarios were designed to uncover.

Judgment Study	Daneva and Ahituv [39]	To evaluate a set of 12 practices based on feedback by ERP practitioners.	*Neutral:* dedicated meeting room with seating around a table.	10 consultants from 7 ERP services firms, selected based on their interest and expertise.	All 12 practices were observed by several of the panel experts.
Sample Study	Neill and Laplante [146]	To investigate the state of practice of requirements engineering in industry.	*Neutral:* web-based questionnaire.	22 questions; participants drawn from a Penn State School of Graduate Professional Studies database; 194 responses from a population of 1,519.	Findings include organization and participant characteristics (various domains; participants held variety of positions); software development life cycle model (agile, waterfall, etc.); RE techniques.
Formal Theory	Nguyen and Shanks [147]	To develop an understanding of the role of creativity in RE.	*Nonempirical:* no empirical observations, but derivation of a conceptual framework from literature.	General creativity literature, requirements engineering creativity literature.	A theoretical framework that offers RE researchers a basis to incorporate creativity in RE methods and techniques.
Computer Simulation	Höst et al. [77]	To investigate bottlenecks and overload in RE processes.	*Nonempirical:* a discrete event simulator was implemented in SDL.	Four simulation scenarios with different parameter values to model different circumstances.	Two ways were identified to avoid congestion: (1) increase number of staff or productivity, (2) decrease the rate of new requirements e.g., through prioritization.

Scales of measurement



Weight, Time, Height

Fahrenheight degrees

Ranking of individuals

Gender, Ethnicity, Company

[www.socialresearchmethods.net 2013]

Empirical Software Engineering manuscript No.
(will be inserted by the editor)

Sampling in Software Engineering Research: A Critical Review and Guidelines

Sebastian Baltes · Paul Ralph

Purpose: *(1) Providing a detailed, SE-specific primer on sampling concerns and techniques; (2) investigating the state of sampling in SE research; (3) providing guidelines for improving sampling in SE research.*

Received: date / Accepted: date

Abstract Representative sampling appears rare in empirical software engineering research. Not all studies need representative samples, but a general lack of representative sampling undermines a scientific field. This article therefore reports a critical review of the state of sampling in recent, high-quality software engineering research. The key findings are: (1) random sampling is rare; (2) sophisticated sampling strategies are very rare; (3) sampling, representativeness and randomness often appear misunderstood. These findings suggest that *software engineering research has a generalizability crisis.* To address these problems, this paper synthesizes existing knowledge of sampling into a succinct primer and proposes extensive guidelines for improving the conduct, presentation and evaluation of sampling in software engineering research. It is further recommended that while researchers should strive for more representative samples, disparaging non-probability sampling is generally capricious and particularly misguided for predominately qualitative research.

Table 2 Reasonable Arguments regarding Representativeness

Name	GPRS*	Argument	Threats
Random Selection	No	Each item in sampling frame has an equal probability of inclusion.	Assumes large sample size, good sampling frame and no coincidences.
Size	No	Sample is so large that missing a subpopulation is unlikely.	Bias toward one or more subpopulations or on important parameters.
Breadth	No	Sample captures a large variance on important parameters.	Point estimates unreliable; groups may not be proportional.
Parameter Matching	No	Sample and population parameters have similar distributions.	Possible bias outside of considered parameters; requires known population parameters.
Universality	No	All possible samples are representative because the phenomenon affects the entire population equally.	Phenomenon is not actually universal.
Postmodern Critique	No	The entire logic of statistically generalizing from a sample to a population is flawed.	Statistical generalization not supported.
Practical Critique	No	Generalizing to a population is not the purpose of this kind of study (e.g. case study, experiment).	Statistical generalization not supported.

*Guaranteed to Produce a Representative Sample?

1. **Identify the population;** that is, in principle, who or what you would like to generalize to (e.g. professional software developers in Turkey, code faults in cyber-physical systems).
2. **Present your sampling algorithm** (see Section 2.5). If the algorithm maps neatly into the tripartite model of a population, a sampling frame and a sample, feel free to use those terms. However, if the algorithm has many stages and does not map neatly into the tripartite model, just focus on providing a replicable, concise, algorithmic account of how another person can generate the same sample. It is not necessary to name every stage or identify one specific stage as the "sampling frame."
3. If you use data collection scripts, provide them.
4. If the sampling strategy has many stages, consider a diagram or flow chart.
5. If your sampling approach was nondeterministic (e.g. theoretical sampling), describe your reasoning.
6. Give an **explicit argument for representativeness** (cf. Table 2). Admit the generalizability threats implied by this argument.
7. **Clearly explain how the sample could be biased.** For complicated sampling strategies, discuss bias for each step in your algorithm or diagram explaining the sampling approach. This could be presented in the sampling section or with limitations.
8. **Publish your sample** as part of a replication package **if and only if** it does not contain sensitive or protected information. Be very careful of the potential for re-identifying de-identified data.

