Preliminary Guidelines for Empirical Research in Software Engineering

Barbara A. Kitchenham et al
IEEE TSE Aug 02

The Problem

“In our view, the standard of empirical software engineering research is poor. This includes case studies, surveys, and formal experiments, whether observed in the field or in a laboratory or classroom. This statement is not a criticism of software researchers in particular; many applied disciplines have problems performing empirical studies.”

Questions

◆ What are the backgrounds of the authors?
◆ With what field did they compare?

The guidelines

◆ Experimental context
◆ Experimental design
◆ Conduct and data collection
◆ Analysis
◆ Presentation of results
◆ Interpretation of results

Goals of Exp Context

The main goals of context guidelines are:
1. To ensure that the objectives of the research are properly defined
2. To ensure that the description of the research provides enough detail for other researchers and for practitioners.

Exp Context Guidelines

C1: Be sure to specify as much of the industrial context as possible. In particular, clearly define the entities, attributes, and measures that are capturing the contextual information.

C2: If a specific hypothesis is being tested, state it clearly prior to performing the study and discuss the theory from which it is derived, so that its implications are apparent.
Exp Context Guidelines

C3: If the research is exploratory, state clearly and, prior to data analysis, what questions the investigation is intended to address and how it will address them.

C4: Describe research that is similar to, or has a bearing on, the current research and how current work relates to it.

Exp Design

The study design describes the products, resources and processes involved in the study, including:
- the population being studied,
- the rationale and technique for sampling from that population,
- the process for allocating and administering the treatments (the term "intervention" is often used as an alternative to treatment), and
- the methods used to reduce bias and determine sample size.

Exp Design Guidelines

D1: Identify the population from which the subjects and objects are drawn.

D2: Define the process by which the subjects and objects were selected.

D3: Define the process by which subjects and objects are assigned to treatments.

D4: Restrict yourself to simple study designs or, at least, to designs that are fully analyzed in the statistical literature. If you are not using a well-documented design and analysis method, you should consult a statistician to see whether yours is the most effective design for what you want to accomplish.

D5: Define the experimental unit.

D6: For formal experiments, perform a pre-experiment or precalculation to identify or estimate the minimum required sample size.

D7: Use appropriate levels of blinding.

D8: If you cannot avoid evaluating your own work, then make explicit any vested interests (including your sources of support) and report what you have done to minimize bias.

D9: Avoid the use of controls unless you are sure the control situation can be unambiguously defined.

D10: Fully define all treatments (interventions).

D11: Justify the choice of outcome measures in terms of their relevance to the objectives of the empirical study.
Conduct and collect

DC1: Define all software measures fully, including the entity, attribute, unit and counting rules.

DC2: For subjective measures, present a measure of interrater agreement, such as the kappa statistic or the intraclass correlation coefficient for continuous measures.

DC3: Describe any quality control method used to ensure completeness and accuracy of data collection.

DC4: For surveys, monitor and report the response rate and discuss the representativeness of the responses and the impact of nonresponse.

DC5: For observational studies and experiments, record data about subjects who drop out from the studies.

DC6: For observational studies and experiments, record data about other performance measures that may be affected by the treatment, even if they are not the main focus of the study.

Analysis Approaches

1. Classical analysis (often referred to as the “frequentist” approach). This approach is adopted by most statistical packages.

2. Bayesian analysis. This approach provides a systematic means of making use of prior information. Prior information may be obtained from previous studies of the phenomenon of interest or from expert opinion.

Analysis

A1: Specify any procedures used to control for multiple testing.

A2: Consider using blind analysis.

A3: Perform sensitivity analyses.

A4: Ensure that the data do not violate the assumptions of the tests used on them.

A5: Apply appropriate quality control procedures to verify your results.

A3: Perform sensitivity analyses.

Presentation

P1: Describe or cite a reference for all statistical procedures used.

P2: Report the statistical package used.

P3: Present quantitative results as well as significance levels. Quantitative results should show the magnitude of effects and the confidence limits.

P4: Present the raw data whenever possible. Otherwise, confirm that they are available for confidential review by the reviewers and independent auditors.

P5: Provide appropriate descriptive statistics.

P6: Make appropriate use of graphics.
Interpretation

11: Define the population to which inferential statistics and predictive models apply.

12: Differentiate between statistical significance and practical importance.

13: Define the type of study.

14: Specify any limitations of the study.

For L6 (Thurs 9/12)

* Laitenberger, etal