Instanciación Plantilla para Especificar Cambios de Replicaciones de Experimentos en Ingeniería del Software

Margarita Cruz, Beatriz Bernárdez, Amador Durán, Antonio Ruiz

Dept. Lenguajes y Sistemas Informáticos, University of Seville,
Avda. Reina Mercedes s/n, 41012, Seville, Spain.
{cruz,beat,amador,aruiz}@us.es

Resumen *Contexto*: AA La replicación de estudios empíricos en Ingeniería del Software es necesaria para consolidar el conocimiento adquirido. No obstante, para incrementar el conocimiento que se genera mediante la replicación, es necesario que la información se publique de forma que permita una comprensión profunda del estudio. *Objetivo*: Al diseñar una replicación, habitualmente surge la necesidad de introducir cambios. El objetivo de este trabajo es facilitar la especificación de dichos cambios proponiendo una plantilla que permita definirlos sistemáticamente y documentarlos de forma homogénea. *Método*: Se ha definido el metamodelo para formalizar la información sobre replicaciones y cambios que son relevantes para la plantilla propuesta. Posteriormente, se ha detallado la plantilla y se ha aplicado a una replicación concreta. *Resultados y Conclusiones*: La aplicación de la plantilla a una replicación concreta ha sido satisfactoria pero se debe aplicar a otras replicaciones o familias de experimentos para su validación y mejora a través de la retroalimentación obtenida. Sistematizar la concepción de los cambios y la manera de documentarlos facilita su comprensión para otros investigadores

Keywords: replicación, plantilla, patrones lingüísticos, estudio empírico, ingeniería del software empírica

1. Introduction

The Empirical Software Engineering (ESE) allows the evaluation of new methods, techniques and tools to know the convenience of using them in the development process.

Once an artifact has been evaluated for the first time, the study needs to be replicated in different contexts and conditions. Not only to consolidate the knowledge acquired, but also to know if its results can be generalized [4].

Despite the importance of replication, and although the practice of replication has increased in recent years [11], the number of replications in Software Engineering (SE) remains low [30].

When it comes to identifying the causes of this situation, there are several problems encountered. On the one hand, the lack of single criteria for reporting replications [10]. On the other hand, tacit knowledge, since the researcher has knowledge that is not explicit in the publication [27]. In addition, there is a lack or incompleteness of the so-called *laboratory packages* that are necessary to be able to make replications [30]. All this is coupled with the effort and resources needed to carry out an experiment [11].

In areas of knowledge other than SE such as psychology, this situation is known as the *replicability crisis* [8]. It also includes problems with the use of statistics, large numbers of published false positives and lack of rigour due to individual biases in evidence collection. In order to mitigate this crisis, in [8] it is suggested to be rigorous in the methodological aspects and an adequate transparency in the transmission of the research results.

According to [15], it is recommended that the first replications of the original experiment be carried out by the same experimenters (internal replications), to avoid the influence of variables that may interfere when changing sites or experimenters. Subsequently, it is valuable to perform external replications as they allow us to know the range of conditions under which the results are maintained and demonstrate that they are independent of the conditions of the original study [28,9].

Whether the replication is internal or external, there may be a need to incorporate changes to the original experiment for various reasons, such as expanding the results of the initial study [28] or adapting an experiment to a different environment than the original [4].

As part of the documentation to be published on the replications, the Carver guidelines [10] highlight the need to describe these changes through a specific section on changes to the original experiment. For each change from the original experiment, it is necessary to record what it is, the situation that caused the change and determine its impact on the threats to the validity of the replication. In addition, an in-depth knowledge of the changes makes it easier to see whether the fundamental principle of independence based on aggregated data analysis according to Kitchenham [22], which promotes the introduction of changes in replication to avoid propagating problems from the original experiment, is met. In contrast, Juristo et al. [18] assert that even the tiniest change in a replication can lead to inexplicable differences in the results.

In this work, a template for the definition of replication changes is applied and evaluated. This template is based on a metamodel to formalize the information and is completed with linguistic patterns that facilitate the writing process [citaJisBD]. The use of the proposed template has a double purpose: on the one hand, it invites the researcher to make the change and its details explicit (reducing tacit knowledge) and, on the other hand, it helps the reader to better understand the replication and to follow the trace of how the original experiment evolves in the succession of experiments of a family.

