

UNA PROPUESTA PARA REPORTAR REPLICACIONES DE EXPERIMENTOS EN INGENIERÍA DEL SOFTWARE

M^a MARGARITA CRUZ RISCO

Tesis Doctoral

**Directores: Dra. Beatriz Bernárdez Jiménez y
Dr. Amador Durán Toro**



Universidad de Sevilla

junio 2020

First published in junio 2020 by

M^a Margarita Cruz Risco

Copyright © MMXX

<http://www.lsi.us.es/~cruz>

cruz@us.es

This is a copyleft document but the content is copyrighted

Esta tesis doctoral ha sido parcialmente financiada por la Comisión Europea (FEDER) y por los proyectos de investigación BELI (TIN2015-70560-R), TAPAS (TIN2012-32273) y OPHELIA (RTI2018-101204-B-C22) del Plan Nacional y COPAS (TIC-1867), THEOS (P10-TIC-5906) y EKIPMENT-PLUS del Plan Andaluz de Investigación

Yo, D. M^a Margarita Cruz Risco con NIF número 28678173W,

DECLARO

mi autoría del trabajo que se presenta en la memoria de esta tesis doctoral que tiene por título:

Una propuesta para reportar replicaciones de experimentos en Ingeniería del Software

Lo cual firmo,

Fdo. D. M^a Margarita Cruz Risco
en la Universidad de Sevilla
01/06/2020

Dra. Beatriz Bernárdez Jiménez, Profesor Contratado Doctor del Área de Lenguajes y Sistemas Informáticos de la Universidad de Sevilla y Dr. Amador Durán Toro, Profesor Titular del Área de Lenguajes y Sistemas Informáticos de la Universidad de Sevilla,

HACEN CONSTAR

que D. M^a Margarita Cruz Risco, Profesor Titular de Escuela Universitaria del Área de Lenguajes y Sistemas Informáticos de la Universidad de Sevilla, ha realizado bajo nuestra supervisión el trabajo titulado

Una propuesta para reportar replicaciones de experimentos en Ingeniería del Software

Una vez revisado, autorizamos el comienzo de los trámites para su presentación como Tesis Doctoral al tribunal que ha de juzgarlo.

Fdo. Dra. Beatriz Bernárdez Jiménez y Dr. Amador Durán Toro
en la Universidad de Sevilla
01/06/2020



AGRADECIMIENTOS

Durante la realización de este trabajo, he recibido orientaciones y he contado con la experiencia y conocimientos de personas a las que agradezco su colaboración.

A mis directores de tesis. A Bea, por despertar en mi el interés por la experimentación, desde aquellas guías de Carver que me recomendó leer, y por estar siempre disponible para responder a mis preguntas. A Amador, por su visión para abordar problemas y aportar soluciones brillantes, especialmente en la parte técnica. Gracias a los dos por haber hecho posible la realización de esta tesis, poniendo a mi disposición vuestros conocimientos.

A Antonio Ruiz por confiar en mí y por sus valiosas sugerencias, y al resto de mis compañeros del grupo ISA, los nombraría a todos, pero en especial a Octavio, y sin olvidar a Bedilia por facilitarme el trabajo de documentación.

Al resto de mis compañeros de departamento, en especial a Carlos Arévalo por sus consejos. A Jose Galindo por ayudarme siempre que lo he necesitado, y a Luisa Romero por el día a día.

Destacar mi agradecimiento a mi buena amiga Carmen Florido, por mostrarme su trabajo y ponerme en contacto con Aránzazu García y Rocío Abia (CSIC), permitiéndome acceder a su investigación. Gracias a las tres por la implicación en la parte de Ciencias.

A mi familia, en especial a mi marido Pepe y a mis hijos Alberto y Arturo, por apoyarme en la dedicación a este trabajo y hacer mas llevadero el confinamiento forzoso que nos ha tocado vivir, coincidiendo con la elaboración de la parte final de esta memoria. Gracias por vuestra paciencia.

RESUMEN

En esta memoria, se presenta el trabajo de investigación llevado a cabo para aumentar el beneficio que se obtiene mediante la replicación de experimentos, desarrollando una propuesta que facilite la especificación sistemática de los cambios de las replications y su documentación de forma homogénea.

La importancia de especificar bien los cambios en las replications viene dada por un doble motivo; por un lado, por el propio autor de la replicación; para tener una guía de cómo reportarla. Parte de los cambios que se propongan permitirán mitigar algunas limitaciones identificadas en el experimento base, o bien son consecuencia de adaptar el experimento al nuevo entorno y es necesario analizar su influencia en la validez del experimento. Por otro lado, a un experimentador *externo* que realiza una replicación, le será de gran ayuda conocer cómo ha evolucionado un experimento original dentro de una *familia*, por qué se han realizado cambios y de qué tipo son. De este modo, se facilita el diseño de experimentos, sin repetir fallos ya identificados y adaptar el experimento al nuevo entorno con más éxito.

Partiendo de la información identificada en la descripción del cambio, se ha propuesto un metamodelo visualizado mediante una plantilla que presenta los datos de forma fija y sirve de guía para evitar que falte información relevante. La plantilla se completa con *patrones lingüísticos* que facilitan la redacción de los cambios.

Basada en la plantilla propuesta, se presenta la herramienta CÆSAR que permite a los experimentadores la definición de los cambios, que pasan a formar parte del repositorio de información, proporcionando una visión global de la *familia de experimentos*.

Para validar la plantilla, se ha instanciado mediante un *estudio de caso múltiple* que abarca las áreas de *Ingeniería del Software*, *Ciencias y Experimentos automáticos*. La utilización comparativa de la plantilla nos ha permitido establecer diferencias entre las áreas debidas, a la distinta terminología y conceptos utilizados al llevar a cabo un experimento.



ABSTRACT

In this dissertation, we report on our contribution to increase the benefit obtained through the replication of experiments, developing a proposal that facilitates the systematic specification of replication changes and their documentation in a homogeneous way.

The importance of specifying replication changes is twofold; on the one hand, so that the author has a guide on how to report the replication. Some of the proposed changes will mitigate some limitations identified in the base experiment or are a consequence of adapting the experiment to the new environment and it is necessary to analyze their influence on the validity of the experiment. On the other hand, an *external* experimenter who carries out a replication needs to know the evolution of the original experiment within a *family*, as well as the type and reason for the changes. This facilitates the design of experiments without repeating failures already identified and adapting the experiment to the new environment successfully.

Based on the information identified in the change description, a metamodel has been proposed which is displayed using a template that shows the data in a fixed form and serves as a guide to avoid the loss of relevant information. The template is completed with *linguistic patterns* that facilitate the writing of changes.

In order to facilitate the definition of changes for experimenters, the CÆSAR tool based on the proposed template is presented. The changes are part of the information repository providing an overview of the *family of experiments*.

To validate the template, it has been instantiated by means of a *multi-case study* that covers the areas of *Software Engineering*, *Science* and *Automatic Experiments*. The comparative use of the template has allowed us to establish differences between the areas due to the different terminology and concepts used when carrying out an experiment.

ÍNDICE GENERAL

I	Introducción	1
1.	Planteamiento del problema	3
1.1.	Motivación	4
1.2.	Relevancia del problema	5
1.3.	Objetivos de la investigación	9
1.4.	Resumen de la Contribución	11
1.5.	Contexto de la Tesis	15
1.6.	Estructura del documento	16
2.	Background	19
2.1.	Introducción	20
2.2.	Terminología	20
2.2.1.	Familia de experimentos	20
2.2.2.	Réplica vs. replicación	21
2.2.3.	Repetición, Replicación y Reproducción	22
2.2.4.	Sujetos y objetos experimentales	23
2.2.5.	Variables	23
2.2.6.	Diseño de bloques	24
2.2.7.	Dimensiones de la configuración experimental	25

2.3.	Amenazas a la validez en experimentación	27
2.3.1.	Determinación de algunas amenazas a la validez	28
2.3.2.	Relación entre validez interna y validez externa	29
2.4.	Tipología de las replicaciones	30
2.4.1.	Clasificación según Shull <i>et al.</i>	30
2.4.2.	Clasificación según Basili <i>et al.</i>	30
2.4.3.	Clasificación según Almqvist	31
2.4.4.	Clasificación según Gómez <i>et al.</i>	31
2.4.5.	Clasificación según Baldassarre <i>et al.</i>	32
2.5.	Resumen	34
II	Estado de la cuestión	37
3.	Estado actual de las replicaciones de experimentos en IS	39
3.1.	Introducción	40
3.2.	¿Qué foros se utilizan para publicar replicaciones?	40
3.3.	¿Quiénes son los autores e instituciones que publican replicaciones?	45
3.4.	¿Cuáles son los estudios más citados que publican replicaciones?	49
3.5.	¿Qué temas de investigación han sido más/ menos replicados?	50
3.6.	¿Qué métodos empíricos son los más utilizados en las replicaciones?	51
3.7.	¿Cómo evoluciona el número de estudios que publican replicaciones?	52
3.8.	¿Cómo se distribuyen los estudios?	53
3.9.	Conclusiones	53
3.10.	Resumen	55
4.	Problemática en la realización de las replicaciones	57

- 4.1. Introducción 58
- 4.2. Falta de guías para reportar replicaciones 59
- 4.3. Carencia de paquetes de laboratorio 60
- 4.4. El conocimiento tácito 61
- 4.5. Diferente terminología 62
- 4.6. Necesidad de clarificar el contexto de las replicaciones 63
- 4.7. El factor humano en las replicaciones 63
- 4.8. Dificultad para realizar y publicar replicaciones 64
- 4.9. Discusión 65
- 4.10. Resumen 66

- 5. Especificación de los cambios en replicaciones 69**
- 5.1. Introducción 70
- 5.2. Revisión de los estudios de replicación 70
- 5.3. Influencia del tipo de replicación en la especificación de los cambios . . . 75
- 5.4. Discusión y hallazgos 77
- 5.5. Resumen 79

- III Propuesta 81**

- 6. Especificación de los cambios mediante plantillas y patrones 83**
- 6.1. Introducción 84
- 6.2. Metamodelo de replicaciones y cambios 84
- 6.3. Plantilla para especificar los cambios 86
- 6.3.1. Especificación de los aspectos generales de las replicaciones . . . 87
- 6.3.2. Especificación del propósito de la replicación 89

6.3.3.	Especificación de los cambios de las replications	91
6.4.	Aplicabilidad de la plantilla	93
6.5.	Una herramienta para definir los cambios	96
6.5.1.	Vista de estudios empíricos	96
6.5.2.	Vista de cambios en las replications	97
6.5.3.	Vista de amenazas a la validez	97
6.6.	Resumen	97
IV	Validación	101
7.	Evaluación de la propuesta en un estudio de caso múltiple	103
7.1.	Introducción	104
7.2.	Experimentación en Ciencias	104
7.2.1.	Principios básicos	104
7.2.2.	Diferencias con la experimentación en IS	107
7.3.	Estudio de caso múltiple	108
7.3.1.	Diseño del estudio de caso	109
7.3.2.	Planificación, recopilación y análisis	111
7.3.3.	Análisis e informe conjunto	112
7.4.	Estudio de caso: Caso-IngSoft	113
7.4.1.	Descripción de las familias	114
7.4.2.	Especificación de los aspectos generales de las replications	115
7.4.3.	Especificación de los cambios de las replications	117
7.4.4.	Discusión	120
7.5.	Estudio de caso: Caso-Ciencias	123

- 7.5.1. Descripción de las familias 123
- 7.5.2. Especificación de los aspectos generales de las replicaciones . . . 126
- 7.5.3. Especificación de los cambios de las replicaciones 127
- 7.5.4. Discusión 128
- 7.6. Estudio de caso: Caso-Automáticos 130
 - 7.6.1. Descripción de las familias 130
 - 7.6.2. Especificación de los aspectos generales de las replicaciones . . . 131
 - 7.6.3. Especificación de los cambios de las replicaciones 133
 - 7.6.4. Discusión 133
- 7.7. Resultados del estudio de caso múltiple 133
 - 7.7.1. Expresividad 134
 - 7.7.2. Precisión 136
 - 7.7.3. Usabilidad 136
 - 7.7.4. Trazabilidad 138
- 7.8. Conclusiones 140
- 7.9. Resumen 141

- V Conclusiones 143**

- 8. Conclusiones y trabajos futuros 145**
 - 8.1. Consecución de objetivos 146
 - 8.2. Publicaciones 148
 - 8.2.1. Revistas internacionales 149
 - 8.2.2. Capítulo de libro 149
 - 8.2.3. Congresos nacionales 150

8.3. Trabajo futuro	151
VI Anexos	153
A. Instanciación de la versión inicial de la plantilla	155
B. Instanciación de la versión final de la plantilla	167
Bibliografía	191

ÍNDICE DE FIGURAS

1.1. Marcas para artículos de investigación en publicaciones ACM	7
1.2. Resumen de contribuciones	11
3.1. Mapa de coautores con al menos cuatro estudios de replicación en 2013–18	47
3.2. Mapa de coautores con al menos un estudio de replicación en 2013–18 .	48
3.3. Número de estudios de replicación publicados en 2013–2018 agrupados por método empírico	52
3.4. Clasificación de estudios de replicación publicados en 2013–2018 según el método empírico y área de conocimiento	53
6.1. Metamodelo de replications y cambios	86
6.2. Plantilla para la especificación de cambios en replications	88
6.3. Correspondencia metamodelo plantilla	89
6.4. Vista de estudios empíricos	97
6.5. Vista de uno de los cambios junto con sus dos amenazas a la validez . .	98
6.6. Vista de amenazas a la validez	98
7.1. Diseño con 6 tratamientos en 3 bloques	107
7.2. Fases del estudio de caso múltiple basado en [197]	109
7.3. Versión inicial y final de la vista de amenaza a la validez	118

