Replication of Empirical Studies in Software Engineering: An Update of a Systematic Mapping Study

Roberta M. M. Bezerra
Centre of Informatics, UFPE
Cidade Universitária, 50.740-560
Recife, PE, Brazil
rmmb@cin.ufpe.br

Fabio Q. B. da Silva
Centre of Informatics, UFPE
Cidade Universitária, 50.740-560
Recife, PE, Brazil
fabio@cin.ufpe.br

Anderson M. Santana
Centre of Informatics, UFPE
Cidade Universitária, 50.740-560
Recife, PE, Brazil
andmedsantana@gmail.com

Cleyton V. C. Magalhães
Centre of Informatics, UFPE
Cidade Universitária, 50.740-560
Recife, PE, Brazil
cvcm@cin.ufpe.br

Ronnie E. S. Santos
Centre of Informatics, UFPE
Cidade Universitária, 50.740-560
Recife, PE, Brazil
ress@cin.ufpe.br

Abstract—Context: Current empirical research highlight the need for replications of empirical studies because replications plays an important role in the construction of scientific knowledge. Objective: Considering the importance of replications in the consolidation of the knowledge produced in the software engineering research, this study aims to update and extend the results produced in a previous mapping study seeking to discuss the current state of the replication work of empirical studies performed in software engineering research between 2011 and 2012. Method: We applied the systematic review method to search and select published papers, to extract, and synthesize data from reported replications. Results: This study analyzed more than 7,000 articles, from which 39 articles that published replications between 2011 and 2012 were selected. Data extracted from these studies were used to update the information about the replications work in software engineering. Conclusion: The number of replications increased significantly in the period, when compared to the previous mapping study. In particular, the percentage of external replications also increased, with respect to internal ones. However, several other limitations identified in the previous mapping studies are still observed in this new set of replications.

Keywords— Replications, Empirical Software Engineering, Systematic Mapping Study.

I. INTRODUCTION

Replications of empirical studies are essential to the construction of scientific knowledge in any empirical Science. In the context of Software Engineering, Daly et al. [1] published the first paper that reports a replication of a study in 1994. Replications of empirical studies in Software Engineering are important to increase and consolidate the body of empirical knowledge in this area, as discussed by Juristo and Vegas [2]. Carver [3] also discusses that "one of the main benefits of an experimental replication is that it provides to researchers the ability to confirm, refute, or deepen the conclusions drawn from an earlier study."

As the research involving replications in software engineering evolve, important issues around these theme have been discussed, such as, the communication among the researchers of the replication and original study [4], the use of lab packages to support the execution of replication [5][6][7], guidelines to perform and report replications [4], the different types of replication [8][9][10], the generation of knowledge by performing replications [2] and the difficulties on performing replications that involve human aspects. Recently, da Silva et al. [11] published the results of a mapping study, in which the state-of-art of replications of empirical studies performed in the context of Software Engineering was presented.

The mapping study performed by da Silva et al. [11] analyzed the replications published since the study of Daly et al. [1] until 2010. This article aims to extend and update the results of da Silva et al.[11], performing a new mapping study in order to identify and analyze the replications of empirical studies in Software Engineering research published between 2011 and 2012. We used the same set of questions investigated previously by da Silva et al. [11] seeking to update the answer to the following main question: *RQ: What is the current state of the replication work of empirical studies performed in software engineering research?* This research question was subdivided in seven specific questions, detailed on the results of this study, which help to answer this main question.

From this introduction, this work is structured as follows: Section 2 summarizes the method used to guide our work. Section 3 presents and discusses the outcomes; and Section 4, briefly provides some conclusions and discussions.

II. METHOD

This study is an update and an extension of a mapping study previously published by da Silva et al. [11]. Thus, we applied the same research protocol used in this first mapping study to update the research results by adding replications performed between 2011 and 2012. The study was performed following three main phases.