The remainder of this paper is organized as follows: Section 2 presents a brief introduction about replications; Section 3 discusses the previous related works to this research area; Section 4 covers the methodology used in this SMS; Section 5 presents results from analyzing the primary studies; Section 6 analyzes the threats to validity; Section 7 include some findings found; and Section 8 presents the concluding remarks and future work.

2. Metamodel on replications and changes

This section presents the metamodel on which the template is based (Fig. 1) using UML class diagram notation.

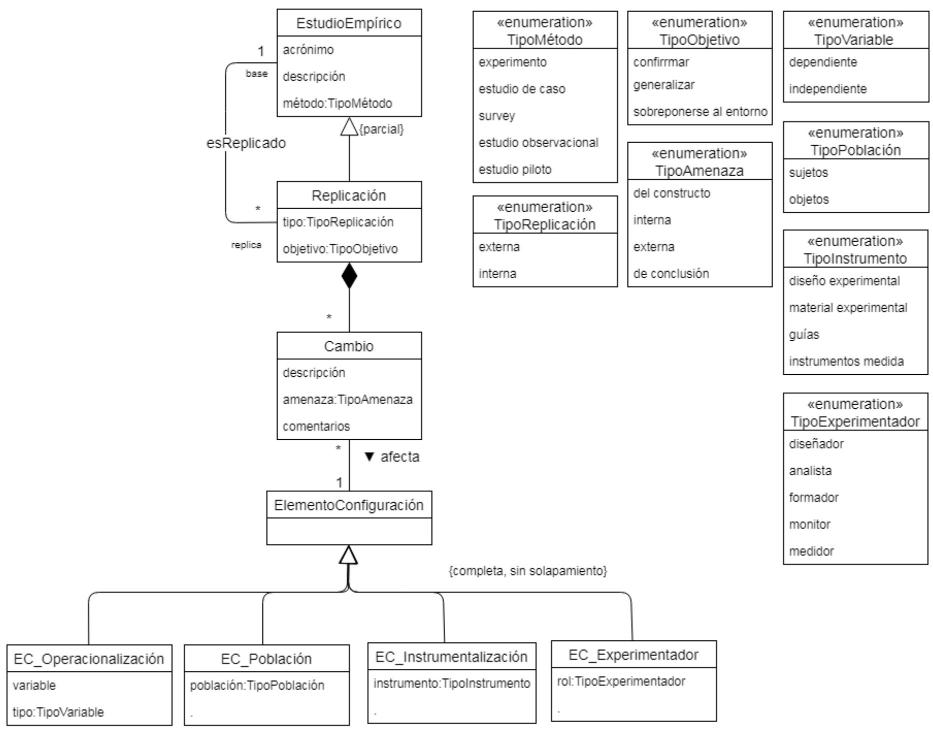

Figura 1. UML class diagram

In the metamodel, a replication is a type of empirical study based on another empirical study (either the original or another replication). The inheritance is partial since objects that are instances of the "original experiment" are considered in the supertype. Among the attributes of the Replication class, they appear:

- **Acronym**. A code or acronym relating to the reference experiment is used.

- **Empirical method**. The three main empirical methods are *controlled experiment*, *survey* y *case study*. The empirical method *quasi-experiment*, in which the assignment of subjects to treatment is not random, are included in the *controlled experiments* [33]. Although less frequent, we also consider two other empirical methods: *pilot study* and *observational study*.
- **Type of replication**.According to who does the replication, these are classified as *internal* (they are performed by the original experimenters) and *external* (they are performed by independent experimenters and therefore their power of confirmation is greater than in the internal ones) [9]. There are other taxonomies based on various criteria, for example a possible classification is according to the degree to which the original experiment procedure is followed. In this regard, there are authors who speak of *exact* and *conceptual* replications [28], *closed* and *differentiated* [1,23,20]. Basili et al. [6] refer to *strict* replications and present a classification into three main groups depending on whether or not the research hypothesis is varied or the theory is expanded. In [16] a classification of the replications in literal, operational and conceptual is proposed depending on the changes carried out and the purpose of the replication. Due to the lack of agreement on the terminology [16] and the differentiating nuances, we have chosen to describe the changes by classifying the replications as *internal* or *external* without entering into other more confusing categories.
- **Target**. For the purpose of replication, a enumerated type has been defined that can take three possible values: i) *Confirm results*, to verify the results obtained, ii) *Generalize the results* includes, among others, changes in the experimenters to analyze if the results are a consequence of the intervention of these experimenters, changes in the population and metrics to see if the results are still met or if the hypothesis is still valid, iii) *Overcoming some limitations of the original experiment environment*, reflects the changes that occur as a result of the constraints imposed and to which the experiment needs to be adapted.