ÍNDICE DE TABLAS

1.1. Actividades de la metodología DSR y secciones correspondientes	14
2.1. Comparación de las clasificaciones presentadas	34
3.1. Revistas en las que se han publicado estudios de replicación en 2013–2018 ordenados por número de publicaciones	41
3.2. Revistas en las que se han publicado estudios de replicación en 2013–2018 junto con el método empírico y área de conocimiento	43
3.3. Conferencias en las que se han publicado estudios de replicación en 2013–2018 ordenadas por número de publicaciones	44
3.4. Conferencias en las que se ha publicado más de un estudio de replicación en 2013–2018 junto con el método empírico y área de conocimiento	45
3.5. Autores con al menos cuatro estudios de replicación publicados en 2013–2018 junto con el método empírico y área de conocimiento	46
3.6. Instituciones con al menos cuatro estudios de replicación publicados en 2013–2018 ordenados por número de publicaciones	49
3.7. Estudios más citados que reportan al menos una replicación en 2013–2018 ordenados por el número de citas según SCOPUS a mayo de 2019	49
3.8. Áreas de investigación y estudios primarios publicados en 2013–2018 ordenados por número de estudios	51
3.9. Métodos empíricos y estudios primarios publicados en 2013–2018 ordenados por número de estudios	52
5.1. Estudios de replicación publicados en 2013–2018 en EMSE y ESEM con sus características	72

5.2. Estudio de replicación publicados en 2013–2018 en EMSE y ESEM agrupados por tipo	75
6.1. Instanciación de la plantilla en la replicación Mind#2	94
7.1. Especificación de un cambio que disminuye la validez (familia <i>Req</i> , replicación Q-2009)	117
7.2. Especificación de los aspectos básicos y de los cambios (familia <i>Req</i> , replicación E-2012B)	119
7.3. Hallazgos utilizando la versión inicial de la plantilla	121
7.4. Resumen de cambios por familia	122
7.5. Instituciones y familias de experimentos en Caso-Ciencias	123
7.6. Especificación de los primeros 5 cambios de la replicación <i>Suelo-2018</i> (familia <i>Suelo</i>) utilizando la versión final de la plantilla	125
7.7. Hallazgos al especificar los cambios en Caso-Ciencias	127
7.8. Replicación SPL–Pr&Com, versión final	131
7.9. Comparación de la nomenclatura utilizada en el <i>Caso-IngSoft</i> , <i>Caso-Ciencias</i> y <i>Caso-Automáticos</i>	132
7.10. Hallazgos al especificar los cambios en Caso-Automáticos	133
7.11. Comparación de la plantilla con las guías seguidas para reportar replications a nivel de expresividad de las propuestas	135
7.12. Uso comparativo de la plantilla	137
7.13. Singularidades de los campos de la plantilla en el Caso-Ciencias	139
7.14. Comparación de la plantilla con las propuestas seguidas para reportar los cambios	140
A.1. Replicación Mind#2	155
A.2. Replicación Mind#3	155
A.3. Replicación Q-2009	156

A.4. Replicación Q-2011	157
A.5. Replicación Q-2012	158
A.6. Replicación E-2012A	159
A.7. Replicación E-2012B	161
A.8. Replicación E-2013	161
A.9. Replicación E-2014	162
A.10. Replicación E-2015	162
A.11. Replicación VV-UPM1	163
A.12. Replicación VV-UPV	164
A.13. Replicación VV-Uds	165
A.14. Replicación VV-ORT	166
B.1. Replicación Mind#2, versión final	168
B.2. Replicación Mind#3, versión final	169
B.3. Replicación Q-2009, versión final	170
B.4. Replicación Q-2011, versión final	171
B.5. Replicación Q-2012, versión final	172
B.6. Replicación E-2012A, versión final	173
B.7. Replicación E-2012B, versión final	175
B.8. Replicación E-2013, versión final	176
B.9. Replicación E-2014, versión final	177
B.10. Replicación E-2015, versión final	177
B.11. Replicación VV-UPM1, versión final	178
B.12. Replicación VV-UPV, versión final	179
B.13. Replicación VV-Uds, versión final	180

B.14. Replicación VV-ORT, versión final	181
B.15. Replicación Suelo–2018, versión final	182
B.16. Replicación Suelo–2019, versión final	184
B.17. Replicación Calidad–2017P, versión final	185
B.18. Replicación Calidad–2017V, versión final	186
B.19. Replicación Oliva–Des, versión final	187
B.20. Replicación Dieta–Hiper, versión final	188
B.21. Replicación SPL–Pr&Com, versión final	189
B.22. Replicación Test–NF, versión final	189
B.23. Replicación Test–F&NF, versión final	190
B.24. Replicación Test–FvsNF, versión final	190

PARTE I

INTRODUCCIÓN

PLANTEAMIENTO DEL PROBLEMA

El principio es la mitad del todo

Pitágoras de Samos (569 a. C.– 475 a. C.),

Filósofo y matemático griego

En este primer capítulo, se presenta el problema que se aborda en la presente memoria justificando su importancia y las razones que han motivado su investigación.

En primer lugar, en la sección §1.1, se expone la motivación y planteamiento del problema. En la sección §1.2 se identifican algunas iniciativas que destacan la importancia en la disponibilidad de los artefactos experimentales para facilitar las replicaciones junto con autores que han resaltado la necesidad de documentar sus cambios. En la sección §1.3 se presenta el objetivo principal de este trabajo de investigación y los objetivos parciales que contribuyen a alcanzar dicho objetivo principal. En la sección §1.4 se describe el método de investigación seguido y se resumen las aportaciones surgidas como consecuencia de los objetivos planteados. En la sección §1.5 se enumeran los proyectos de investigación en los que se enmarca el presente trabajo. Finalmente, en la sección §1.6 se describe la estructura del resto de la memoria.

1.1 MOTIVACIÓN

En *Ingeniería del Software (IS)*, y específicamente en el desarrollo software, es frecuente que surjan técnicas y herramientas para mejorar el proceso software y, en estos casos, la experiencia o la teoría existente no son suficientes para discernir las prácticas más adecuadas.

La IS empírica permite la evaluación de estos nuevos métodos, técnicas y herramientas para conocer la conveniencia de usarlos en el proceso de desarrollo [21]. Su evaluación comparativa en un entorno controlado permite ir variando ciertas condiciones y analizar su influencia en los resultados esperados. Una vez evaluado empíricamente, el experimento necesita ser replicado en diferentes contextos y condiciones no solo para consolidar el conocimiento adquirido, sino también para saber si sus resultados son generalizables [19]. En [145] se destaca la importancia de las replications no solo para confirmar resultados sino también para establecer el rango de condiciones en las que se cumplen los resultados.

El motivo que nos lleva a investigar en los cambios introducidos en las replications respecto al experimento original y, en concreto, en cómo se reportan, surge como consecuencia de un experimento y su posterior replicación llevados a cabo en nuestro grupo de investigación por investigadores con experiencia en *Ingeniería del Software Empírica (ISE)*. La duda surge a la hora de reportar la replicación ya que existen directrices para reportar el experimento original [118, 125, 138, 196, 245] sin embargo, no está claro cómo reportar su replicación. Únicamente Carver [42] presenta una propuesta inicial de directrices sobre el contenido de las publicaciones que reportan replications.

En este caso, el experimento original o experimento base, ya está publicado en el momento en que se quiere publicar su replicación. La primera pregunta que se nos plantea es qué contar del experimento original ya publicado. Por otro lado, en el caso de que se publiquen en el mismo artículo dicho experimento original y su replicación tampoco está claro la estructura del documento. Al año siguiente, se realiza una segunda replicación y el problema de qué y cómo reportar la replicación surge de nuevo. En cualquier caso, los investigadores implicados quieren destacar las diferencias o cambios al experimento original ya que afecta a las conclusiones a extraer y al diseño de replications posteriores.

La segunda pregunta que se nos plantea es, por tanto, cómo contar cada cambio introducido en la replicación. Puede que sea suficiente su especificación en lengua-

je natural pero pensamos que una tabla ayuda a su descripción y salva, en parte, el inconveniente en la ambigüedad del lenguaje natural. En cualquier caso, además de explicar la situación antes y después del cambio, se debe de justificar el motivo de llevar a cabo cada cambio ya sea, entre otros, impuesto por el medio o para analizar la influencia de alguna nueva *variable*.

Documentar los cambios es útil tanto para el experimentador que realiza el experimento, ya que le permite dejar constancia de los ajustes o cambios realizados, como para el investigador que necesita entender el experimento, conocer los cambios que se han hecho en cada replicación y seguir la traza de sus cambios en las sucesivas replicaciones. Esto es especialmente importante para un investigador que quiere hacer una replicación externa: conocer en detalle el trabajo anterior es esencial para sobreponerse a algunos fallos o adaptar el experimento a otro entorno.

Pensamos que la mejora en la definición de los cambios aumenta la comprensión de la replicación y, en concreto, de sus cambios y puede contribuir a aumentar el número de replicaciones.

1.2 RELEVANCIA DEL PROBLEMA

Para justificar la importancia de las replicaciones, a continuación, se presentan algunas organizaciones y específicamente investigadores que han publicado o se han interesado por el tema de las replicaciones y sus aspectos relacionados.

En [1], la *sociedad científica y educativa sobre computación ACM* (Association for Computing Machinery) incluye entre sus "Directrices y procedimientos de publicación", un apartado dedicado a la "Revisión e identificación del *artefacto*" donde propone clasificar las publicaciones en *artefactos evaluados*, *artefactos disponibles* y *resultados validados*.

Un *artefacto* u *objeto digital* es software, *scripts* para ejecutar experimentos y/o analizar resultados o en general datos, creado para ser usado como parte de un estudio o generado por el experimento. Tal como se comenta en ICSE'2020 (International Conference on Software Engineering), la revisión de los *artefactos* de los trabajos aceptados aumenta la probabilidad de que los resultados puedan ser replicados y reproducidos por otros investigadores.

Según la clasificación de la ACM, los *artefactos evaluados* pueden ser *funcionales* o *reutilizables* y, a su vez, los *resultados validados* pueden ser *replicados* o *reproducidos*.

De esta forma quedaría:

- *Artefactos evaluados*. Son los que han superado con éxito en una auditoría independiente. Se distinguen dos niveles:
 - *Funcionales*. Los artefactos están documentados, son consistentes, completos, aplicables e incluyen pruebas apropiadas de verificación y validación.
 - *Reutilizables*. Además de ser funcionales, están cuidadosamente documentados y bien estructurados facilitando la reutilización.
- *Artefactos disponibles*. Si además de ser funcionales son de acceso público y por tanto están *disponibles*. Se proporciona un DOI o enlace a este repositorio junto con un identificador único para el objeto.
- *Resultados validados*. Se distinguen dos niveles:
 - *Replicados*. Experimentadores distintos del equipo original han sido capaces de obtener los mismos resultados del trabajo utilizando los artefactos proporcionados por el autor.
 - *Reproducidos*. Experimentadores distintos del equipo original y sin utilizar los artefactos del autor, han sido capaces de obtener los mismos resultados.

Para el tema de las replications interesan los *artefactos disponibles* y *resultados validados*. En concreto, la especificación completa de los cambios podría formar parte del *paquete de laboratorio* junto con el conjunto de datos de entrada y su análisis, constituyendo un *artefacto* que podría ser candidato a *artefacto replicado*.

ACM propone añadir *marcas* a las publicaciones indicando si el estudio (o artefacto) es *funcional*, *reutilizable*, *disponible replicado* o *reproducido*. En la figura §1.1 aparecen las marcas correspondientes.

ACM destaca que un resultado experimental no está plenamente establecido a menos que pueda ser *reproducido* de forma independiente, señalando además, los beneficios obtenidos al poner los *artefactos de la investigación* a disposición del público para facilitar las replications que verifiquen la solidez de los resultados originales.

En la misma línea, destaca la iniciativa *Open Science* de la revista EMSE (Empirical Software Engineering), un movimiento para hacer públicos los datos de las investigaciones aumentando así la transparencia y la reproducibilidad de los estudios [78]. Está

Functional	Reusable	Available	Replicated	Reproduced
No Badge				
Artifacts documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation	Functional + very carefully documented and well-structured to the extent that reuse and repurposing is facilitated. In particular, norms and standards of the research community for artifacts of this type are strictly adhered to.	Functional + placed on a publicly accessible archival repository. A DOI or link to this repository along with a unique identifier for the object is provided.	Available + main results of the paper have been obtained in a subsequent study by a person or team other than the authors, using, in part, artifacts provided by the author.	Available + the main results of the paper have been independently obtained in a subsequent study by a person or team other than the authors, without the use of author-supplied artifacts.

Figura 1.1: Marcas para artículos de investigación en publicaciones ACM

basada en los siguientes términos: *i) Open Access*: acceso libre a los artículos, *ii) Open Data*: datos disponibles y *iii) Open Source Software*: software de código abierto. Los paquetes de replicación que contienen los datos en bruto y todo el material utilizado son un ejemplo de compartición de datos para facilitar las replicaciones que están dentro del apartado ii).

Desde ESEM'2018 (International Symposium on Empirical Software Engineering and Measurement), se sigue la política *Open Science* y se anima a los autores a la presentación de estudios de replicación.