A. Data Search Procedure

The search strategy combined automatic and manual search. We performed the manual search in the same journals and conferences used by da Silva et al. [11] and used the following

978-1-4673-7899-4/15/$31.00 ©2015 IEEE

engines in the automatic search: *ACM Digital Library, IEEEXplore, Scopus, Science Direct,* and *Springer*.

The automatic search was carried using a search string based on two main terms related to the research questions: *replication* and *software engineering*. We refined the previous string used by da Silva et al. [11] seeking to increase precision of the search process. We compared the results obtained with the new string with those from the previous one in other to check their consistent coverage. The results confirmed that the two strings retrieved the same studies, thus providing the same sensitivity whereas the new one is more precise. This is the new refined string used in this work: *("Software Engineering") AND ("Replication" OR "Replications" OR "Replicate" OR "Replicated" OR "Replicating")*

B. Study Selection

We selected papers reporting two categories of studies: a) Replications of empirical studies in software engineering; b) Conceptual and theoretical works about replications, including theories, definitions, taxonomies, lessons learned, etc. We excluded papers that met at least one of these exclusion criteria: a) published in years that were not 2011 or 2012; b) Not written in English; c) Not accessible on the Web; d) Invited papers, keynote speeches, workshop reports, theses, books and dissertations; e) Incomplete documents, drafts, slides of presentations, and extended abstracts; f) Secondary and tertiary studies, and meta-analyses; g) Addressing areas of computer science that were clearly not Software engineering (e.g., database systems, human-computer interaction, computer networks, etc.), and; h) Addressing replication only as part of future work.

Altogether, the processes of manual and automatic search retrieved about 7,000 papers. First, we performed a preliminary selection of studies by analyzing their titles and abstracts and we selected 57 papers potentially relevant. Then, we analyzed the studies considered as potentially relevant, by reading their abstract, introduction and conclusion. At the end, we identified 39 papers reporting 51 replications of 35 original studies previously published.

C. Data Extraction

Two researchers, working independently, performed data extraction to collect data to answer the research questions. Conflicts of extraction were solved in consensus meetings in which at least three researchers were involved.

III. Results and Discussions

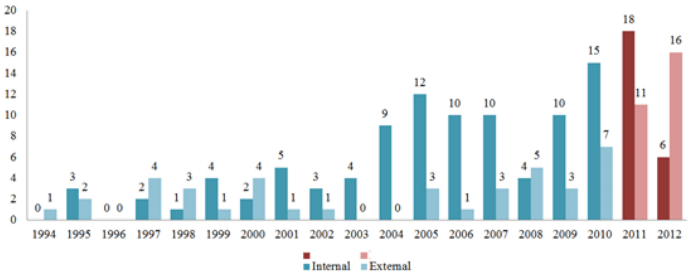
This mapping study identified articles reporting replications or conceptual work about replications published between 2011 and 2012. These studies can be subdivided in three groups: a) 39 studies that performed replications of empirical studies; b) 35 studies (original studies) that were replicated during this period; c) 10 papers presenting theoretical and conceptual research concerning issues about replications in software engineering. This third group is not analyzed in this paper. Magalhães et al. [12] presented an extensive analysis of these papers about replication. In the following sections, we provide answers to the detailed research questions.

RQ1: What is the evolution in the number and type (internal and external) of replications over the years?

The first mapping study [11] identified 96 papers reporting 133 replications of empirical studies published between 1994 and 2010, an average of 6.9 replications per year. The present mapping study identified 39 papers reporting 51 replications, published during 2011 and 2012, representing an average of 25.5 replications per year. The complete list of 39 papers can be found in: https://goo.gl/wx7NAC.

Our analysis considered the type of these replications: a) internal replications, those performed by the same researchers of the original study; b) external replications, those performed by a different group of researchers. Considering internal and external replications, the vast majority of replications published between 1994 and 2010 were internal replications. On the other hand, the replications identified on this current mapping study tend to a balance between internal and external ones, as presented in Figure 1.