To reflect the relationship between the replication class and the class that describes its changes, a composition has been used since the changes are an intrinsic part of the replication. Its attributes are:

- **Description**. It allows to describe what the change is about.
- **Validity threat addressed**. It's a type enumerated with the values:*increases the validity of the construct*, *external*, *internal* and of *conclusion*. This attribute serves to indicate how the change introduced mitigates the threat to the validity of a given type.
- **Remarks**. Additional information about the change.

Each change is carried out on an element of the experimental configuration. In [16] the elements of the experimental configuration that can be changed in a replication are identified. These elements are grouped into dimensions. In the class diagram, a hierarchy has been used to classify the configuration elements into the four dimensions. The classification is *complete* and *without overlap*.

- **Dimension and elements**. The dimensions identified in [16] and represented in the metamodel by the subtypes are:
  - *Operationalization*: includes changes related to *dependent variables* (changes in metrics and measurement procedures) and *independent variables* (changes in the way treatment is applied). It corresponds to the EC_Operationalization subtype with the attributes: variable (name of the variable to be modified) and type that can take the values *dependent* or *independent*.
  - *Population*: includes changes in the properties of the *experimental subjects* (e.g. different experience, age, etc.) and *objects* (e.g. different programming language, difficulty level, etc.). It corresponds to the EC_Population subtype with the population attribute that can take the values *subjects* or *objects*.
  - *Protocol*: includes changes in the *experimental design*, in the *experimental material* (when using different instances of the same type of experimental object, e.g. changing the formulation of one problem to another of equal difficulty), in the *guides* (e.g. change the instructions provided to subjects) and in the *measuring instruments* (e.g. change the questionnaire for data collection). It corresponds to the EC_Protocol subtype with the instrument attribute that can take the values *experimental design*, *experimental material*, *guides* o *measuring instruments*.
  - *Experimenters*: regarding changes in the roles of experimenters. It corresponds to the EC_Experimenter subtype with the role attribute that can take the values *designer*, *analyst*, *trainer*, *monitor* o *measurer*.

3. Template for replication changes

This section presents, in summary form, the template to facilitate the specification of replication changes already proposed in [citaJisBD].

The use of templates helps to structure the information in a fixed form, reduces ambiguity, facilitates reusability and also serves as a guide to avoid missing relevant information [13]. For some fields in the templates, you have identified phrases that are common and have been parameterized. These phrases are called lingustic patterns (L–patterns) [31].

In the notation used to describe the L-patterns, words or phrases between $<$ and $>$ must be properly replaced, words or phrases between $\{$ and $\}$ and separated by $|$ represents options; only one option must be chosen and words between [and] are optional, that is, they may or may not appear when the template is instantiated. Filling blanks in pre–written sentences, i.e. L–patterns, is easier and faster than writing a whole paragraph. Table 2

Table 2 shows the proposed template. The meaning of the fields is as follows:

- **Acronym for Replication**: In order to obtain a quick identification of the different experiments of a family, it is useful that the description of the set of

<table>
<tr><th><i><Acronym></i> #<<i>n</i>></th><th>Replication of experiment <<i>Acronym of experiment</i> >#<<i>m</i>></th></tr>
<tr><td>Empirical method</td><td>{Controlled experiment | Survey | Case study |
Observational study | Pilot study }</td></tr>
<tr><td>Type of replication</td><td>{Internal | External }</td></tr>
<tr><td>Target</td><td>{Confirm results |
Generalize the results |
Overcoming some limitations of the original experiment environment |
Custom target }</td></tr>
<tr><td>Change #<<i>i..j</i>></td><td>Originally, <<i>description of the situation in the original experiment</i>>
In replication <<i>description of the situation in replication</i>>
{in order to | because of } <<i>cause of change</i>></td></tr>
<tr><td>[Modified Dimension]</td><td>{Operationalization [(, in particular, of the {dependent |independent} variable <<i>variable name</i>>)] |
Population[(, in particular, experimental {subjects |objects})]|
Protocol [(, in particular, {experimental design |experimental material |guides |measuring instruments })] |
Experimenters [(, in particular, { designer |analyst |trainer |monitor |measurer})] }</td></tr>
<tr><td>Validity threat addressed</td><td>The change increases {the construct |external |internal |conclusion } validity</td></tr>
<tr><td>[Remarks]</td><td><<i>Remarks</i>></td></tr>
</table>

Cuadro 1. Template for specifying changes in a replication

replications of the same experiment begins with a code or acronym relative to the reference experiment and followed by a sequential number.