Desde ICSE'2019 se invita a los autores de contribuciones aceptadas a presentar los artefactos asociados para su evaluación y recibir una (y sólo una) de las marcas, indicando si es *reutilizable*, *disponible replicado* o *reproducido* (ver figura §1.1). Asimismo, los autores pueden realizar cortas presentaciones en el festival ROSE (Recognizing and Rewarding Open Science in Software Engineering). Los festivales ROSE se han celebrado en todas las grandes conferencias¹ recientes en IS tales como ICSE'2019, FSE'2018, FSE'2019 (European Software Engineering conference and Symposium on the Foundations of Software) o ESEM'2019 y están previstas para FSE'2020 o ICSE'2020 entre

¹Debido a la crisis del COVID-19, la mayor parte de los congresos en 2020 se celebrarán de forma virtual

otras.

Además de las organizaciones presentadas anteriormente, ante la importancia de llevar a cabo replicaciones, diferentes investigadores destacan la importancia de especificar los cambios introducidos y disponer de directrices para reportar las replicaciones y específicamente sus cambios:

- En [59] Magalhães *et al.* estudian el estado de las replicaciones en IS destacando la necesidad de directrices y metodologías para apoyar el proceso de replicación.
- La trascendencia de documentar los cambios queda también patente en Gómez *et al.* [97] donde se aborda en profundidad las características de los distintos tipos de cambios que se pueden llevar a cabo en una replicación. La naturaleza del cambio determina la *dimensión* de la configuración experimental afectada, el propósito de la replicación y su nivel de verificación sobre los resultados originales.
- Brooks *et al.* en [34] señalan la necesidad de describir el proceso experimental como si fuera una receta (*chef's recipe*), proporcionando tantos detalles como sea posible, incluyendo la descripción de los detalles menores, para facilitar nuevas replicaciones.
- En [127, 128], Juristo y Vegas señalan que las replicaciones múltiples de un experimento aumentan la confianza en sus resultados y animan a los investigadores a introducir *cambios* en las replicaciones ya que permiten identificar nuevas variables que influyan en los resultados. Destacan, así mismo, la importancia de registrar dichos cambios y analizar su impacto.
- Baldassarre *et al.* [19] señalan que cuando un investigador planea una replicación es importante considerar aspectos como: quién está replicando el estudio, qué diferencias se están introduciendo y por qué se está replicando el estudio.
- Lindsay *et al.* [145] destacan que cuanto más explícita es la descripción de las variaciones entre replicaciones, más convincentes son los resultados.
- Dentro de la documentación a publicar sobre las replicaciones, las guías de Carver [42], las únicas específicas para reportar replicaciones, ponen de manifiesto la necesidad de describir los cambios introducidos con respecto al experimento original.

Respecto del número de cambios, hay autores que sugieren hacer pequeños cambios al experimento original para llegar a resultados concluyentes. Así por ejemplo, Vegas *et al.* en [240] señalan la conveniencia de evitar cambios no deseados en las replications ya que podrían dificultar la agregación de los resultados.

Por otro lado, Kitchenham [135], promueve la introducción de cambios en las replications para evitar propagar problemas del experimento original añadiendo que cuando se cambian sujetos, escenarios y materiales aumenta el poder de confirmación de los resultados.

A nuestro modo de ver, el hecho de que existan foros, algunos de ellos específicos para replications, donde se están publicando este tipo de trabajos revela la importancia del problema para la comunidad investigadora. Dichos foros, congresos y revistas, se identifican en la sección §3.2 donde se analiza el estado actual de las replications.

1.3 OBJETIVOS DE LA INVESTIGACIÓN

En el presente trabajo se analiza el proceso de replicación de experimentos controlados y concretamente sus cambios como base para proponer una plantilla que facilite la especificación de los cambios en los informes de replications.

El objetivo principal de esta tesis es **facilitar la especificación sistemática de los cambios para apoyar el proceso de replicación de experimentos controlados.**

Con este objetivo, se proponen las siguientes preguntas de investigación (PI):

- **PI₁**: ¿Cuál es el estado actual de las replications? (capítulo §3).
- **PI₂**: ¿Cuáles son los principales problemas identificados en la práctica de las replications? (capítulo §4).
- **PI₃**: ¿Cómo se están reportando actualmente los cambios? ¿Afecta el tipo de replicación (interna, externa) o la inclusión del experimento original en el informe de la replicación a la forma de definir los cambios? (capítulo §5).
- **PI₄**: ¿Se puede sistematizar la definición de los cambios? ¿Se pueden definir una plantilla con patrones lingüísticos para facilitar su definición y detectar la falta de información relevante? (capítulo §6).

- **PI₅**: ¿Se puede extender la propuesta a otras áreas tales como *Agrobiología y Tecnología de alimentos* y otros tipos de experimentos tales como *experimentos automáticos*? (capítulo §7).

A su vez, la pregunta de investigación **PI₁** se puede concretar en las siguientes preguntas:

- **PI_{1.1}**: ¿Qué foros se utilizan para publicar replications? (sección §3.2).
- **PI_{1.2}**: ¿Quiénes son los autores y las instituciones que publican replications? (sección §3.3).
- **PI_{1.3}**: ¿Cuáles son los estudios más citados que publican replications? (sección §3.4).
- **PI_{1.4}**: ¿Qué temas de investigación han sido más/menos replicados? (sección §3.5).
- **PI_{1.5}**: ¿Qué métodos empíricos son los más utilizados en las replications? (sección §3.6).
- **PI_{1.6}**: ¿Cómo evoluciona el número de estudios que publican replications? (sección §3.7).
- **PI_{1.7}**: ¿Cómo se distribuyen los estudios según los temas de investigación y los métodos empíricos? (sección §3.8).

Del mismo modo, la pregunta de investigación **PI₄** se puede concretar en los distintos aspectos de la plantilla que se van a analizar.

- **PI_{4.1}**: *Expresividad*. ¿Qué proporción de cambios ha sido posible definir? ¿La *expresividad* de la plantilla es mayor que la de otras propuestas seguidas? (sección §7.7.1).
- **PI_{4.2}**: *Precisión*. El uso de los patrones-L, ¿hace más precisa la definición de los cambios que el lenguaje natural y/o en forma de tablas? ¿Permite detectar la información que falta? (sección §7.7.2).
- **PI_{4.3}**: *Usabilidad*. ¿Cuántos campos de la plantilla han presentado problemas de comprensión? ¿Están los investigadores familiarizados con la terminología? (sección §7.7.3).

- PI_{4.4}: Trazabilidad. Al registrar los cambios, ¿se facilita la trazabilidad entre replications? (sección §7.7.4).

1.4 RESUMEN DE LA CONTRIBUCIÓN

Para explicar las aportaciones de este trabajo que se pueden resumir mediante el diagrama mostrado en la figura §1.2, vamos a utilizar la metodología de la *Ciencia del diseño* (DSR, por sus siglas en inglés de *Design Science Research*) que ha sido la metodología seguida en la investigación.

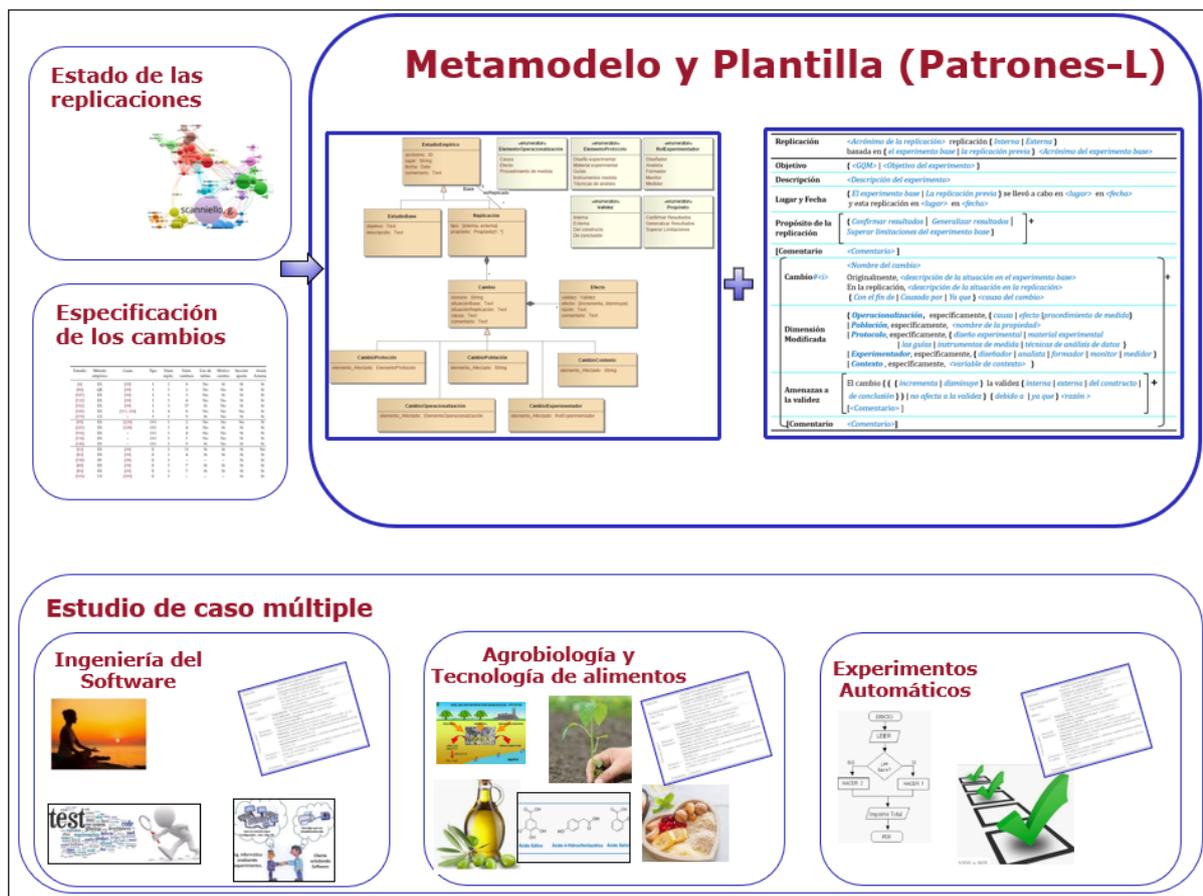


Figura 1.2: Resumen de contribuciones

DSR crea y evalúa artefactos para resolver los problemas organizacionales identificados [243]. En dicha metodología, se definen las siguientes actividades propuestas en [174]:

1. *Identificación del problema y motivación.* Se define el problema que da lugar a una

propuesta de investigación.

En nuestro caso, se analizan las dificultades encontradas al tratar de reportar la replicación y en concreto sus cambios debido a la falta de directrices. Se identifica la necesidad de abordar la definición de los cambios de una manera sistemática.

2. *Definición de los objetivos de la solución.* Es necesario conocer otras propuestas y especificar qué se espera del artefacto a desarrollar.

Para ello, se revisa la literatura para: *i)* analizar cómo se están reportando los cambios en el área de replications de IS; *ii)* identificar la información involucrada; y, *iii)* conocer otras propuestas.

Las contribuciones en esta fase son:

- *Análisis del estado de las replications.* Tras seleccionar y revisar estudios publicados entre 2013 y 2018 reportando replications, se han identificado los principales foros donde se publican replications, autores más prolíficos, instituciones y países más destacados en la publicación de replications. Se han utilizado *mapas de calor* para identificar los clústeres o grupos de coautores destacados. Así mismo, se han clasificado los estudios según las áreas temáticas abordadas y los métodos empíricos utilizados, identificando tanto las tendencias actuales así como los vacíos de investigación.
- *Revisión del informe de cambios.* Tomando como base los artículos identificados en la revisión anterior publicados en dos de los principales foros del área de IS empírica: *Empirical Software Engineering Journal (EMSE)* y *International Conference on Empirical Software Engineering and Measurement (ESEM)*, se ha analizado cómo se han reportado los cambios ya sea en forma de tablas o textual. Se ha identificado la información reportada y cuando la descripción es mediante tablas, se ha analizado la estructura de dichas tablas.

3. *Diseño y desarrollo.* Se da forma al artefacto.

El artefacto propuesto para formalizar la información sobre replications y cambios es el *metamodelo*. Una posible instanciación del *metamodelo* es la *plantilla* completada con patrones lingüísticos (*patrones-L*) que facilitan la redacción.

Las contribuciones en esta fase son:

- *Metamodelo de replications.* A partir de la información sobre las replications identificada en la revisión anterior, se ha formalizado el *metamodelo* que representa la información a gestionar y la relación entre los distintos conceptos

relevantes. Sirve como base para la definición de la *plantilla* sobre replicaciones y cambios.

- *Patrones lingüísticos*. Se han definido frases parametrizadas que proponen varias alternativas o texto libre para facilitar la cumplimentación de la *plantilla*.
- *Plantilla propuesta*. Es la contribución principal de la tesis. Ha pasado por varias versiones. Debido a las limitaciones encontradas en la versión inicial, se han ajustado los patrones lingüísticos, el *metamodelo* y la *plantilla* consecuentemente. Con la versión actual se han podido definir todos los cambios de los experimentos seleccionados.

4. *Demostración*. Implica el uso del artefacto desarrollado en una o más instancias del problema analizando su aplicabilidad.

Se instancia la primera versión de la *plantilla* en la replicación que ha dado origen a la propuesta y se constata su aplicabilidad. Para ello, se utiliza la herramienta CÆSAR que ha sido desarrollada para facilitar la definición de los cambios que forman parte de la replicación.

5. *Evaluación*. Mediante evidencias empíricas, se analiza si el uso del artefacto cubre los objetivos esperados. Puede ser necesario realizar ajustes en el artefacto y volver, mediante un proceso iterativo, a la actividad de *diseño y desarrollo*.