Fig. 1. Evolution of Replications over the Years



It is possible to identify three distinct periods regarding the number of published replications. The first period (1994-2003), in which the interest of the theme started, had an average of 4.1 studies published per year. The second period (2004-2009) represents an increase of replications in software engineering, in which an average of 11.7 replications per year were published and the increase of internal replications is evident. Finally, the third period is related to the years 2010-2012 when an average of 24.3 replications per year were published and we observed the equilibrium between internal and external replications.

RQ2: Which individuals and organizations are most active in replications?

The first mapping study identified 194 researchers that published replication of empirical studies in software engineering between 1994 and 2010. In that set of studies, only 7% (14/194) of the researchers published both internal and external replications, and from the remaining researchers, 26% (50/194) published only external replications and 67% (130/194) only internal replications.

This current mapping study identified 125 researchers publishing replications between 2011 and 2012, in which 59% (74/125) of them published only internal replications, 38% (47/125) published only external replications and only 3% (4/125) of researchers published both internal and external replications. It shows that, regarding replication types, the researchers still form two almost disjoint sets.

We identified that only 21% (26/125) of the researchers published replications in the period between 1994 and 2010 and also in 2011 and 2012. Therefore, it seems that the development of replications is not a systematic activity of a

large group of researchers. We also identified 61 organizations that have published replications between 2011 and 2012, in which 52% (32/61) published only internal replications, 41% (25/61) only external replications and 7% (4/61) published both external and internal replications.

RQ3: What software engineering topics have been addressed by replications?

Following the analysis performed on the first mapping study, we classified the replications according to the topic areas of software engineering defined on the SWEBOK [13]. The first mapping study found *Software Requirements* as the most frequent topic, with 32 replications (24%) addressing studies in this topic. The second and third more researched topics were *Software Quality* and *Software Construction*, with 20 replications (15%) each. These three topics represented about 55% (72/133) of the replications identified by the authors of the first mapping study.

We identified *Software design* as the most replicated topic, followed by *Software testing* and *Software engineering models and methods*. These three topics represent over 55% (28/51) of the replications in this study. There is a shift in the focus of replication work, as none of the three most addressed topics remain the same between the two mapping studies. Table I shows the distribution of replications over the topics.

TABLE I. REPLICATIONS X SWEBOK CHAPTERS

SWEBOK Chapter	Internal	External
2. Software Requirement	0	1
3. Software Design	13	0
4. Software Construction	2	2
5. Software Testing	2	8
6. Software Maintenance	1	4
7. Software configuration management	0	0
8. Software engineering management	4	2
9. Software engineering process	0	0
10. Software engineering models and methods	0	7
11. Software Quality	2	3
Total	24	27

RQ4: What research methods are being replicated?

We classified the research method of each replication using the classification of empirical studies in software engineering proposed by Easterbrook et al. [14], also used on the first mapping study.

Similar to the first mapping study, the main research method replicated was quasi-experiments, as shown in Figure 2. We also did not find replications of ethnographies or action research, which according to da Silva et al. [11] are methods difficult to replicate due to the tacit knowledge needed to perform these studies and its context dependency.

We also identified the unity of analysis of each replication, and we classified the unity of analysis as *Artifacts* (software artifacts as unit of analysis); *Academics* (students as unit of analysis); and/or *Professionals* (practitioners as unit of analysis). Over 59% (31/51) of the replications identified between 2011 and 2012 used academic students as unity of analysis (Figure 3). Similar results were found by Sjøberg et al. [15] and Almqvist [15]. However, we observed an increasing participation of professionals in the replications when compared to the first mapping study.

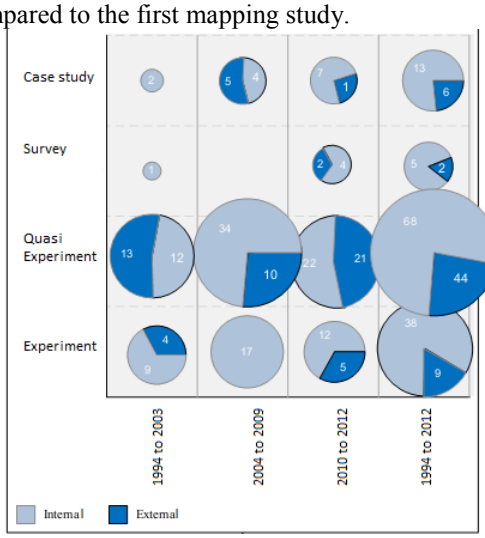


Fig. 2. Distribution of replications by research method.