- **Acronym of experiment**: Same as the replication. When the reference experiment is the original experiment, the sequential number following it should be 1, so that it can be easily identified when the reference experiment is already a replication.
- **Environment**: Site and date of replication.
- **Empirical method**: One of the possible values is selected: *controlled experiment*, *survey*, *case study*, *pilot study* and *observational study.*
- **Type of replication**: You select between: *internal* or *external.*
- **Target**: Allows you to select between three possible values or free text: *confirm results*, *Generalize the results*, *Overcoming some limitations of the original experiment environment* or *Custom target.*
- **Description of the change**: The changes are numbered sequentially and their description contains a L-pattern that must be completed. The situation in the original experiment, the change itself and the cause or consequence are recorded. In this last part of the pattern, choose between "to the end of" or "caused by" to complete the sentence.
- **Modified dimension and elements**: The L- pattern is completed by choosing one of the proposed dimensions. Within the selected dimension, you can specify the modified element by choosing one of the options presented.

Both the dimension and the element concerned are optional. In other words, for simplicity's sake, it is possible to describe a change without having to identify the dimension and the element affected by the change.

- **Validity threat addressed**: One of the threats is chosen: *the construct*, *external*, *internal* or *conclusion* validity [33].
- **Remarks**: Additional information about the change.

4. Application of the template

In this section the template is applied to the definition of the changes made in the replications carried out in three series of experiments in the SI area that deal with: i) Mindfulness; ii) Requirements analysis, and; iii) Code evaluation techniques.

4.1. A replication of a Mindfulness experiment

A Series of Mindfulness experiments. Añadir la segunda replicación bernardezTse2017.

This is an internal replication carried out at the University of Seville and described in [7] by some of the authors of the present article that deals with the effect of the practice of *Mindfulness* on the performance of students when developing conceptual models.

This is the first instantiation of the template. With its current structure we are able to make explicit the changes in the replication cited. Although it is clear that template validation needs to be applied to other replications.

The acronym used to refer to the original experiment and replication is mind#1 and mind#2 respectively.

Lessons learned

- We realize the need to add the date and location of the replication to the template. A new "site "dimension is identified in [15]; if replication is carried out at a site other than the original experiment, the site dimension is changed.

4.2. Family of experiments on Requirements analysis

This is an empirical study of the influence of analyst domain knowledge and experience on the effectiveness of requirements analysis. The family consists of a series of experiments in which postgraduate students from the Polytechnic University of Madrid participate, as well as professionals from different countries and institutions. The study is described in [3].

According to the author, the acronyms for the experiments are: Q-2007, Q-2009, Q-2011, Q-2012, E-2012A, E-2012B, E-2013, E-2014 and E-2015. Each experiment is a replication of the previous experiment. In other words, the template has been applied defining the changes in each replication with respect to the previous experiment except the E-2015 experiment which is a replication of the E-2013.

Lessons learned

- When a change involves the suppression or appearance of a dependent or independent variable, it is considered to affect the operationalisation dimension. Regarding the threat to validity, it does not affect the construct validity.
- There are changes that affect the protocol dimension but it is difficult to identify the affected element.
- There are replications with a specific target and therefore they do not fit into any of the three options proposed in the template. It is necessary to include the option"custom target".
- The target "To overcome some limitations of the original environment of the experiment"has been deleted. It has been added as one of the possible causes for change. A change can be motivated by environmental conditions but is not the goal of replication.
- Although replication is identified in the template by a code or acronym relating to the reference experiment and followed by a sequential number, it is advisable to follow the author's notation if different.
- Although when defining the template, quasi-experiments are included in the controlled experiments, if the author defines replication as a quasi-experiment, this will be reflected in the template.

4.3. A series of experiments on Code evaluation techniques

It consists of an experiment and its four replications. The original experiment is conducted with students at the Polytechnic University of Madrid to evaluate the effectiveness of three verification and validation techniques. This experiment has been replicated in four different sites: the Polytechnic University of Madrid, the Polytechnic University of Valencia, the University of Seville and the ORT University of Uruguay. The study is described in [17,21,19].

