

Does Subject Type Influence Software Engineering Experiment Results?

Sira Vegas¹, Patricia Riofrío¹, Natalia Juristo^{1,2}

¹ Escuela Técnica Superior de Ingenieros Informáticos, Universidad Politécnica de Madrid, Madrid, Spain

² Department of Information Processing Science, University of Oulu, Oulu, Finland

{svegas, priofrío, natalia}@fi.upm.es

Abstract— Context: A key issue when dealing with the generalization threat of software engineering experiments is to use different subject types. **Objective:** In this paper, we aim to investigate which subject types are used in experiments and their impact on results. **Method:** We have performed a systematic mapping study by manually searching experiments published from January 2014 to June 2016 in six leading software engineering conferences and journals. **Results:** Out of the 833 papers published in the period covered, we have identified 93 papers reporting experiments with subjects. Of these, 27 papers report experiments that have two subject types (professionals and students). We have studied the impact of subject type on the results of experiments reported in 11 of these papers. **Conclusion:** We have observed contradictory results. Only in some cases subject type influences experimental results. This suggests that further research is needed in order to find an explanation.

Keywords- experimentation; experiments with subjects; experiments with professionals.

I. INTRODUCTION

Simplifications of reality are a must in laboratory experiments. A very common one is to use students instead of professional developers in software engineering (SE) laboratory experiments. Several authors have reached similar findings with respect to the usage rates of students and professionals in SE experiments. Höfer and Tichy [1] report that 60% of experimental studies in SE use students, whereas 22% employ professionals and only 14% use both subject groups. Sjøberg et al. [4] report that, out of 103 papers reporting controlled experiments conducted between 1993 and 2002, 87% use students and only 9% employ professionals. Besides, undergraduates are used more often than graduate students. But experiments run with students suffer from the *interaction of causal relationship with units external validity threat*¹ [3].

Despite the importance of thoroughly investigating the impact of the use of students in SE experiments, no research has, to the best of our knowledge, been undertaken to date.

We have identified a need to synthesize the evidence about the impact of using professionals or students in SE experiments by thoroughly examining their results. For this purpose, we have reviewed the literature to study the impact of using professionals versus students on the results of

published SE experiments. We have conducted a systematic mapping study in conformance with [2].

The paper is organized as follows. Section II details the research method that we have followed. Section III answers our research questions. Section IV outlines the conclusions.

II. RESEARCH METHOD

Our systematic mapping study looks for answers to the following questions:

- RQ1: What types of subjects are used in SE experiments?
- RQ2: How are subject types incorporated into the experiment analysis?
- RQ3: What is the impact of using subject types on experimental results?

A. Search and Selection Process

We conducted a manual search of all the papers published from January 2011 to June 2016 in leading SE journals and conferences, namely, IEEE Transactions on Software Engineering (TSE), the Empirical Software Engineering Journal (EMSE), ACM Transactions on Software Engineering and Methodology (TOSEM), the International Conference on Software Engineering (ICSE), the Foundations on Software Engineering (FSE), and the International Symposium on Empirical Software Engineering and Measurement (ESEM)².

We skimmed through a total of 833 published articles in search of papers reporting experiments satisfying the following inclusion/exclusion criteria:

- a) Papers reported controlled experiments or quasi-experiments.
- b) Experiments compared at least two treatments.
- c) Experiment used subjects to apply treatments.

We identified 93 papers reporting experiments with subjects.

B. Data Extraction

Data were extracted as part of an iterative three-round process:

In Round 1, we collected data on the type of subject used in each experiment reported in the 93 identified papers. We then selected papers reporting: (1) experiments in which both

¹ An effect identified with one kind of unit might not hold for studies with another kind of unit [3].

² For the year 2016, the search includes six issues of TSE, three issues of EMSE, one issue of TOSEM and ICSE papers. Note that papers from ESEM 2016 and FSE 2016 had not been published at the time of writing.

subject types (professionals and students) participate, or (2) replications of a single experiment accounting for different subject types. Additionally, we removed papers reporting the same experiment (only one duplicate was found)³. We were left with 27 papers.

