



# Investigations about replication of empirical studies in software engineering: A systematic mapping study<sup>☆</sup>



Cleyton V.C. de Magalhães<sup>\*</sup>, Fabio Q.B. da Silva, Ronnie E.S. Santos, Marcos Suassuna

Centre for Informatics, Federal University of Pernambuco, Recife 50.740-560, Brazil

## ARTICLE INFO

### Article history:

Received 11 September 2014

Received in revised form 28 January 2015

Accepted 2 February 2015

Available online 14 February 2015

### Keywords:

Replications

Experiments

Empirical studies

Mapping study

Systematic literature review

Software engineering

## ABSTRACT

**Context:** Two recent mapping studies which were intended to verify the current state of replication of empirical studies in Software Engineering (SE) identified two sets of studies: empirical studies actually reporting replications (published between 1994 and 2012) and a second group of studies that are concerned with definitions, classifications, processes, guidelines, and other research topics or themes about replication work in empirical software engineering research (published between 1996 and 2012).

**Objective:** In this current article, our goal is to analyze and discuss the contents of the second set of studies about replications to increase our understanding of the current state of the work on replication in empirical software engineering research.

**Method:** We applied the systematic literature review method to build a systematic mapping study, in which the primary studies were collected by two previous mapping studies covering the period 1996–2012 complemented by manual and automatic search procedures that collected articles published in 2013.

**Results:** We analyzed 37 papers reporting studies about replication published in the last 17 years. These papers explore different topics related to concepts and classifications, presented guidelines, and discuss theoretical issues that are relevant for our understanding of replication in our field. We also investigated how these 37 papers have been cited in the 135 replication papers published between 1994 and 2012.

**Conclusions:** Replication in SE still lacks a set of standardized concepts and terminology, which has a negative impact on the replication work in our field. To improve this situation, it is important that the SE research community engage on an effort to create and evaluate taxonomy, frameworks, guidelines, and methodologies to fully support the development of replications.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Replications of empirical studies play important roles in the construction of knowledge. According to Schmidt, a replication that demonstrates the same findings obtained by other experiment "... is the proof that the experiment reflects knowledge that can be separated from the specific circumstances (such as time, place, or persons) under which it was gained" [2]. Replications are also important to identify the range of conditions under which findings from one experiment hold and the possible exceptions [3].

Considering the importance of replications in the advance of science in general, Schmidt [2] expected that one would find a

body of knowledge that provide clear and unambiguous definitions for central questions like 'what exactly is a replication experiment?', 'what exactly is a successful replication?', and 'what are all types of replication and their corresponding roles?'. Furthermore, one would expect to find empirically evaluated guidelines on how to perform and report replications complementing existing guidelines to perform experiments and other empirical studies.

However, Schmidt argues that this is not true for most of scientific disciplines [2]. The published replications and the theoretical works about replication research have not used clear-cut definitions of terms and concepts, and there is no generally accepted taxonomy to distinguish between types of replications and their roles in generating scientific knowledge. According to Schmidt, "the word replication is used as a collective term to describe various meanings in different contexts" [2]. Carver et al. [4] report that a similar situation is also found in empirical software engineering

<sup>☆</sup> *Article Notes:* Preliminary and partial results of this study have been presented at the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE'2014) and published in the Conference Proceedings [1].

<sup>\*</sup> Corresponding author.

research. Our findings reinforce the need to address these issues in software engineering.

The goal of this article is to contribute to the advance of the replication work in empirical software engineering. We expect that the results presented in our study will stimulate and provide support for a debate in the scientific community to central questions related to replications. Although we do not expect to fully answer these questions in this article, we believe our work will contribute to some of the answers:

*What should be considered a replication?*  
*What should be considered a successful replication?*  
*What are the types of replications and their functions?*  
*How should replications be performed?*  
*How should replications be reported?*

In a recent mapping study, da Silva et al. [5] studied the current state of published replications of empirical studies in software engineering research. The mapping study selected and analyzed papers reporting replications of empirical studies published until 2010 and also found a second set of studies addressing several topics about replication work. The papers about replication were not further analyzed by da Silva et al. [5]. More recently, the same research group performed an update of the mapping study previously published, covering material published in 2011 and 2012 [6]. Also in this update, the same type of papers about replication were collected and saved for future analysis.