The acronyms used to reference the original and the replications are VV-UPM, VV-UPM1, VV-UPV, VV-Uds and VV-ORT.

Lessons learned

- The order in which the different steps are executed is considered the new "order of application" element of the protocol dimension.
- In the analyzed replications, no changes have been found in the experimental dimension since neither the experimenter nor his role is specified.

<table>
<tr><th><<i>Acronym</i>> #<<i>n</i>></th><th>Replication of experiment <<i>Acronym of experiment</i> >#<<i>m</i>></th></tr>
<tr><td>Environment</td><td><<i>Date</i>>
<<i>Place</i>></td></tr>
<tr><td>Empirical method</td><td>{Controlled experiment | Quasi-experiment | Survey | Case study | Observational study | Pilot study }</td></tr>
<tr><td>Type of replication</td><td>{Internal | External }</td></tr>
<tr><td>Target</td><td>{Confirm results |
Generalize the results |
<<i>Custom target</i>>}</td></tr>
<tr><td>Change #<<i>i..j</i>></td><td>Originally, <<i>description of the situation in the original experiment</i>>
In replication <<i>description of the situation in replication</i>>
{in order to | because of } {Overcoming some limitations of the original experiment environment |<<i>cause of change</i>>}</td></tr>
<tr><td>[Modified Dimension]</td><td>{Operationalization [(, in particular, of the {dependent |independent} variable <<i>variable name</i>>)] |
Population[(, in particular, experimental {subjects |objects})]|
Protocol [(, in particular, {experimental design |experimental material |guides |measuring instruments |order of application })] |
Experimenters [(, in particular, { designer |analyst |trainer |monitor |measurer})] }</td></tr>
<tr><td>Validity threat addressed</td><td>The change increases {the construct |external |internal |conclusion } validity</td></tr>
<tr><td>[Remarks]</td><td><<i>Remarks</i>></td></tr>
</table>

Cuadro 2. Template for specifying changes in a replication after applying lessons learned

4.4. Analysis of the changes

This section presents the changes identified in each replication. We analyze whether the change fits the template or the difficulties we have had in trying to define each change according to the template and language patterns defined. Table 6. Los tipos de cambios Table 3

Type	Description
T0	Fits the template
T1	It is a change in the level (factor) of the independent variable (operationalization dimension) but it can also be considered a change in the protocol dimension
T2	A dependent variable is suppressed or added. The operationalization dimension is modified but no validity threat is addressed
T3	A independent variable is suppressed or added. The operationalization dimension is modified but no validity threat is addressed
T4	It's a compound change. It is convenient its decomposition in simple changes
T5	It is not clear whether the change affects the operationalization or protocol dimension
T6	The change modifies the protocol dimension; the affected element is not identified
T7	The change modifies the protocol dimension; the affected element is the change in the order of actions. This element is new
T8	The change modifies the operationalization dimension; the affected variable is not identified

Cuadro 3. Types of problems identified in the changes

Id	Baseline	Change	Description	Type
Mind# 2	Mind# 1	Change-1	Increase in the number of sessions	T0
		Change-2	Random subject assignment	T0
		Change-3	The public speaking workshop is held at the end of the workshop.	T0

Cuadro 4. Analysis of the changes

5. Related works

This section analyzes related work that: i) presents guidelines on how to report changes in SI replications; ii) presents such changes in an easy-to-interpret way, e.g., through tables; and iii) defines or uses templates or patterns.

Carver [10] presents an initial proposal for guidelines on the content of publications reporting replications. In this article, the content that should be