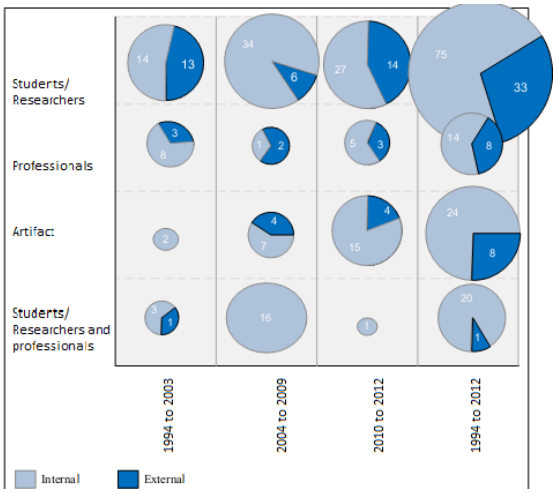


Fig. 3. Distribution of replications by unity of analysis

RQ5: What sets of replications were found?

We applied the same concept defined by da Silva et al. [11] to classify the sets of replications identified on this study, in which a set of replications is determined by the original study and the replications performed after and based on that. We identified 35 sets of replications. The majority, 80% (28/35), of these sets are composed only by an original study and one replication. Only 6% (2/35) of the sets include more than four replications. da Silva et al [11] obtained 61% of sets composed of a single replication, thus we observed an increase in proportion of sets with only one replication.

RQ6: Did the replications confirm the results of the original studies?

We compared the original studies and replications to verify if the results from original studies were confirmed in the replications. We used the labels *Yes*, *No* and *Partially* to categorize the replications with respect to the confirmation of results of the original study. This classification was based on

the conclusions presented by the authors of the papers and no further analysis was performed. Both this study and the first mapping study found similar results with a high level of confirmations on internal replications and low levels of confirmations presented on external replications (Table II).

TABLE II. DISTRIBUTION OF REPLICATIONS BASED ON THE CONFIRMATION OF ORIGINAL STUDY RESULTS

Confirmation of results	Internal Replication		External Replication		Total	
	1994 - 2010	2011 - 2012	1994 - 2010	2011 - 2012	1994 - 2010	2011 - 2012
Yes	82%	84%	26%	33%	65%	57%
Partially	9%	8%	28%	22%	15%	16%
No	9%	8%	46%	45%	20%	27%

RQ7: What was the elapsed time between the replications and corresponding original studies?

To answer this question, we observed the year in which both the original study and the replications were published. In this context, we did not consider replications and original studies published in the same paper (called single-report replications by da Silva et al. [11]). This means that the original study and replication were published in the same year and the time elapsed between the original study and replication was not mentioned. We observed similarities between the first mapping study and this study, considering the set of internal replications. We observed that the time between the external replications and the originals studies increased over the years, from 4.4 years on average in the first mapping study to 7.5 years on this study (TABLE III).

TABLE III. ELAPSED TIME

	1994 a 2009	2010 a 2012
Internal replications	4	3.3
External replications	4.4	7.5
All papers	4.2	5.9

IV. CONCLUSION AND FUTURE WORKS

We presented an extension of a mapping about replications of empirical studies engineering, in which 39 papers reporting 51 replications, published between 2011 and 2012, were identified and analyzed. The number of replications increased in this period with an average of nearly 26 replications published by year against an average of only 7 replications published by year in the period between 1994 and 2010. Further, the relative number of external replications has also increased from 29% (39/133) in the first period to 53% (27/51) in 2011 and 2012.