In this current article, we analyze and discuss the content of the papers about replications (hereafter referred to as ABO papers) published in the Software Engineering literature to increase our understanding about the current state of the work on replication in empirical software engineering research. We expect that this analysis will shed some light in the issues related to the five questions raised above.

Our goal is twofold. First, to classify the set of ABO studies in Software Engineering into categories related to the topics in which the articles focused on (recommendations, frameworks, guidelines, among others). Second, to analyze how the replications performed between 1994 and 2012 have cited and used the ABO studies, in order to verify the impact of these studies in recent replication work.

The set of papers analyzed in this article is composed of those selected by da Silva et al. [5], those found in the update of the mapping study [6], and papers found through a search process performed to cover work published in 2013. We systematically structured and analyzed data extracted from these articles to answer the following six research questions:

- RQ1: What was the evolution in the number of ABO studies over the years?
- RQ2: Which individuals and organizations are most active in publishing ABO studies?
- RQ3: How the ABO studies define replication?
- RQ4: What topics or themes have been addressed by the ABO studies?
- RQ5: Which ABO studies are cited by the papers that reported replications?
- RQ6: How the results or propositions presented in the cited ABO studies have been used in papers that report replications?

This article is organized as follows. In Section 2, we present a background with discussion on concepts and related works. In Section 3, we present the method used in this study. In Section 4, we present a comprehensive set of results of our review and in Section 5 we discuss these results. Finally, in Section 6, we present some conclusions and proposals for future works.

## 2. Background and related work

As briefly discussed in the Introduction, there is little agreement about nomenclature and definition of concepts about replication in many empirical sciences and also in empirical software engineering. In this article, we expect to shed some light on the debate about some theoretical and practical issues related to performing, classifying, and reporting replications in SE research. In this section, we start by providing some preliminary definitions, we then briefly describe the two mapping studies on replication that originated this current study and clarify some terminology issues. Finally, we show how this article improves the preliminary results published by Magalhães et al. [1].

### 2.1. Definition of replication

According to La Sorte, “replication refers to a conscious and systematic repeat of an original study” [7]. This definition implies that a replication must be explicitly related (conscious repetition) to a previous study. Similarly, A Dictionary of Social Sciences [8] defines replication as “a repetition of a research procedure to check the accuracy or truth of the findings reported”. In fact, most definitions found in the scientific literature consider a replication to be a repetition of a research procedure already performed in another study, usually called the *original* or the *baseline study*.

This definition is a starting point in precisely characterizing what should be considered a replication. According to this characterization, empirical studies that address similar questions or hypothesis, but without explicit reference to a previous study that can be considered the original study, should not be considered replications. For this reason, da Silva et al. [5] do not consider as replications the studies that Krein and Knutson (2010) [ABO022] classify as independent replications. Similarly, we also do not consider replications the type of study that Baldassarre et al. [20] call conceptual replications. The reason in both cases is that the (very similar) definitions of independent and conceptual replication admit studies to be called replications without a reference (direct or indirect) to an original study.

However, because of the variations that may be intended or unintended introduced in the replication design, the definitions presented above are not precise enough to characterize unambiguously what should be considered a replication and what should be seen as an entirely different study. We expect that this article motivates the research community to engage on an effort in building standardized and consistent set of definitions and corresponding terminology related to replication work in SE research.

### 2.2. A brief summary of the mapping studies

The first article that explicitly reported a replication of an empirical software engineering study was published in 1994 [9]. The mapping study presented by da Silva et al. [5] analyzed 96 articles reporting 133 unique replications of 72 original studies published between 1994 and 2010. Bezerra and da Silva [6] updated da Silva’s work and found 39 new articles, reporting 51 replications of 35 original studies, published in 2011 and 2012.

Using the definition of internal and external replication proposed by Brooks et al. [ABO036] to classify the replications, Fig. 1 shows the evolution of the number of replications found in the two mapping studies (da Silva et al. [5] is presented in blue and Bezerra and da Silva [6] in red).

da Silva et al. [5] and Bezerra and da Silva [6] raise several questions about the replication work in SE research. According to both studies, no clear cut definition of replication has been used in the studies, there is little standardization on how to report the

Download English Version:

<https://daneshyari.com/en/article/550993>

Download Persian Version:

<https://daneshyari.com/article/550993>

[Daneshyari.com](https://daneshyari.com)