Your research methods?

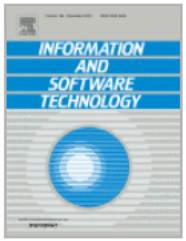
Let's discuss!

- PQ4.1. Which research methods have you used in your research so far? Which ones are you planning to use? Why were these methods chosen?
- PQ4.2. In the studies you have done so far or are planning to do in the near-term, what are the top-3 threats to the validity of your results? What did you do or can you do to mitigate them (in a better way) going forward?

ELSEVIER

Information and Software Technology

Volume 164, December 2023, 107329



Threats to validity in software engineering research: A critical reflection

Roberto Verdecchia [a], Emelie Engström [b], Patricia Lago [c], Per Runeson [b], Qunying Song [b]

Worldviews, Research Methods, and their Relationship to Validity in Empirical Software Engineering Research

Kai Petersen
School of Computing
Blekinge Institute of Technology
Karlskrona, Sweden
Email: kai.petersen@bth.se

Cigdem Gencel
Facult of Computer Science
Free University of Bolzano/Bozen
Bolzano/Bozen, Italy
Email: cigdem.gencel@unibz.it

TABLE I. RESEARCH METHODS AND NATURE OF DATA IN DIFFERENT WORLDVIEWS

World views	Truth	Research method	Data collection
Positivist/ Postpositivist	Objective (independent from participants)	Controlled experiment, case study, and survey	Quantitative (Random sampling)
Constructivist/ Interpretivist	Subjective (constructed by participants)	Case study, survey, interview	Quantitative and Qualitative (Random and purposeful)
Advocacy/ Participatory	Subjective (constructed by participants and the observer)	Action research (intervention driven)	Quantitative and Qualitative (Random and purposeful)
Pragmatist	Depends on what works at the time	Multi method research	Quantitative and Qualitative (Random and purposeful)

World Views

	Positivist	Interpretivist	Participatory (Critical Theory)	Pragmatist
Data Collection (record, store, review, and revise)	**Theoretical Validity** - Reliability of measures - Reliability of treatment implementation - Heterogeneity of subjects - Constructs (e.g. theory) not well defined, unclear - Mono-method bias (use single measure) - Evaluation apprehension - Selection of subjects - Ambiguity about direction of causal influence - History affects results - Maturation (behavior change over time) - Poor instrumentation for data collection - Compensatory rivalry between subjects with different treatments - Researcher expectations **Generalizability** - Interaction of selection and treatment - Interaction of setting and treatment - Interaction of history and treatment	**Theoretical Validity** - Constructs (e.g. theory, object of study) not well defined, unclear - Mono-method bias - Evaluation apprehension - Selection of subjects - Ambiguity about direction of inferences derived from data - History affects results - Maturation (behavior change over time) - Poor instrumentation for data collection **Generalizability** - Saturation, i.e. non-purposeful selection study context/questions **Descriptive Validity** - poor recording of data - too many steps of interpretation	**Theoretical Validity** - Constructs (e.g. theory, object of study) not well defined, unclear - Mono-method bias - Descriptive accuracy - Selection of subjects - Ambiguity about direction of inferences derived from data - History affects results - Maturation (behavior change over time) - Learning effects - Poor instrumentation for data collection - political and social threats - Resistance of change - few iterations **Generalizability** - Saturation, i.e. non-purposeful selection study context/questions **Descriptive Validity** - poor recording of data - too many steps of interpretation	Combinations of the views, depending on how research stances are combined (multi-method)
Data Analysis (interpret, report, review, and revise)	**Interpretative Validity** - Low stat. power - Invalid assumptions of stat. tests	**Interpretative Validity** - Researcher bias in drawing conclusions **Generalizability** - Lack of context definition and awareness in interpretation	**Interpretative Validity** - Researcher bias in drawing conclusions - Researcher suffers from organizational/contextual blindness **Generalizability** - Lack of context definition and awareness in interpretaiton	Combinations of the views, depending on how research stances are combined (multi-method)

Feedback, criticism/praise, ideas for improvement,
Please email:

robert.feldt@chalmers.se

Thank you!