Both findings may be an (direct or indirect) effect of the Workshop RESER, which was held for the first time in 2010, and again in 2011 and 2013. RESER stimulated the discussion of topics related to replication and also the development of joint replications. The joint replications had an impact on the number of replications and an even larger impact on the relative increase in the external ones.

The number of sets of replication composed of a single replication increased in 2011 and 2012. Further, the confirmatory tendency of internal replications remained (as well as the non-confirmatory tendency of the external ones), despite a small increase in the confirmation of results provided by external replications.

Finally, although the increase in the number of external replications may indicate that more researchers are becoming aware of the importance of this type of research, the absolute numbers are still small and several of the potential biases and limitations found by da Silva et al. [11] still remain.

ACKNOWLEDGMENT

Fabio Q. B. da Silva holds a research grant from CNPq #314523/2009-0. Cleyton V. C. de Magalhães and Ronnie E. S. Santos are PhD students and receive a scholarship from CNPq.

REFERENCES

[1] J. Daly. et al, Verification of Results in Software Maintenance Through External Replication. IEEE International Conference on Software Maintenance. Glasgow: 1994.

[2] N. Juristo, S. Vegas, Using Differences among Replications of Software Engineering Experiments to Gain Knowledge. International Symposium on Empirical Software Engineering and Measurement, p. 356–366, 2009.

[3] J. C. Carver, Towards Reporting Guidelines for Experimental Replications□: A Proposal. International Workshop on Replication in Empirical Software Engineering Research. Cape Town: 2010.

[4] S. Vegas, et al. Analysis of the Influence of Communication between Researchers on Experiment Replication International Symposium on Empirical Software Engineering. Rio de Janeiro: 2006

[5] F. J. Shull et al. The role of replications in Empirical Software Engineering. Empirical Software Engineering, v. 13, n. 2, p. 211–218, 29 jan. 2008.

[6] B. Kitchenham, The role of replications in empirical software engineering—a word of warning. Empirical Software Engineering, v. 13, n. 2, p. 219–221, 29 jan. 2008.

[7] L. P. Scatalon, R. E. Garcia, R. C. M. Correia, Packaging Controlled Experiments Using an Evolutionary Approach Based on Ontology. International Conference on Software Engineering and Knowledge Engineering. Miami: 2011

[8] O. Gómez, N. Juristo, S. Vegas, Replications Types in Experimental Disciplines. International Symposium on Empirical Software Engineering and Measurement. Bolzano-Bozen: 2010a

[9] O. Gómez, N. Juristo, S. Vegas, Replication, Reproduction and Re-analysis□: Three ways for verifying experimental findings. 2010b.

[10] N. Juristo, S. Vegas, Using Differences among Replications of Software Engineering Experiments to Gain Knowledge. International Symposium on Empirical Software Engineering and Measurement, p. 356–366, 2009.

[11] F. Q. B. da Silva et al., Replication of empirical studies in software engineering research: a systematic mapping study. Empirical Software Engineering, p. 1–57, 1 set. 2012.

[12] C. V. C. de Magalhães, F. Q. B. da Silva.; R. E. S. Santos; M. Suassuna, Investigations about replication of empirical studies in software engineering: A systematic mapping study. Information and Software Technology, Volume 64, August 2015, p. 76-101, ISSN 0950-5849, http://dx.doi.org/10.1016/j.infsof.2015.02.001.

[13] SWEBOK. Guide to the Software Engineering Body of Knowledge - SWEBOK. 4. ed. USA: Angela Burgess, 2004. p. 174.

[14] S. Easterbrook et al. Selecting Empirical Methods for Software Engineering Research. In: F. Shull, J. Singer, D. I. K. Sjøberg (Eds.). Guide to Advanced Empirical Software Engineering. London: [s.n.]. p. 285–311.

[15] D. I. K. Sjøberg et al., A Survey of Controlled Experiments in Software Engineering. IEEE Transactions on Software Engineering, v. 31, n. 9, p. 733–753, 2005.

[16] J. P. F. Almqvist, Replication of Controlled Experiments in Empirical Software Engineering — A Survey. Master Dissertation. Lund University, 2006