La contribución en esta fase es:

- *Validación de la plantilla*. Se ha evaluado mediante un *estudio de caso múltiple* que abarca: *i*) el área de conocimiento de IS (familias de experimentos de Mindfulness, análisis de requisitos y técnicas de evaluación del código); *ii*) las áreas de conocimiento de *Agrobiología y Tecnología de alimentos* (familias de experimentos sobre descontaminación de suelos, calidad del aceite de oliva virgen, biodegradación en suelos agrícolas, extracción de componentes del aceite de oliva virgen e influencia de la dieta en la acumulación de colesterol); y *iii*) otros tipos de experimentos como *experimentos de computación automáticos* (familias de experimentos sobre pruebas automatizadas de software y pruebas de Líneas de Productos software). Debido a las limitaciones encontradas, se han modificado el *metamodelo* y la *plantilla* y se han ajustado los *patrones-L*.
- *Validación del metamodelo*. El uso de la herramienta CÆSAR y la realimentación que ha generado han permitido mejorar el metamodelo. Tras especifi-

carse los cambios de la replicación de descontaminación de suelos, CÆSAR ha sido valorada de forma muy positiva por la autora de dicha replicación, mostrando su satisfacción con la herramienta.

6. *Comunicación*. Se publican los resultados de la investigación.

La tabla §1.1 muestra estas actividades resumidas y su correspondencia con las secciones de esta memoria.

Tabla 1.1: Actividades de la metodología DSR y secciones correspondientes

Actividad	Descripción	Sección
<i>Identificación del problema y motivación</i>	Necesidad de abordar una adecuada definición de cambios para el informe de las replicaciones de experimentos controlados	§1.1, §1.2
<i>Definición de los objetivos de la solución</i>	Facilitar la documentación clara de los cambios de forma sistemática	§3, §5
<i>Diseño y desarrollo</i>	Se propone el <i>metamodelo</i> . Una posible implementación es la <i>plantilla</i> completada con <i>patrones-L</i>	§6.2, §6.3
<i>Demostración</i>	Se instancia la primera versión de la <i>plantilla</i> en las replicaciones que han dado origen a la propuesta mediante la herramienta CÆSAR	§7.4.1
<i>Evaluación</i>	Se evalúa el artefacto, es decir la <i>plantilla</i> , en un estudio de caso múltiple que abarca las áreas de conocimiento de IS y <i>Agrobiología y Tecnología de alimentos</i> y otros tipos de experimentos como <i>experimentos automáticos</i> . Se utiliza CÆSAR como prueba de concepto	§7.3, §7.4, §7.5, §7.6, §7.7
<i>Comunicación</i>	Se publican los resultados de la investigación	

1.5 CONTEXTO DE LA TESIS

Esta tesis ha sido desarrollada en el contexto del grupo de investigación de Ingeniería de Software Aplicado (ISA) de la Universidad de Sevilla.

El trabajo relacionado con esta tesis se encuadra en el ámbito de los siguientes proyectos de investigación:

- COPAS: *ECosystems for Optimized Process As a Service*. Proyecto de Excelencia de la Junta de Andalucía (P12-TIC-1867). Fecha de Inicio: 30-01-2014. Fecha de Finalización: 16-02-2019.

Específicamente, la participación en este proyecto ha estado relacionada con el paquete de trabajo 1: *Posibilitar el desarrollo de sistemas de información basados en PRaaS (Process as a Service)*, que consiste en desarrollar un conjunto de técnicas y metodologías que, de forma automática o semiautomática, apoyen el desarrollo de sistemas de información basados en PRaaS.

- BELI: Tecnologías para Servicios Cloud Híbridos, Altamente Configurables y Regulados por acuerdo de nivel de servicio (SLA, por sus siglas en inglés de *Service Level Agreement*). Proyecto del Ministerio Economía y Competitividad (TIN2015-70560-R). Fecha de Inicio: 01-01-2016. Fecha de Finalización: 31-12-2019.

He participado en los paquetes de trabajo 3 y 5: *Pruebas de servicios* y *Casos de uso de servicios* en concreto, en el desarrollo de técnicas y herramientas para la generación automática de casos de prueba en servicios web preferiblemente regulados por SLA.

- THEOS: Tecnologías Habilitadoras para EcOsistemas Software. Proyecto de excelencia financiado por la Junta de Andalucía (P10-TIC-5906). Fecha de Inicio: 15-03-2011. Fecha de Finalización: 30-04-2016.

He participado en la integración de las tareas de análisis y del modelado de procesos de negocio en *entornos de gestión del ciclo de vida de la aplicación*.

- TAPAS: Tecnologías Avanzadas para Procesos como Servicios. Proyecto del Ministerio de Economía y Competitividad (TIN2012-32273), financiado por el Gobierno de España. Fecha de Inicio: 01-01-2013. Fecha de Finalización: 31-12-2015.

He participado en el paquete de trabajo 2: *Sistema de gestión de conformidad para la auditoría de procesos de negocio* diseñando un almacén de evidencias para la auditoría escalable.

- OPHELIA: *Optimización de servicios basados en conocimiento usando aplicaciones basadas en servicios*. Proyecto del Ministerio de Ciencia, Innovación y Universidades (RTI2018-101204-B-C22). Fecha de Inicio: 01-01-2019. Fecha de Finalización: 31-12-2021. Participación en el paquete de trabajo 5: *Validación empírica*.
- EKIPMENT-PLUS: *Mejora del rendimiento de procesos basados en conocimiento: Un enfoque empírico multidisciplinar basado en personas, equipos, software y datos*. Proyecto del Plan Andaluz de I+D+i, financiado por la Junta de Andalucía. Fecha de inicio: 01-01-2020. Fecha de finalización: 31-12-2023.

1.6 ESTRUCTURA DEL DOCUMENTO

La organización de la memoria es la siguiente:

Parte I. Introducción. Consta de dos capítulos; en el presente capítulo §1, se plantea la motivación y relevancia del problema y se encuadra la investigación. En el siguiente capítulo, §2, se presentan aspectos generales sobre la replicación de experimentos y se explican los conceptos y terminología utilizados en el desarrollo de la propuesta.

Parte II. Estado de la cuestión. Se distinguen tres capítulos. En el capítulo §3 se muestra el estado actual de las replications identificando, entre otros, las áreas temáticas de investigación más abordadas y los métodos empíricos utilizados. En el capítulo §4 se presentan los problemas detectados al abordar la replicación de experimentos. En el capítulo §5 se analiza cómo se están especificando los cambios en los informes de replicación y se identifican otras propuestas.

Parte III. Propuesta. Es el núcleo de la tesis. Incluye el capítulo §6 en el que se define el *metamodelo* y la *plantilla* que estructura la información involucrada en la replicación. Basada en la *plantilla* propuesta, se presenta la herramienta CÆSAR.

Parte IV. Validación. En el capítulo §7, mediante un estudio de caso múltiple se evalúa la *plantilla* en las áreas de *Ingeniería del Software* y *Agrobiología y Tecnología de alimentos* y en experimentos de tipo *automático*.

Parte V. Conclusiones. En el capítulo §8 se analiza el grado de cumplimiento de los objetivos propuestos, se resumen las publicaciones derivadas de este trabajo, y se indican las principales líneas para el trabajo futuro.

Parte VI. Anexos. El anexo §A contiene la instanciación de la versión inicial de la *plantilla* en las familias de experimentos del primer estudio de caso. La instanciación final de la *plantilla* en las familias de experimentos del estudio de caso múltiple se muestra en el anexo §B.

BACKGROUND

El aprendizaje es hijo de la repetición

*Robin Sharma (1965 –),
Profesor, escritor y orador nepalés*

En este capítulo se analizan aspectos generales de las repeticiones de experimentos en Ingeniería del Software tales como su tipología y haciendo especial hincapié en los conceptos y terminología utilizados a lo largo de la memoria y específicamente en la propuesta.

En la sección §2.1 se destaca la falta de acuerdo en la terminología a utilizar. En la sección §2.2 se presentan los conceptos principales, ilustrando la explicación mediante un escenario de experimentación. En la sección §2.3 se analizan las amenazas a la validez del experimento. Las principales clasificaciones de las repeticiones se presentan y comparan en la sección §2.4 y en la sección §2.5 se resume el capítulo.

2.1 INTRODUCCIÓN

Dada la falta de acuerdo existente entre la terminología en experimentación y en replicaciones en particular, en este capítulo se comenta la forma en qué se va a usar la terminología en el resto de la memoria. Para ello, se presenta una breve introducción sobre replicaciones sin necesidad de análisis detallado, sino con objeto de aclarar los términos que presentan más inconsistencias según la bibliografía consultada y abordando los principales conceptos utilizados en experimentación, que se interpretan de forma distinta según los autores.

Para los términos y principales conceptos tratados, a continuación, se presentan las definiciones de los principales autores y, cuando es posible, se ilustran con un ejemplo. Cuando varios conceptos son equivalentes, se determina el término a utilizar en la presente memoria.

2.2 TERMINOLOGÍA

En esta sección, se abordan los principales conceptos y terminología sobre experimentos controlados y en consecuencia para sus replicaciones.

Para ilustrar la explicación de algunos de los conceptos, se utiliza un escenario de experimentación simplificado (*Exp-Código*) conocido por haber instanciado la plantilla que se presenta en la sección §6.3, en algunas de sus replicaciones. El objetivo de *Exp-Código* es la evaluación de la efectividad de tres técnicas de verificación del código fuente por parte de alumnos de máster. Para ello, cada alumno aplica cada técnica sobre el código de un programas y se contabilizan la cantidad de errores encontrados por cada una de las técnicas.

2.2.1 Familia de experimentos

Basili [22] introduce el concepto de *familia de experimentos* como un marco que permite organizar los experimentos relacionados, facilitando las replicaciones y el trabajo entre experimentadores. Las conclusiones que se obtienen en una *familia de experimentos* refuerzan y amplían las obtenidas en experimentos aislados.

En [201] se define una *familia de experimentos* como grupo de replicaciones con al menos tres experimentos que evalúan los efectos de al menos dos tecnologías diferentes

sobre la misma variable de respuesta y donde se conocen los ajustes en los experimentos y se puede acceder a los datos antes de ser procesados.

En [50] aparece el término *familias de experimentos concertada* que se caracteriza por un marco común para todos los estudios que la componen. Incluyendo variables y normas comunes para el diseño y recopilación de datos de todos los experimentos aunque se examinan distintos contextos. Los datos de los estudios pueden combinarse fácilmente y además de validar los resultados, se analizan las variaciones entre los diferentes experimentos.

2.2.2 Réplica vs. replicación

Ante la duda de utilizar en la presente tesis el término *réplica* o bien *replicación*, se ha buscado la expresión utilizada en el principal foro nacional de referencia para los investigadores en Ingeniería del Software y Bases de Datos (SISTEDES). Desde la edición de 2017, incluye un área de interés o “*track*” sobre *Métodos Empíricos en Ingeniería del Software y Sistemas de Información (MEISSI)*. Se han analizado los estudios publicados en español en dicho *track* y se ha encontrado que:

- En 2017, el estudio [47] utiliza el término *replicación*.
- En 2018, en la conferencia invitada a cargo de los profesores Marcela Genero y Mario Piattini sobre *Presente, pasado y futuro de la Ingeniería del Software Empírica* de la presentación del *track* MEISSI, se utilizó el término *réplica*.
- En 2019, los artículos [8, 101] utilizan el término *réplica*.

Se han analizado algunas tesis presentadas en la ETSI Informáticos de la Universidad Politécnica de Madrid: [15, 38, 91, 187, 237] y en todas se utiliza el término *replicación*. De forma análoga, en la tesis [11] de la Universidad de la República (Montevideo, Uruguay) se utiliza el término *replicación*.

Teniendo en cuenta lo anterior y que en el foro nacional se aceptan ambos términos al ser sinónimos, hemos optado por utilizar el término *replicación* ya que, además, es más parecido al término en inglés “*replication*”.

2.2.3 Repetición, Replicación y Reproducción

Hay autores que utilizan los términos *replicación* y *reproducción* como sinónimos para referirse a la repetición de un experimento [122], sin embargo, otros autores consideran que existen diferencias. Según [40, 96, 122], en una *replicación* se sigue “muy de cerca” el protocolo experimental utilizado en el experimento base mientras que en la *reproducción* el protocolo experimental es diferente. Es decir, en la *reproducción* se prueba la misma hipótesis que en el experimento base pero con un protocolo experimental diferente.

En la misma línea, como ya se ha comentado en §1.2, en [1], se definen los términos:

- *Repetibilidad*. Los mismos experimentadores y la misma configuración experimental.
- *Replicabilidad*. Diferentes experimentadores y la misma configuración experimental.
- *Reproducibilidad*. Diferentes experimentadores y diferente configuración experimental.

Otros términos relacionados que suelen aparecer y que conviene tener claros son:

- *Re-Análisis*. Consiste en utilizar los datos de un experimento ya ejecutado para verificar los resultados en lugar de volver a ejecutar el experimento. Se pueden utilizar nuevas técnicas de análisis o se repiten las técnicas ya utilizadas [95].
- *Meta-Análisis* o síntesis cuantitativa de experimentos. Consiste en combinar los resultados de varios experimentos para generar nuevo conocimiento [10].
- *Replicación Parcial*. Solo se replica una parte del experimento base [43]. Según [182], las replicaciones parciales tienen el mismo objetivo que el experimento base, pero se modifican el diseño, procedimiento experimental, etc. Según [109] en una replicación parcial se añaden o suprimen variables, se suprime alguna parte no esencial de los procedimientos o se llevan a cabo pequeños cambios para agilizar la recogida de datos.