Id	**Baseline**	**Change**	**Description**	**Type**
Q-2009	Q-2007	Change-1	Effectiveness is not measured	T2
		Change-2	Retention capacity is not measured	T2
		Change-3	The influence of the development experience is analysed	T3
		Change-4	Interviews are conducted in English	T6
		Change-5	The respondent is changed	T6
Q-2011	Q-2009	Change-1	Group interviews	T6
		Change-2	The influence of skill on requirements and skill in interviews is analyzed	T3, T4
		Change-3	Increases the interview time	T5, T6
		Change-4	Decreases the time to present information	T6
		Change-5	The consolidation time is limited	T6
		Change-6	The respondent is changed	T6
Q-2012	Q-2011	Change-1	The subjects are professionals	T0
		Change-2	The influence of development skill is analyzed	T3
		Change-3	Decreases consolidation time	T6
		Change-4	No training period	T3, T8
E-2012A	Q-2012	Change-1	The influence of knowledge is analysed	T3
		Change-2	Change in experimental design	T0
		Change-3	Individual interviews	T6
		Change-4	The language is a blocking variable.	T0
		Change-5	The respondent is a blocking variable	T0
		Change-6	There are two responders	T6
		Change-7	Groups are formed	T0
		Change-8	Decreases elicitation time	T6
		Change-9	Increases consolidation time	T6
		Change-10	The influence of the difficulty of the problem is analysed	T3
E-2012B	E-2012A	Change-1	The problem domains are modified	T0
		Change-2	The order of the problems is changed	T7
		Change-3	The experiment is conducted at the end of the course	T5, T6
E-2013	E-2012B	Change-1	Change in experimental design	T0
		Change-2	Training is provided	T8
E-2014	E-2013	Change-1	There are one responders	T6
		Change-2	Increases training time	T8
E-2015	E-2013	Change-1	Increases training time	T8

Cuadro 5. Analysis of the changes

Id	**Baseline**	**Change**	**Description**	**Type**
VV-UPM1	VV-UPM	Change-1	The visibility of the error is analysed	T3, T8
		Change-2	Two versions of each program	T3
		Change-3	All types of failures are duplicated	T0
		Change-4	Test cases are provided	T0
		Change-5	A program is discarded	T0
		Change-6	Each subject applies the three techniques	T0
VV-UPV	VV-UPM	Change-1	A reading technique is omitted	T0
		Change-2	The duration of the sessions is limited	T6
		Change-3	The duration of training is reduced	T0
		Change-4	Change the order of application	T7
		Change-5	Changes in the application of techniques	T6
		Change-6	Changes in the application of test cases	T6
VV-Uds	VV-UPM	Change-1	The duration of the sessions is limited	T6
		Change-2	Change at time of test case execution	T7
		Change-3	The subjects work in pairs	T6
		Change-4	The duration of training is reduced	T0
		Change-5	Change the order of application	T5
VV-ORT	VV-UPM	Change-1	A reading technique is omitted	T0
		Change-2	A program is discarded	T0
		Change-3	There is only one session	T6
		Change-4	Changes in the application of techniques to programs.	T6
		Change-5	No test cases are executed	T6

Cuadro 6. Analysis of the changes

included about the original study and about replication is indicated. It is recommended to include a section on changes to the original experiment but without specifying how to report such changes.

Apart from Carver's methodological proposal, several studies have been found that document changes in replication using *ad-hoc* tables. Among them we highlight the following:

Solari [29] refers to identifying experimental incidents that occur during replication. Examples of incidents are unexpected changes in any of the variables defined for the experiment. In the study, five replications of the same experiment carried out by different experimenters in different locations are documented. Incidents are classified by replication and incident category. With the information stored, a summary table is presented that facilitates the analysis and comparison of the replications.

The work of Lung et al. [24], is the most related to our proposal. A replication of an experiment with human subjects is documented by presenting a summary table of differences and changes made to adapt the procedure to local circumstances. For each change introduced, the situation in the original experiment, the situation in replication and the reasons for making the change are clearly explained.

In [1], Almqvist presents a Systematic Literature Review about Replications. It includes a template to collect the configuration of the replications and in one of its rows it jointly describes all the changes made in the replication.

In [2], Apa et al. present a replication of a failure detection experiment. The changes are described, in a section on changes in the original experiment, in textual form and in a table where for each change the situation in the original experiment and in replication is compared.

In [14], Fucci et al. document an external replication of an experiment on evidence-based development. It contains a section of changes to the original configuration where it presents a table with the adjustments made to the base experiment.

In [17], Juristo et al. document eight replications of an experiment to evaluate the effectiveness of three verification and validation techniques. A table is presented where for each change, its situation in each replication and the design decision adopted is analyzed.

In [25], Quesada et al. describe a controlled experiment on function point analysis and its two replications. The differences in the experimental configuration between the experiment and replications are explained using a table and in textual form.

On the other hand, there are the studies on template definition in other areas within SI. The idea of using templates and patterns to define changes in the replication of experiments is based on the templates proposed by Durán et al. [13] for the elicitation of requirements. The templates and L-patterns have been successfully applied in several areas; Del Rio et al. [12] have used them for the definition of business process performance indicators and Segura et al. [26] have used them for the definition of metamorphic relationships.