In Round 2, we examined the remaining 27 papers (hereinafter primary studies) and gathered data on the characteristics of the subjects and how they were incorporated into the experiment. We then selected the papers that provided information regarding the influence of subject type on the response variable under study. We were left with 11 primary studies.

In Round 3, we gathered information from the 11 selected primary studies to study the impact of subject type on the results.

C. Analysis

The response to RQ1 is based on the 93 selected papers. The response to RQ2 is based on the 27 primary studies -out of the 93 selected papers- that use students and professionals as experimental subjects. The response to RQ4 is based on the last 11 primary studies.

III. RESULTS

A. RQ1: Type of Subjects Used

Out of the 93 selected papers, 50 use students only, 16 use professionals only, and 27 use both. This means that 77 papers use students and 43 professionals. The numbers that we have found for experiments run with students only are consistent with the figures reported by Höfer and Tichy in 2007 [1]. The numbers for experiments run with professionals only or both subject types are the inverse with respect to the figures given by Höfer and Tichy.

B. RQ2: Incorporation of Subject Type into the Experiment Analysis

Out of the 27 primary studies, there are some cases in which we could not study the impact of subject type:

- In about one third of the primary studies (10 out of 27), the data corresponding to all subjects are analysed together, and they make no distinction between subject types.
- There are another two primary studies for which the impact of subject type cannot be assessed. One of studies does not perform statistical analyses, whereas one of the treatments is confounded with subject type in the other.
- There are four primary studies in which the data corresponding to all subjects are analysed together, where subject characteristics instead of subject type are incorporated as a factor in the analysis of the experiment.

In the remaining 11 primary studies, we have found different ways of incorporating subject type into the analysis

of the experiments (note that some primary studies perform more than one type of analysis):

- Five studies run separate analyses for different subject types. Typically, the subject types used in the analyses are students versus professionals, but they may occasionally be more detailed (some primary studies make a distinction between undergraduate, master's and PhD students and professionals).
- The data corresponding to all subjects is analysed together in eight studies, where subject type is incorporated as a factor in the analysis of the experiment.

Therefore, the analysis of the impact of subject type on the results of experiments will be different depending on how it has been incorporated into the analysis of the experiment.

C. RQ3: Mapping the Impact of Subject Types to Results

When separate analyses are run for different subject types:

- Subject type does not have an influence on the results in two out of the five experiments examined.
- Subject type does have an influence on the results for at least one response variable in the remaining three experiments. More precisely:
 - Two studies show a difference in the sign of effect.
 - One study shows a difference in the size of effect.

When subject type is incorporated as a factor in the analysis, we have found that:

- Subject type does not have an influence on the results in three out of the eight experiments examined. Note that one of the studies failed to explore the interaction between subject type and treatments.
- In the remaining five experiments:
 - One study shows contradictory results.
 - Two studies found that subject type has an influence on the results of at least one response variable, but not on the interaction.
 - Two studies found that subject type has an influence on the results of at least one response variable, but did not explore the interaction.

These contradictory results suggest that further research is needed in order to investigate if there are other issues that could be influencing the results rather than the subject type.

ACKNOWLEDGMENT

Research funded by the Spanish Ministry of Economy and Competitiveness research grant TIN2014-60490-P.

REFERENCES

- [1] A. Höfer, W. F. Tichy. Status of Empirical Research in Software Engineering. *Empirical Software Engineering Issues. Critical Assessment and Future Directions*. Lecture Notes in Computer Science Volume 4336, 2007, pp 10-19, 2007.

³ Note that we have not taken into consideration experiments where the results for different subject types are reported in separate papers.

- [2] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson. Systematic mapping studies in software engineering. In *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE'08)*. pp. 71–80, 2008.
- [3] W.R. Shadish, T.D. Cook, D.T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Cengage Learning, 2001.
- [4] D.I.K. Sjøberg, J. E. Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N.-K. Liborg, A.C. Rekdal. A survey of controlled experiments in software engineering. *IEEE Transactions on Software Engineering*. 31(9): 733–753, 2005.