La similitud entre una *replicación* y el experimento base depende de los cambios que se introducen en la *replicación*. Los cambios vienen determinados por el objetivo

que tenga la *replicación*. En la sección §2.2.7 se identifican algunos de los cambios que se pueden definir en una replicación.

2.2.4 Sujetos y objetos experimentales

Según Wohlin *et al.* [245], el *sujeto experimental* es la persona que participa en un estudio empírico para evaluar un *objeto experimental*. Los tratamientos se aplican a la combinación de *objetos y sujetos experimentales*.

Según Jedlitschka *et al.* [117], cuando los sujetos o en general participantes no son humanos, se denominan *unidades experimentales*.

Para ampliar nuestra propuesta a otras áreas, nos referiremos a *sujetos experimentales y unidades experimentales*.

Kitchenham [138], destaca la importancia de identificar la *población* de la que se seleccionan los *sujetos experimentales* y definir el proceso de selección seguido ya que para poder generalizar los resultados, la selección o *muestra* debe ser representativa de la *población* a estudiar.

Runeson y Höst en [196] señalan que en los *estudios de casos*, los fenómenos se estudian en su contexto natural y por tanto la selección de *sujetos* no se realiza sobre una muestra estadísticamente representativa.

- En *Exp-Código* los *sujetos experimentales* o *población* son los alumnos que evalúan las técnicas.

2.2.5 Variables

Las *variables independientes* son aquellas que podemos controlar y variar en el experimento [245]. Dicha variación en los valores de las *variables independientes* se refleja en los resultados cuantificados mediante la *variable dependiente*. La principal *variable independiente* es la que es objeto de estudio.

Las *variables independientes* deben tener varios *niveles*, al menos dos. A cada uno de los *niveles* también se le suele denominar *tratamiento*. Según Juristo y Moreno [124], los términos *tratamiento* y *niveles* provienen de los experimentos en agricultura donde tienen más sentido al hablar de tratamientos con fertilizantes y distintas concentraciones de reactivos. Según [138] la asignación de los *sujetos* a los *tratamientos* debe ser objetiva

para no comprometer la *validez* del experimento. Las *variables independientes*, también se denomina *factores*, *variables predictoras* o simplemente *predictoras*. Los *niveles* también se denominan *alternativas* [124]. Según [76], las *variables independientes* son *factores* que tienen influencia en los resultados.

La *variable dependiente*, también se denomina *variable de respuesta* (“*outcome variable*”) [124, 245]. Representa las métricas definidas para medir el efecto.

- En *Exp-Código* la *variable independiente* son las técnicas de verificación del código y los *niveles* son cada una de las tres técnicas.
- En *Exp-Código* la *variable dependiente* se puede cuantificar contabilizando la cantidad de errores encontrados por cada una de las técnicas. En este caso, la cantidad de errores encontrados es la *variable dependiente*.

A las características que rodean al experimento, y pueden influir pero se van a controlar mediante el diseño experimental, se les denomina *variables de contexto* o *parámetros*. Según [245] son *variables independientes* que se mantienen a un nivel fijo durante el experimento. Según [124] un *parámetro* es cualquier característica invariable durante el experimento.

- En *Exp-Código* la edad y experiencia son ejemplos de *variables de contexto*.

La identificación de los *parámetros* es importante en la especificación de un experimento y sus replicaciones no sólo para contextualizar al lector/experimentador, sino también, para ver en que circunstancias se produce el resultado y saber si un cambio podría arrojar distintos resultados.

En las sucesivas replicaciones de una familia, puede ser interesante variar dichos *parámetros* y analizar su efecto en la *variable dependiente* [124].

2.2.6 Diseño de bloques

Según [124], para controlar las variaciones en algunas de las características de un experimento controlado que pueden afectar a los resultados del experimento, se definen las *variables de bloqueo* que requieren un tipo especial de diseño experimental llamado *diseño de bloques*.

El *diseño de bloques* evita la influencia indeseada de algún *factor*. Así por ejemplo, si se quiere analizar el efecto de un *factor* A en la variable respuesta y se sabe que otro factor B también puede influir en la variable respuesta, es necesario anular su influencia mediante el equilibrio del factor B en todas los niveles del factor A (es decir, se reparte por igual). En el caso de que el factor A tenga dos niveles y el factor B otros 2 niveles, resulta un diseño 2X2 bloques [124].

De esta forma, en caso de que en *Exp-Código* la experiencia influya en los resultados, se puede bloquear dividiendo a los alumnos en dos grupos según tengan o no experiencia en desarrollo software. En este caso el factor que se quiere analizar, es decir, la técnicas de verificación del código tiene 3 niveles (cada una de las tres técnicas) y el factor cuya influencia se quiere evitar tiene 2 niveles (con y sin experiencia), lo que da lugar a un diseño de 3x2 bloques. Así, los grupos son homogéneos y comparables. Del experimento, se podría discernir, por ejemplo, la mejor técnica para alumnos con experiencia.

Otra posible opción es asignar los sujetos a los grupos de forma aleatoria para que los grupos estén equilibrados [245].

2.2.7 Dimensiones de la configuración experimental

Los *sujetos experimentales*, el diseño o los tipos de *variables* descritos anteriormente son, entre otros, *elementos* que forman parte de la *configuración experimental*.

Gómez *et al.* [97], clasifican los *elementos de la configuración experimental* en grupos llamados *dimensiones*:

- *Operacionalización*. Incluye las *variables independientes*. y *dependientes*. La *operacionalización* es la selección de dichas variables para reflejar la *causa* y el *efecto* a medir. La *causa* viene representado por la *variable independiente* que puede tener varios *niveles* y el *efecto* por las *variables dependientes*. Los elementos que se pueden distinguir en la *operacionalización* son la *causa*, el *efecto* y los *procedimientos de medida*.
- *Población*. Son los *sujetos* o *unidades experimentales*. En [97] se incluye además, a los *objetos experimentales*, sin embargo en [201] sólo incluye en la *dimensión población* a los *sujetos experimentales*.
- *Protocolo*. Los elementos que forman parte del *protocolo experimental* son el *diseño*

experimental, *material experimental*, *guías* o instrucciones proporcionadas a los sujetos, los *instrumentos de medición* y las *técnicas de análisis de datos*. El *protocolo experimental* es la disposición de estos elementos para observar los efectos de los *tratamientos* [123].

Nos hemos planteado añadir las *tareas* como un elemento más, susceptible de cambios, dentro del *protocolo*. No lo hemos hecho por seguir los *elementos* identificadas por Gómez *et al.* [97] para la *dimensión protocolo*.

- *Experimentadores*. Esta dimensión incluye a los investigadores que pueden participar en el experimento en sus diferentes roles: *diseñador*, *analista*, *formador*, *monitor* o *medidor*.
- *Contexto*. Se ha añadido una quinta dimensión (que no aparece en la propuesta de Gómez *et al.* [97]). El contexto es el entorno donde se realiza el experimento. Si se quiere estudiar si alguna característica del entorno afecta a los resultados, se define mediante la *variable de contexto* (parámetro)

En el metamodelo y plantilla propuestos en las secciones §6.2 y §6.3, se asocia cada cambio que se produce en una replicación con la *dimensión experimental* afectada y se identifica el *elemento de la configuración experimental* que cambia.

En el *estudio de caso múltiple* presentado en la sección §7, al cumplimentar la plantilla, se ha observado que uno de los campos más costosos de cumplimentar es la *dimensión afectada* por el cambio. Para ayudar a su comprensión, a continuación se presenta un posible cambio en *Exp-Código* para cada *elemento* de cada *dimensión experimental*.

- Cambios en la *dimensión operacionalización*:

En la *causa*, implica cambiar la forma de aplicar el tratamiento o en general una *variable independiente*. Un ejemplo de cambio puede consistir en aumentar las sesiones en que se aplica el tratamiento.

En el *efecto*, implica cambiar las métricas. Un posible cambio es medir el tiempo empleado en encontrar tres errores en vez de tiempo empleado en encontrar un error de cada tipo.

En el *procedimiento de medida*. Un cambio en el procedimiento es, por ejemplo, medir el tiempo con un cronómetro manual en vez de por un programa de forma automática.

- Cambios en la *dimensión población*:

Es un cambio de alguna de las propiedades de los *sujetos experimentales*. Así por ejemplo, al realizar el experimento en profesionales en lugar de alumnos de máster, cambia la propiedad *experiencia*.

- Cambios en la *dimensión protocolo*:

En el *diseño experimental*, es por ejemplo, hacer aleatoria la asignación de sujetos al tratamiento.

En *material experimental*, cambiar algún formulario de recopilación de datos.

En *guías*, cambiar las instrucciones proporcionadas.

En *instrumentos de medición*, cambiar alguno de los programas sobre los que se evalúan las técnicas de verificación del código.

En las *técnicas de análisis de datos*, cambiar el análisis estadístico utilizado.

- Cambios en la *dimensión experimentador*:

Se cambia alguna de las personas (o bien aumenta o disminuye el número de personas) que realiza alguno de los roles (*monitor, diseñador, analista, formador o medidor*).

- Cambios en la *dimensión contexto*:

Un cambio en una *variable de contexto*, es por ejemplo, cambiar el momento de la realización del experimento. De esta forma, se podría realizar el experimento durante la época de exámenes en lugar de al comienzo del curso. Es importante dejar constancia porque puede afectar a los resultados.

2.3 AMENAZAS A LA VALIDEZ EN EXPERIMENTACIÓN

Las *amenazas a la validez* son el conjunto de situaciones o debilidades que presenta un determinado estudio empírico y que podrían poner en tela de juicio los hallazgos del mismo. Las *amenazas a la validez* identificadas por Wohlin [245] son:

Validez de la conclusión: Esta relacionada con la capacidad del análisis estadístico de sacar conclusiones correctas sobre la relación entre el tratamiento y el resultado de un experimento.

Puede verse afectada cuando no se cumplan los requisitos para usar una prueba estadística determinada o la potencia estadística de una prueba aplicada es baja.

Validez interna: Se refiere a la relación causal entre tratamiento y resultado. Es decir entre las variables independientes y dependientes. Se presenta amenazas a la validez interna cuando el resultado viene influido por algún factor no deseado, es decir, el resultado no se debe solo al tratamiento, que es la operacionalización de la causa.

Validez del constructo: Está relacionado con la capacidad de las variables independientes y dependientes para reflejar el constructo que se va a estudiar. Es decir, que el tratamiento, métricas y procedimientos de medida reflejen realmente los constructos de causa y de efecto.

Validez externa: Está relacionado con la capacidad de poder generalizar los resultados. Algunos ejemplos de amenazas a la validez externa son el hecho de escoger un grupo de sujetos experimentales que no es representativo de la población que se quiere estudiar o un entorno experimental que no es adecuado.

Respecto a la *validez externa*, las replicaciones demuestran que los resultados experimentales no dependen de las condiciones específicas del experimento base y, respecto a la *validez interna*, permiten conocer el rango de condiciones bajo las cuales se mantienen los resultados experimentales (Shull *et al.* [218]).

Existen otros tipos de amenazas, aunque menos utilizadas en Ingeniería del Software. En [163] se refiere a *validez ecológica* relacionada con la similitud entre el entorno experimental y la realidad investigada. Otro tipo de amenaza identificada en [164] aunque de menor importancia desde una perspectiva científica, es *face validity* o *validez aparente* referida al grado en que una medida mide lo que se quiere medir según el observador. De esta forma, importa más la apariencia de validez que presenta un estudio que su validez real, es decir el estudio *debe de parecer válido* a los demás.

2.3.1 Determinación de algunas amenazas a la validez

En este apartado, a partir de [97, 245], se presentan una serie de pautas que de forma orientativa, pueden servir de ayuda a la hora de determinar como afecta un cambio a las *amenazas a la validez* del experimento:

- Cambios en la *dimensión operacionalización*. Suelen estar relacionados con la *validez del constructo* y probablemente también con la *validez interna*.
- Cambios en la *dimensión población*. La *validez externa* aumenta cuando se generalizan los resultados para la nueva población.
- Cambios en el *dimensión protocolo*. Afectan a la *validez interna*. Si tras el cambio, se confirman los resultados, se demuestra que éstos son independientes del protocolo seguido y aumenta la *validez interna*.
- Cambios en la *dimensión experimentador*. Si se confirman los resultados, demuestra que éstos son independientes del experimentador y aumenta la *validez interna*.
- Cambios en la *dimensión contexto*. Si influye en los resultados, puede disminuir la *validez interna*. También puede afectar a otros tipos de amenazas.

En cualquier caso, el *efecto* (aumento o disminución) a la *validez del experimento* depende de cada cambio concreto. Asimismo, puede haber cambios que no afecten a la *validez*, no esté claro el tipo de *amenaza* afectada, ni el efecto sobre la *validez del experimento* e incluso, es complicado identificar la *dimensión* de la configuración experimental afectada.

2.3.2 Relación entre validez interna y validez externa

En [220] se analiza la relación entre *validez interna* y *validez externa*. Cuando los investigadores se centran en la *validez interna* intentan asegurar que el resultado sólo es causado por el tratamiento recibido. Si se centran en la *validez externa*, analizan el efecto en el mundo real, pero sin saber qué factores causan realmente el resultado observado. De esta forma, cuando aumenta la *validez interna*, la *validez externa* disminuye y al contrario. Si se quiere generalizar los resultados para ganar en *validez externa*, se puede cambiar el entorno y utilizar participantes con distinto nivel de conocimiento, en ese caso, puede que haya nuevas variables que afectan a los resultados y representa una amenaza a la *validez interna*. Por otro lado, si en la replicación se mantienen, tanto como sea posible, las condiciones originales puede representar una amenaza a la *validez externa* ya que aunque se asegura que no hay nuevas variables que afecten a los resultados, no es generalizable.