In the description of experiments, the Goal-Question-Metric (GQM) template by Basili [5], recommended by Wohlin [33], should be highlighted for the definition of experiment objectives. Finally, in the Design Science Research (DSR) methodology, Wieringa [32] defines a template for the specification of a problem that will lead to future research.

Cuadro 7. Comparación de trabajos relacionados

Related works	**Topic**	**Empirical SI**	**Use template**	**Use pattern**
Carver [10]	Guidelines	✓	✗	✗
Solari [29]	Reports a replication	✓	✗	✗
Lung et al. [24]	Reports a replication	✓	✗	✗
Almqvist [1]	SLR replications	✓	✗	✗
Apa et al. [2]	Reports a replication	✓	✗	✗
Fucci et al. [14]	Reports a replication	✓	✗	✗
Juristo et al. [17]	Reports a replication	✓	✗	✗
Quesada et al. [25]	Reports a replication	✓	✗	✗
Durán et al. [13]	Elicitation of requirements	✗	✓	✓
Del Río et al. [12]	Performance indicators	✗	✓	✓
Segura et al. [26]	Metamorphic relationships	✗	✗	✓
Basili [5]	GQM	✓	✗	✓
Wieringa [32]	DSR	✗	✗	✓
This work	Templates for changes	✓	✓	✓

The table 7 analyses, for each of the above-mentioned works: i) the subject matter, ii) their belonging to the Empirical SI area, iii) the use of templates to facilitate reuse and visual presentation, and iv) the use of patterns to facilitate writing.

Referencias

1. Johan Per Fredrik Almqvist. Replication of controlled experiments in empirical software engineering - a survey, 2006. Master's thesis, Department of Computer Science, Faculty of Science, Lund University, Sweden, 2006.
2. Cecilia Apa, Oscar Dieste, et al. Effectiveness for detecting faults within and outside the scope of testing techniques: an independent replication. *Empirical Software Engineering*, 19(2):378–417, 2014.
3. Alejandrina Aranda López King. *Estudio empírico de la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educción de requisitos*. PhD thesis, ETSI_Informatica, 2016.
4. Maria Teresa Baldassarre, Jeffrey Carver, Oscar Dieste, and Natalia Juristo. Replication types: Towards a shared taxonomy. In *Proceedings of EASE '14*, pages 18:1–18:4, 2014.
5. Victor R Basili, Gianluigi Caldiera, and H Dieter Rombach. Goal question metrics paradigm. *Encyclopedia of Software Engineering*, pages 528–53, 1994.