Los tipos de amenazas a la *validez interna* y *validez externa* determinan si el experimento tiene validez. No hay consenso en la comunidad sobre cuándo centrarse en la

validez interna o en la *validez externa* (Siegmund *et al.* [220]). En dicho estudio, se incluye una revisión sistemática de la literatura llevada a cabo en 2015 en los principales foros del área concluyendo que sólo el 54% de los estudios presentados analizan las amenazas a la validez y sólo el 23% identifica los diferentes tipos de amenazas.

2.4 TIPOLOGÍA DE LAS REPLICACIONES

Una primera clasificación de las replications en *replicaciones internas* y *replicaciones externas*, según sean llevadas a cabo por los mismos investigadores del experimento base o investigadores distintos, es debida a Brooks *et al.* [35].

Otro posible criterio de clasificación es según el grado en que se siga el procedimiento el experimento base. En este sentido, existen distintas clasificaciones como por ejemplo la clasificación en *Cerradas (Closed)* y *Diferenciadas (Differentiated)* de Lindsay *et al.* [145].

A continuación, aparecen algunas de las principales clasificaciones identificadas.

2.4.1 Clasificación según Shull *et al.*

Según Shull *et al.* [218], se pueden clasificar las replications en *Exactas (Exact)* y *Conceptuales (Conceptual)* y a su vez, estas últimas, se pueden clasificar en *Dependientes (Dependent)* e *Independientes (Independent)*:

- Replicaciones *Exactas*. Siguen los procedimientos de un experimento base sin variaciones.
- Replicaciones *Conceptuales*. Se mantiene la hipótesis y se varía el procedimiento experimental.
 - Replicaciones *Dependientes*. Se mantienen las condiciones similares al experimento base.
 - Replicaciones *Independientes*. Se varían las condiciones del experimento.

2.4.2 Clasificación según Basili *et al.*

Basili *et al.* [22] clasifica las replications según las siguientes categorías:

- Replicaciones que *no varían ninguna hipótesis de investigación*. Se dividen en:
 - Replicaciones Estrictas (*Strict*): intentan duplicar el experimento base.
 - Replicaciones que varían la forma de llevar a cabo el experimento para abordar ciertas *amenazas a la validez*.
- Replicaciones que *varían las hipótesis de investigación*. Se dividen en:
 - Replicaciones que varían las variables *dependientes* e/o *independientes*.
 - Replicaciones que varían variables en el entorno.
- Replicaciones para *extender la teoría*. Determinan los límites en los que se mantienen los resultados.

2.4.3 Clasificación según Almqvist

Teniendo en cuenta el grado en que se sigue el experimento base y la clasificación en replications *internas* y *externas* propuesta por Brooks *et al.* [35], Almqvist [9] define los siguientes tipos de replicación:

- *Similar-externa (Similar-external)*. El mismo experimento llevado a cabo por otros investigadores.
- *Mejorada-interna (Improved-internal)*. Con variaciones en el experimento y con los mismos investigadores.
- *Similar-interna (Similar-internal)*. El mismo experimento y los mismos investigadores
- *Diferenciada-externa (Differentiated-external)*. Con variaciones en el experimento y llevado a cabo por otros investigadores.

2.4.4 Clasificación según Gómez *et al.*

Gómez *et al.* [97] comparan diferentes tipos de replications en IS y en otras disciplinas.

Identifican las *dimensiones* de la configuración experimental que se ven afectadas cuando se lleva a cabo algún cambio en una replicación: *operacionalización, población,*

protocolo y experimentadores. Estas dimensiones se analizan en la sección §2.2 y son las que se utilizan en la sección §6.3 como parte de la plantilla propuesta.

A partir de las *dimensiones* modificadas establecen los siguiente tres tipos de replications:

- *Literal*. El objetivo es llevar a cabo una replicación lo más exacta posible al experimento base.
- *Operacional*. El objetivo es variar algunas (o todas) las *dimensiones* de la configuración experimental.
- *Conceptual*. Se cambia el *protocolo* y *operacionalización* con diferentes *experimentadores* para definir la replicación. Solo se conoce un *hecho empírico* establecido por un experimentador con anterioridad.

2.4.5 Clasificación según Baldassarre *et al.*

A partir de las encuestas y debates llevadas a cabo en el taller celebrado en ISERN¹ 2013, Baldassarre *et al.* [19] proponen una clasificación de las replications que consolide la terminología utilizada.

La clasificación tiene en cuenta: *i)* el procedimiento o pasos seguidos en el estudio, se suele analiza el grado en que se sigue el experimento base; *ii)* los experimentadores que llevan a cabo la replicación y *iii)* la combinación de ambos factores.

Teniendo en cuenta estos factores y que los términos más utilizados para describir las replications son *Interna*, *Externa*, *Cerrada*, *Diferenciada* y *Conceptual*, clasifica las replications en:

- *Interna*. Se mantienen los mismos (o la mayoría) de los experimentadores del experimento base.
- *Externa*. Los experimentadores son distintos a los del experimento base.
- *Cerrada*. Se mantienen, tanto como sea posible, las condiciones del experimento base. Un ejemplo de una replicación de tipo *Cerrada* se reporta en [140]. El experimento se replicó en cuatro universidades mediante una replicación distribuida basada en instancias estrechamente coordinadas (“*replicación conjunta*”).

¹International Software Engineering Research Network - ISERN

- *Diferenciada*. Se añaden cambios de forma intencionada.
- *Conceptual*. Solo se mantiene la hipótesis. Un ejemplo de un estudio en el que se reporta una replicación *conceptual* es [110]. La replicación sobre predicción de defectos está basado en varios experimentos previos y compara los resultados con los estudios anteriores.

En [59], Magalhães *et al.* comparan diferentes clasificaciones concluyendo que cualquier intento de establecer una tipología de replicaciones debe hacerse con cuidado ya que, tal como dicen Juristo *et al.* [96], hay autores que designan con el mismo término a diferentes tipos de replicaciones y al contrario; usan diferentes términos para referirse al mismo tipo de replicación. Además, expresiones como “*con la mayor precisión posible*” o “*tanto como sea posible*” hacen que no queden suficientemente claras algunas definiciones.

Con ánimo de aclararlo, se presenta la tabla §2.1 que muestra cada tipo de replicación con el nombre que le han dado en las distintas taxonomías recogidas.

Puede observarse que las replicaciones sin cambios, es decir, las replicaciones en las que el experimento base se duplica con la mayor precisión posible, se denominan *Exactas*, *Cerradas*, *Literales* o *Estrictas* según los autores. El término utilizado es distinto y sin embargo, el significado es equivalente.

La distinción entre replicaciones *Internas* y *Externas* propuesto por Brooks *et al.* [35] coincide con los utilizados por Almqvist [9] y Baldassarre *et al.* [19] tanto en el significado como en el término utilizado.

El término *Conceptual* es similar, tanto Shull *et al.* [218] como en Baldassarre *et al.* [19], utilizado para replicaciones en las que solo se mantiene la hipótesis. Gómez *et al.* [97] se refieren a replicaciones llevadas a cabo con diferentes experimentadores con nuevos protocolos y operacionalización.

El término *Diferenciada* es utilizado por Baldassarre *et al.* [19] para replicaciones en las que se cambia el diseño, hipótesis, contexto o medidas del experimento base. De forma equivalente, Lindsay *et al.* [145] se refiere a variaciones deliberadas en aspectos importantes de las condiciones del estudio. Por su parte Gómez *et al.* [97] denomina *Operacional* a este tipo de replicaciones y las relaciona con cambios en alguna de las dimensiones de la configuración experimental. Los términos *Mejorada* y *Diferenciada* utilizados por Almqvist [9] tienen un significado similar. Por último, las replicaciones según la clasificación de Basili *et al.* [22] cuyos objetivos son *abordar alguna amenazas* o

Tabla 2.1: Comparación de las clasificaciones presentadas

Definición	Brooks [35]	Shull [218]	Lindsay [145]	Basili [22]	Almqvist [9]	Gómez [97]	Baldassarre [19]
Sin cambios		Cerrada	<i>Exacta</i>	<i>Estricta</i>		<i>Literal</i>	<i>Cerrada</i>
Abordan amenazas				Abordan amenazas			
Varía hipótesis				Varía hipótesis			
Extender teoría				Extender teoría			
= experimen. ≠ investig.					<i>Similar-Externa</i>		
≠ experimen. = investig.					<i>Mejorada-Interna</i>		
= experimen. = investig.					<i>Similar-Interna</i>		
≠ experimen. ≠ investig.					<i>Diferenciada-externa</i>		
Variações importantes			Diferenciada ¹			<i>Operacional</i> ²	<i>Diferenciada</i>
Sólo mantiene hipótesis		<i>Conceptual</i>				<i>Conceptual</i> ³	<i>Conceptual</i>
= condiciones exp. base		<i>Dependiente</i>					
≠ condiciones exp. base		<i>Independiente</i>					
= investig.	<i>Interna</i>						<i>Interna</i>
≠ investig.	<i>Externa</i>						<i>Externa</i>

¹Implica variaciones de las condiciones del estudio

²Implica cambio en alguna de las dimensiones

³Solo se establece un hecho empírico

extender la teoría estarían también aquí incluidas.

2.5 RESUMEN

En este capítulo, se han presentado los principales conceptos sobre experimentos controlados utilizados en esta memoria, destacando la terminología relacionada con las replicaciones y eligiendo la terminología que se va a utilizar, justificando los motivos en caso de que haya varias alternativas.

Se ha utilizado un escenario de experimentación para ilustrar los conceptos de: *i*) sujetos experimentales; *ii*) variables dependientes, independientes, de contexto y de bloqueo; *iii*) *dimensión* experimental y *iv*) *amenazas a la validez*.

Asimismo, se han presentado algunas pautas para inferir las *amenazas a la validez* a

partir de la *dimensión experimental* afectada por cada cambio definido en la replicación.

Para completar la introducción a las replicaciones, se ha revisado la tipología de las replicaciones presentando algunas de las principales clasificaciones destacando la de Basili *et al.* [22].

PARTE II

ESTADO DE LA CUESTIÓN

ESTADO ACTUAL DE LAS REPLICACIONES DE EXPERIMENTOS EN IS

No hay nada como mirar, si queréis encontrar algo

J. R. R. Tolkien (1892–1973),

Escritor, poeta, filólogo y profesor británico

En este capítulo, se responde a cada una de las preguntas de investigación que componen la pregunta PI_1 sobre el estado actual de las replications. Asimismo, se identifican vacíos de investigación tanto en temas de investigación como en métodos empíricos utilizados por los experimentadores que pueden contribuir a fomentar nuevas replications.

En la sección §3.1 se presentan las preguntas de investigación. En la sección §3.2 se contesta a la pregunta sobre revistas y congresos donde se han publicado replications. En la sección §3.3 se identifican los autores más prolíficos en el tema de replications y las instituciones a las que pertenecen. Se acompaña de los mapas de calor de coautorías. En la sección §3.4 se presentan los estudios más citados que reportan alguna replication. En las secciones §3.5 y §3.6 se identifican respectivamente las áreas de investigación y los métodos empíricos más utilizados para llevar a cabo replications. En la sección §3.7 se presenta la evolución anual en el número de replications. En la sección §3.8 se analiza la distribución del número de replications según áreas de investigación y métodos empíricos. La sección §3.9 presenta las conclusiones. Por último, la sección §3.10 resume el capítulo.

3.1 INTRODUCCIÓN

Antes de aterrizar en cómo se especifican los cambios, se ha llevado a cabo una revisión bibliográfica sobre las replications en general, cuyo estudio completo aparece publicado en [51].

Se han consultado los repositorios SCOPUS¹ y *Web of Science*² (WoS) y se han seleccionado 137 estudios que reportan al menos una replicación y que han sido publicados entre 2013 y 2018. Sólo se han considerado los estudios publicados desde 2013 ya que hay revisiones sobre replications que cubren años anteriores [29, 57, 59].

La pregunta de investigación **PI₁**: ¿Cuál es el estado actual de las replications? se ha concretado en las siguientes preguntas:

- **PI_{1.1}**: ¿Qué foros se utilizan para publicar replications?
- **PI_{1.2}**: ¿Quiénes son los autores y las instituciones que publican replications?
- **PI_{1.3}**: ¿Cuáles son los estudios más citados que publican replications?
- **PI_{1.4}**: ¿Qué temas de investigación han sido más/menos replicados?
- **PI_{1.5}**: ¿Qué métodos empíricos son los más utilizados en las replications?
- **PI_{1.6}**: ¿Cómo evoluciona el número de estudios que publican replications?
- **PI_{1.7}**: ¿Cómo se distribuyen los estudios según los temas de investigación y los métodos empíricos?

En las siguientes secciones, se abordan cada una de las preguntas de investigación utilizando gráficos para visualizar los datos recopilados y ofreciendo las interpretaciones consiguientes.

3.2 ¿QUÉ FOROS SE UTILIZAN PARA PUBLICAR REPLICACIONES?

De los 137 estudios primarios identificados, 75 (55%) se publicaron en revistas y 62 (45%) en actas de congresos.

¹www.scopus.com

²www.webofknowledge.com

La tabla §3.1 presenta las revistas donde se han publicado los estudios primarios, ordenado según el número de estudios publicados, incluyendo información relativa a su indexación en los rankings de calidad de revistas tales como JCR—y su cuartil 2018—o SCImago.