6. Victor R Basili, Forrest Shull, and Filippo Lanubile. Building knowledge through families of experiments. *IEEE Transactions on Software Engineering*, 25(4):456–473, 1999.
7. Beatriz Bernárdez, Amador Durán, José A. Parejo, and Antonio Ruiz-–Cortés. An experimental replication on the effect of the practice of mindfulness in conceptual modeling performance. *Journal of Systems and Software*, 136:153–172, 2018.
8. Fernando Blanco, José César Perales López, and Miguel A Vadillo. Puede la psicología rescatarse a sí misma? incentivos, sesgos y replicabilidad. 2017.
9. Andrew Brooks, John Daly, James Miller, Marc Roper, and Murray Wood. Replication of experimental results in software engineering. *Technical Report ISERN-96-10, University of Strathclyde*, 1996.
10. Jeffrey C Carver. Towards reporting guidelines for experimental replications: a proposal. In *1st International Workshop on Replication in Empirical Software Engineering*, 2010.
11. Fabio QB Da Silva, Marcos Suassuna, A César C França, Alicia M Grubb, Tatiana B Gouveia, Cleviton VF Monteiro, and Igor Ebrahim dos Santos. Replication of empirical studies in software engineering research: a systematic mapping study. *Empirical Software Engineering*, 19(3):501–557, 2014.
12. Adela del Río-Ortega, Manuel Resinas, Amador Durán, and Ruiz-Cortés Antonio. Defining process performance indicators by using templates and patterns. In *Proceedings of BPM'12*, pages 223–228. Springer, 2012.
13. Amador Durán Toro, Beatriz Bernárdez Jiménez, Antonio Ruiz Cortés, and Miguel Toro Bonilla. A requirements elicitation approach based in templates and patterns. In *Proceedings of WER'99*, pages 17–29, 1999.
14. Davide Fucci, Giuseppe Scanniello, Simone Romano, Martin Shepperd, Boyce Sigweni, Fernando Uyaguari, Burak Turhan, Natalia Juristo, and Markku Oivo. An external replication on the effects of test-driven development using a multi-site blind analysis approach. In *Proceedings of EMSE'16*. ACM, 2016.
15. Omar S Gómez, Natalia Juristo, and Sira Vegas. Replications types in experimental disciplines. In *Proceedings of ESEM'10*. ACM, 2010.
16. Omar S Gómez, Natalia Juristo, and Sira Vegas. Understanding replication of experiments in software engineering: A classification. *Information and Software Technology*, 56(8):1033–1048, 2014.
17. N Juristo, Sira Vegas, Martín Solari, Silvia Abrahao, and Isabel Ramos. Comparing the effectiveness of equivalence partitioning, branch testing and code reading by stepwise abstraction applied by subjects. In *2012 IEEE Fifth International Conference on Software Testing, Verification and Validation*, pages 330–339. IEEE, 2012.
18. Natalia Juristo and OmarS. Gómez. Replication of software engineering experiments. In Bertrand Meyer and Martin Nordio, editors, *Empirical Software Engineering and Verification*, volume 7007 of *Lecture Notes in Computer Science*, pages 60–88. Springer Berlin Heidelberg, 2012.
19. Natalia Juristo and Sira Vegas. Functional testing, structural testing and code reading: What fault type do they each detect? In *Empirical Methods and Studies in Software Engineering*, pages 208–232. Springer, 2003.
20. Natalia Juristo and Sira Vegas. The role of non-exact replications in software engineering experiments. *Empirical Software Engineering*, 16(3):295–324, 2011.
21. Natalia Juristo, Sira Vegas, Martín Solari, Silvia Abrahão, and Isabel Ramos. A process for managing interaction between experimenters to get useful similar replications. *Information and Software Technology*, 55(2):215–225, 2013.

22. Barbara Kitchenham. The role of replications in empirical software engineering—a word of warning. *Empirical Software Engineering*, 13(2):219–221, 2008.
23. R Murray Lindsay and Andrew SC Ehrenberg. The design of replicated studies. *The American Statistician*, 47(3):217–228, 1993.
24. Jonathan Lung, Jorge Aranda, Steve Easterbrook, and Gregory Wilson. On the difficulty of replicating human subjects studies in software engineering. In *Proceedings of ICSE'08*, pages 191–200. IEEE, 2008.
25. Christian Quesada-López, Denisse Madrigal, and Marcelo Jenkins. An empirical evaluation of automated function points. In *Ibero-American Conference on Software Engineering*, pages 151–165, 2016.
26. Sergio Segura, Amador Durán, Javier Troya, and Antonio Ruiz Cortés. A template-based approach to describing metamorphic relations. In *Metamorphic Testing (MET), 2017 IEEE/ACM 2nd International Workshop on*, pages 3–9. IEEE, 2017.
27. Forrest Shull, Victor Basili, Jeffrey Carver, José Carlos Maldonado, Guilherme Horta Travassos, Manoel Mendonça, and Sandra Fabbri. Replicating software engineering experiments: addressing the tacit knowledge problem. In *Proceedings of EMSE'2002*, pages 7–16. IEEE, 2002.
28. Forrest J Shull, Jeffrey C Carver, Sira Vegas, and Natalia Juristo. The role of replications in empirical software engineering. *Empirical Software Engineering*, 13(2):211–218, 2008.
29. Martin Solari. Identifying experimental incidents in software engineering replications. In *Proceedings of EMSE'2013*, pages 213–222. IEEE, 2013.
30. Martín Solari, Sira Vegas, and Natalia Juristo. Content and structure of laboratory packages for software engineering experiments. *Information and Software Technology*, 2017. In Press.
31. Amador Durán Toro and Beatriz Bernárdez Jiménez. Metodología para la elicitación de requisitos de sistemas software. *Informe Técnico LSI-2000-10. Facultad de Informática y Estadística Universidad de Sevilla*, 2000.
32. Roelf J. Wieringa. *Design science methodology for information systems and software engineering*. Springer, 2014.
33. Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in software engineering: an introduction*. Springer Publishing Company, Incorporated, 2012.