Tabla 3.1: Revistas en las que se han publicado estudios de replicación en 2013–2018 ordenados por número de publicaciones

# ^a	Nombre de la revista	Acrónimo	SCImago	JCR	Q ₂₀₁₈ ^b
21	Empirical Software Engineering	EMSE	✓	✓	Q1
14	Information and Software Technology	IST	✓	✓	Q1
6	IEEE Transactions on Software Engineering	TSE	✓	✓	Q1
5	ACM Transactions on Software Engineering and Methodology	TOSEM	✓	✓	Q1
5	Journal of Systems and Software	JSS	✓	✓	Q1
3	Journal of Visual Languages and Computing	JVLC	✓	✓	Q3
3	Software Quality Journal	SQJ	✓	✓	Q2
3	Software and Systems Modeling	SoSyM	✓	✓	Q1
2	Advanced Materials Research	AMR	✓	✗	–
2	Science of Computer Programming	SCP	✓	✓	Q3
2	Journal of Software — Evolution and Process	JSEP	✓	✓	Q3
1	Applied Soft Computing	ASC	✓	✓	Q1
1	Future Generation Computer Systems	FGCS	✓	✓	Q1
1	Requirements Engineering	RE	✓	✓	Q2
1	Computer Journal	CJ	✓	✓	Q4
1	Advances in Intelligent Systems and Computing	AISC	✗	✗	–
1	Journal of Theoretical and Applied Information Technology	JTAIT	✓	✗	–
1	International Journal of Software Engineering and Knowledge Engineering	IJSEKE	✓	✓	Q4
1	e-Informatica Software Engineering Journal	EISEJ	✓	✗	–
1	Expert Systems with Applications	ESA	✓	✓	Q1

^aNúmero de estudios publicados. ^b Cuartil 2018 en el índice JCR.

Las dos principales revistas, EMSE e IST, incluyen casi la mitad de los artículos publicados en revistas, es decir, representan el 25% de todos los estudios incluidos en esta revisión. Cabe destacar la presencia de la revista *Advanced Materials Research* (AMR), cuyo ámbito de aplicación es muy diferente al de IS. Se verificó manualmente que los dos artículos publicados en esta revista [246, 247] se seleccionaron correctamente.

Para conocer el *método experimental* utilizado y el *tema de investigación* tratado en cada revista, los estudios se han clasificado según el *método experimental* utilizado en:

- *Experimento* (EX)
- *Cuasi-experimento* (QE)

- *Estudio de caso* (CS)
- *Encuesta* (SV).

Asimismo, según el *tema de investigación*, los estudios se han clasificado teniendo en cuenta las *Áreas de conocimiento* de SWEBOK [32], a saber:

- *Requisitos* (REQS)
- *Diseño* (DESG)
- *Construcción de software* (CONS)
- *Pruebas* (TEST)
- *Mantenimiento del software* (MAIN)
- *Gestión de la configuración* (CONF)
- *Gestión de la IS* (MNGT)
- *Procesos* (PROC)
- *Modelos y métodos* (METH)
- *Calidad del software* (QUAL)
- *Práctica profesional* (SKIL)
- *Economía* (ECON).

Es necesario tener en cuenta que algunos de los estudios se han clasificado en más de un *área temática de investigación*.

Para cada revista, la tabla §3.2 presenta el número de estudios clasificados según el método empírico utilizado y el área temática de investigación. Como se puede ver, el *experimento controlado* (EX) es el método empírico preferido y hay una gran variedad de temas de investigación, aunque *requisitos* (REQS) y *pruebas* (TEST) destacan sobre el resto. Los temas *gestión de la configuración* (CONF) y *economía* (ECON) no están incluidos en la tabla §3.2 ya que no se han encontrado estudios de replicación que aborden dichos temas de investigación en los años cubiertos en la revisión.

Tabla 3.2: Revistas en las que se han publicado estudios de replicación en 2013–2018 junto con el método empírico y área de conocimiento

Revista	Método empírico				Área de conocimiento									
	SV	QE	EX	CS	REQS	DESG	CONS	TEST	MAIN	MNGT	PROC	METH	QUAL	SKIL
EMSE	1	2	17	1	4	1	10	5	6	0	0	1	4	2
IST	2	1	9	2	6	2	1	2	1	0	1	2	1	3
TSE	0	0	5	1	3	2	0	0	1	0	0	1	2	0
TOSEM	0	0	5	0	2	1	2	2	1	0	0	0	0	0
JSS	0	0	5	0	2	2	0	2	1	0	0	1	0	2
JVLC	0	0	3	0	1	2	1	0	0	1	0	0	0	0
SQJ	1	0	1	1	0	1	0	3	0	0	0	0	1	0
SoSyM	0	0	3	0	1	2	0	0	1	0	0	2	0	0
AMR	0	0	1	1	0	1	0	1	0	1	0	0	0	0
SCP	0	0	2	0	1	0	0	1	0	0	0	1	0	0
JSEP	0	0	2	0	0	0	1	1	1	0	0	0	1	0
ASCJ	0	0	0	1	0	0	1	1	0	0	0	0	1	0
FGCS	0	1	0	0	0	0	0	0	0	1	0	0	1	0
RE	0	0	0	1	1	0	0	0	0	0	0	0	0	0
CJ	0	0	1	0	0	0	0	1	0	0	0	0	0	0
AISC	0	0	1	0	1	1	0	0	0	0	0	0	0	0
JTAIT	0	0	1	0	1	0	0	0	0	0	0	0	0	0
IJSEKE	0	0	1	0	0	0	0	0	0	0	0	1	0	0
EISEJ	0	0	1	0	0	0	0	1	0	0	0	0	0	0
ESA	0	0	0	1	0	0	0	1	0	0	0	0	1	0
Total	4	4	58	9	23	15	16	21	12	3	1	9	12	7

La tabla §3.3 muestra las conferencias en las que se han presentado algún estudio de replicación en 2013–2018, destacando la conferencia ESEM con más del 15% de todos los estudios incluidos en la presente revisión.

En el *International Workshop on Replication in Empirical Software Engineering Research ICSE (RESER)*, a pesar de ser un taller sobre replications, sólo se presentaron 3 estudios de replications. Este taller se celebró en 2010, 2011 y 2013, por lo que debido al rango de años cubiertos en la revisión, sólo se incluye la edición de 2013.

De forma similar a la tabla §3.2, la tabla §3.4 muestra para cada conferencia con más de un estudio de replicación, el número de estudios junto al método empírico y área de conocimiento.

Tabla 3.3: Conferencias en las que se han publicado estudios de replicación en 2013–2018 ordenadas por número de publicaciones

# ^a	Nombre de la conferencia	Acrónimo
9	International Symposium on Empirical Software Engineering and Measurement	ESEM
6	International Conference on Software Engineering	ICSE
4	International Conference on Evaluation and Assessment in Software Engineering	EASE
4	International Conference on Software Engineering and Knowledge Engineering	SEKE
3	International Workshop on Replication in Empirical Software Engineering Research	RESER
3	Requirements Engineering Conference	RE
3	Ibero–American Conference on Software Engineering	CIBSE
2	Americas Conference on Information Systems	AMCIS
2	ACM Symposium on Applied Computing	SAC
2	IEEE International Conference on Program Comprehension	ICPC
1	International Conference on Product-Focused Software Process Improvement	PROFES
1	IEEE International Workshop on Machine Learning Techniques for Software Quality Evaluation	MaLTeSQuE
1	ACM Research in Adaptive and Convergent Systems	RACS
1	Brazilian Symposium on Software Components, Architectures and Reuse	SBCARs
1	IEEE International Conference on Software Engineering and Service Sciences	ICSESS
1	International Conference on Software Testing, Verification and Validation	ICST
1	CM Joint European Software Engineering Conference and Symposium on the Foundations of Software Eng.	ESEC/FSE
1	Working Conference on Reverse Engineering	WCRE
1	Student Research Workshop at the Conference of the European Chapter of the Association for Comp. Ling.	EACLsRW
1	Working IEEE/IFIP Conference on Software Architecture	WICSA
1	International Workshop on CrowdSourcing in Software Engineering	CSI-SE
1	International Workshop on Emerging Trends in Software Metrics	WETSOM
1	International Workshop on Conducting Empirical Studies in Industry	CESI
1	Annual Conference on Innovation and Technology in Computer Science Education	ITiCSE
1	International Conference on Software Quality. The Future of Systems and Software Development	SWQD
1	CSI International Conference on Software Engineering	CONSEG
1	IEEE International Conference on Software Maintenance and Evolution	ICSME
1	International Conference on Agile Software and Systems Development	XP
1	IEEE International Working Conference on Mining Software Repositories	MSR
1	Winter Simulation Conference	WSC
1	International Conference on Augmented Cognition. Neurocognition and Machine Learning	AC
1	IEEE Symposium on Computational Intelligence and Data Mining	CIDM
1	ACM Southeast 2018 Conference	ACMSE
1	International Conference on Conceptual Modeling	ER

^aNúmero de estudios publicados.

Tabla 3.4: Conferencias en las que se ha publicado más de un estudio de replicación en 2013–2018 junto con el método empírico y área de conocimiento

Conferencia	Método empírico				Área de conocimiento									
	SV	QE	EX	CS	REQS	DESG	CONS	TEST	MAIN	MNGT	PROC	METH	QUAL	SKIL
ESEM	2	0	6	1	0	2	2	3	0	3	0	2	4	0
ICSE	0	0	5	1	0	1	3	2	1	0	0	0	1	1
EASE	0	0	4	0	2	1	0	2	0	0	0	0	1	0
SEKE	2	0	2	0	1	0	1	1	0	0	0	2	1	0
RESER	0	0	2	1	1	0	1	1	0	0	0	1	0	0
RE	1	0	1	1	3	0	0	0	0	0	0	0	0	0
CIBSE	0	0	3	0	0	0	1	1	0	2	0	0	0	1
AMCIS	1	0	1	0	0	0	1	0	0	0	1	0	1	0
SAC	0	0	1	1	0	0	1	1	0	0	0	0	0	0
ICPC	0	0	2	0	0	1	0	0	0	0	0	1	0	0
Total	6	0	27	5	7	5	10	11	1	5	1	6	8	2

En el caso de las conferencias, *pruebas* (TEST) y *construcción de software* (CONS) son los temas que se abordan con más frecuencia, aunque *calidad del software* (QUAL) y *requisitos* (REQS) tienen valores cercanos. Al igual que en las revistas, en las conferencias no se encontraron estudios de replicación para los temas *gestión de la configuración* (CONF) y *economía* (ECON) y por lo tanto, no se incluyen en la tabla §3.4. Con respecto a los métodos empíricos más utilizados, el *experimento controlado* (EX), al igual que en las revistas, destaca claramente del resto de métodos empíricos.

3.3 ¿QUIÉNES SON LOS AUTORES E INSTITUCIONES QUE PUBLICAN REPLICACIONES?

En la tabla §3.5 aparecen los autores con al menos cuatro estudios de replicación publicadas en el periodo de la revisión junto al método empírico y área de conocimiento. Destacan Giuseppe Scanniello de la Universidad de Basilicata (Italia) y Natalia Juristo y Oscar Dieste de la Universidad Politécnica de Madrid (España) como los autores más prolíficos. También se puede observar que el método empírico más utilizado es el *experimento controlado* (EX) y el tema de investigación más abordado es *requisitos* (REQS).

Otros autores de prestigio como Claes Wohlin del Blekinge Institute of Technology (Sweden), tienen publicaciones relevantes sobre IS empírica en el período de revisión,

Tabla 3.5: Autores con al menos cuatro estudios de replicación publicados en 2013-2018 junto con el método empírico y área de conocimiento

Autor	Método empírico				Área de conocimiento									
	SV	QE	EX	CS	REQS	DESG	CONS	TEST	MAIN	MNGT	PROC	METH	QUAL	SKIL
Scanniello, G.	0	0	15	0	7	7	5	3	2	1	0	2	1	0
Juristo, N.	0	1	7	0	2	1	0	2	1	0	0	3	3	2
Dieste, O.	1	0	6	0	3	0	1	3	0	0	0	2	2	2
Fernández, D.M.	5	0	0	1	6	0	0	0	0	0	0	0	0	0
Ricca, F.	0	0	6	0	2	2	1	0	1	0	0	2	0	0
Abrahão, S.	0	0	6	0	2	3	0	1	0	0	1	3	0	0
Gravino, C.	0	0	6	0	4	4	2	0	2	0	0	0	0	0
Tortora, G.	0	0	6	0	4	4	2	0	2	0	0	0	0	0
Genero, M.	0	0	5	0	5	1	1	0	3	0	0	0	0	0
Spínola, R.O.	5	0	0	0	3	0	1	0	0	1	0	1	0	0
Insfran, E.	0	0	5	0	2	3	0	1	0	0	1	2	0	0
Torchiano, M.	1	0	4	0	1	2	1	0	1	0	0	2	0	0
Carver, J.C.	1	0	4	0	2	1	2	0	2	0	0	0	1	0
Conte, T.	3	0	2	0	3	1	0	1	0	0	1	0	1	0
Vegas, S.	0	0	4	0	0	1	0	0	1	0	0	3	1	1
Risi, M.	0	0	4	0	1	2	2	2	1	0	0	0	0	0
Fucci, D.	0	0	4	0	0	0	1	3	0	0	0	0	3	1
Wagner, S.	4	0	0	0	4	0	0	0	0	0	0	0	0	0
Prikladnicki, R.	3	0	1	0	3	0	0	0	0	0	1	0	1	0
Reggio, G.	0	0	4	0	1	2	0	0	1	0	0	2	0	0
Total	23	1	87	1	55	34	19	16	17	2	4	22	13	6

por ejemplo [244] con 121 citas en SCOPUS, a fecha mayo de 2019, pero no aparece en la tabla §3.5 ya que, aunque contribuyen significativamente al campo, no hemos encontrado ninguna replicación explícitamente en el período 2013–2018.

Para analizar las colaboraciones entre autores, hemos utilizado la herramienta VOSviewer [238]. La figura §3.1 muestra un mapa de coautores generado con VOSviewer en el que los autores de la tabla §3.5, es decir, los autores con al menos 4 publicaciones, se muestran como burbujas. El tamaño de las burbujas depende del número de publicaciones, mientras que el grosor de los enlaces entre dos burbujas, depende del número de publicaciones comunes entre dichos autores. De esta forma, los autores con publicaciones en común se muestran agrupados.

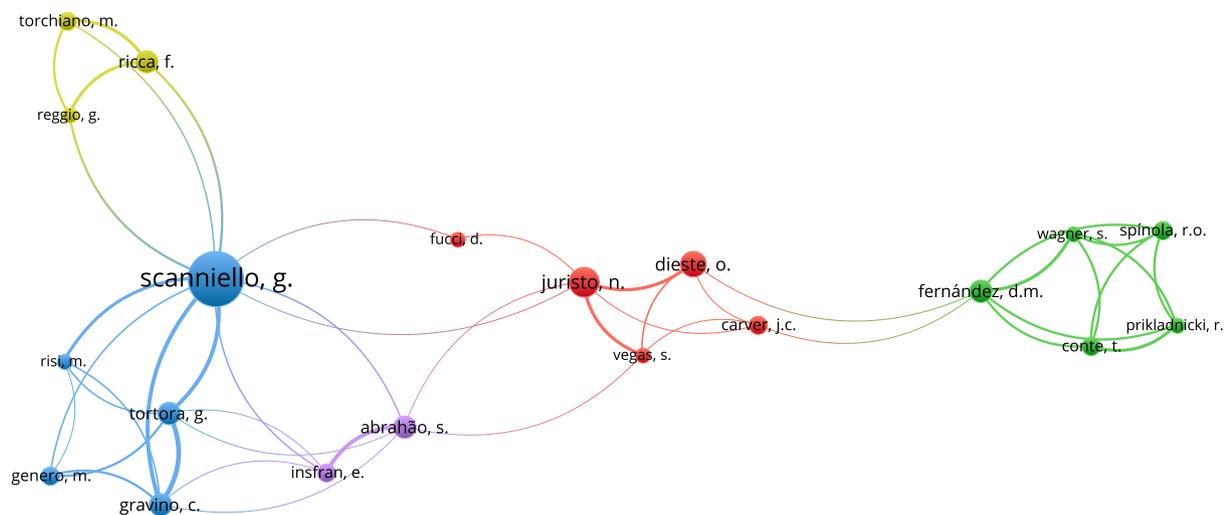


Figura 3.1: Mapa de coautores con al menos cuatro estudios de replicación en 2013–18

De acuerdo con los datos en la tabla §3.5, tres agrupaciones principales están representados en la figura §3.1 en torno a los autores más prolíficos, es decir Scanniello, Juristo y Daniel Méndez Fernández de la Technical university of Munich (Alemania). También, se identifican otras dos pequeñas agrupaciones en torno a Silvia Abrahão de la Universidad Politécnica de Valencia, (España) y Filippo Ricca de la Universidad de Genova (Italia). La figura §3.2 muestra el mapa de coautores considerando todos los autores, es decir, no sólo aquellos con al menos cuatro publicaciones.

La tabla §3.6 muestra las instituciones que actualmente albergan a los autores más activos y por lo tanto lideran los avances en replicaciones de IS empírica. Los datos se han obtenido contabilizando las instituciones de los autores de cada estudio primario. En el caso de que varios autores del mismo estudio pertenecieran a la misma institución, dicha institución se ha contabilizado solo una vez.

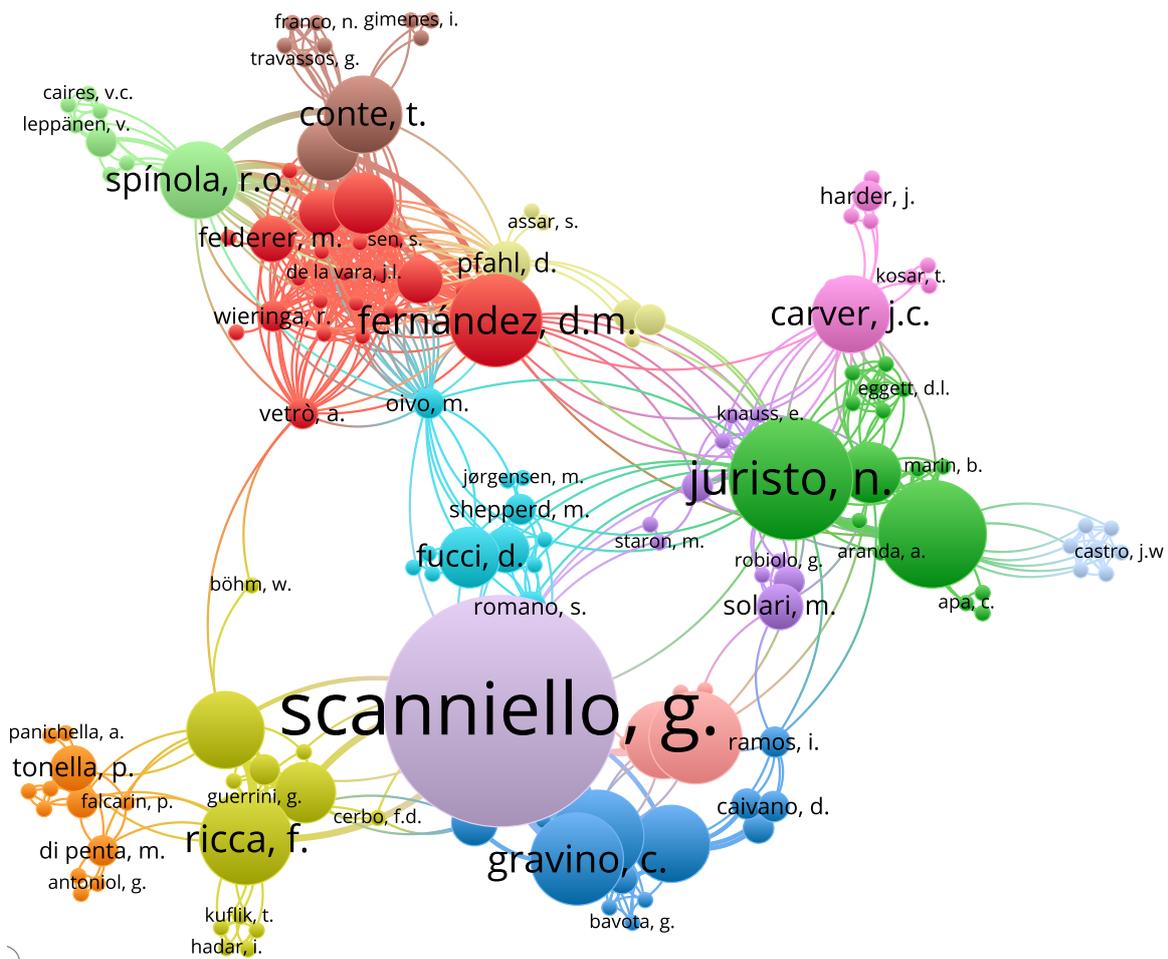


Figura 3.2: Mapa de coautores con al menos un estudio de replicación en 2013–18

Las universidades europeas lideran claramente el ranking, siendo Italia y España los países más activos en este campo. Este hecho es una clara continuación de los resultados de [57], donde las instituciones más activas fueron la Universidad de Castilla-La Mancha (España), la Universidad de Sannio (Italia), el Simula Research Laboratory (Noruega) y la Universidad de Valladolid (España).

Las cinco principales instituciones son las universidades italianas (Universidad de Basilicata y Universidad de Salerno) y las universidades españolas (Universidad de Castilla-La Mancha, Universidad Politécnica de Madrid y Universidad Politécnica de Valencia).

3.4. ¿CUÁLES SON LOS ESTUDIOS MÁS CITADOS QUE PUBLICAN REPLICACIONES?

Tabla 3.6: Instituciones con al menos cuatro estudios de replicación publicados en 2013–2018 ordenados por número de publicaciones

# ^a	Institución	Web oficial	País
15	University of Basilicata	http://portale.unibas.it	Italia
8	University of Salerno	https://www.unisa.it/	Italia
7	Universidad de Castilla - La Mancha	https://www.uclm.es/	España
6	Universidad Politécnica de Madrid	http://www.fi.upm.es	España
6	Universidad Politécnica de Valencia	http://www.upv.es/	España
5	Technical University of Munich	https://www.tum.de/	Alemania
5	University of Genova	https://unige.it/	Italia
5	Federal University of Amazonas	https://ufam.edu.br/	Brasil
5	University of São Paulo	https://icmc.usp.br/	Brasil
4	University of Oulu	https://www.oulu.fi/	Finlandia
4	University of Lund	https://lunduniversity.lu.se/	Suecia
4	Fondazione Bruno Kessler	https://www.fbk.eu/en/	Italia
4	Politecnico di Torino	https://www.polito.it/	Italia

^aNúmero de estudios publicados.

3.4 ¿CUÁLES SON LOS ESTUDIOS MÁS CITADOS QUE PUBLICAN REPLICACIONES?

La tabla §3.7 presenta los estudios más citados, incluyendo información del foro donde fueron publicados, año de publicación, número de citas según SCOPUS a fecha

Tabla 3.7: Estudios más citados que reportan al menos una replicación en 2013–2018 ordenados por el número de citas según SCOPUS a mayo de 2019

Estudio primario	Foro	Año ^a	#Citas	#Citas/Año	Método empírico	Área de conocimiento
Gothra <i>et al.</i> [93]	ICSE	2015	109	27,3	CS	QUAL
Pearson <i>et al.</i> [173]	ICSE	2017	49	24,5	EX	TEST
Binkley <i>et al.</i> [30]	EMSE	2013	44	7,3	EX	TEST
Ceccato <i>et al.</i> [45]	EMSE	2014	41	8,2	EX	CONS
Abrahão <i>et al.</i> [2]	TSE	2013	40	6,7	EX	REQS, DESC
Fernández D.M. <i>et al.</i> [80]	IST	2015	34	8,5	SV	REQS
Fernández A. <i>et al.</i> [77]	JSS	2013	34	5,7	EX	TEST, METH
Fernández D.M. <i>et al.</i> [81]	EMSE	2017	28	14	SV	REQS
Scanniello <i>et al.</i> [205]	TOSEM	2014	27	5,4	EX	REQS, CONS
Ribeiro <i>et al.</i> [190]	ICSE	2014	24	4,8	EX	MAIN
Hadar <i>et al.</i> [106]	IST	2013	24	4	EX	REQS

^aAño de publicación.

mayo de 2019, así como el método empírico utilizado y las áreas temáticas de investigación correspondientes.

Se puede observar que los dos estudios más citados utilizan métodos automatizados para la predicción de defectos y la localización de fallos, es decir, los sujetos no son humanos. Aunque nuestro principal interés son las replications de experimentos con sujetos humanos, siguiendo criterios similares a [57] no se han excluido estos estudios. Destacar también que los dos artículos más citados de D. M. Fernández tratan de una *familia de encuestas* sobre el estado de la práctica de la Ingeniería de Requisitos en Alemania.

En cuanto a los foros en los que se han publicado estos estudios más citados, la única conferencia es ICSE con 3 estudios y el resto son revistas destacando EMSE (3 estudios) e IST (2 estudios).

3.5 ¿QUÉ TEMAS DE INVESTIGACIÓN HAN SIDO MÁS/MENOS REPLICADOS?

La tabla §3.8 presenta los estudios primarios agrupados por áreas temáticas de investigación.

Se puede observar que el número de estudios en la tabla §3.8 (215) es mayor que el número de estudios primarios (137) ya que algunos estudios se han clasificado en más de un área temática.

En concordancia con los datos presentados en las secciones anteriores, las áreas de investigación que han atraído más atención son *pruebas* (TEST), *requisitos* (REQS) y *construcción de software* (CONS) con más del 25% de los estudios primarios que abordan cada uno de estas áreas.

Se han identificado algunas lagunas en la investigación relacionadas con *economía* (ECON), *gestión de la configuración* (CONF) y *procesos* (PROC). Así, sólo 3 estudios de replicación se han publicado en el área PROC y ninguno en las de ECON y CONF en el periodo cubierto en la revisión. Hemos encontrado algunos estudios empíricos relacionados con las áreas de CONF, como por ejemplo [108] y de PROC, como [120] que posteriormente podrían haber sido replicados pero no hemos podido encontrar ningún estudio relacionado con el área de ECON.

Tabla 3.8: Áreas de investigación y estudios primarios publicados en 2013–2018 ordenados por número de estudios

Área	Estudios primarios	#Estudios
TEST	[5, 12, 13, 16, 18, 24, 33, 46, 61, 70, 71, 72, 77, 87, 89, 90, 100, 104, 114, 150, 153, 159, 167, 170, 173, 183, 198, 203, 209, 213, 219, 222, 226, 229, 236, 242, 246, 248]	38
REQS	[2, 6, 13, 14, 23, 27, 49, 67, 80, 81, 82, 83, 85, 98, 106, 112, 132, 133, 134, 142, 152, 156, 160, 166, 175, 176, 178, 189, 191, 204, 205, 207, 210, 223, 249]	35
CONS	[30, 33, 37, 45, 48, 94, 99, 102, 103, 105, 107, 112, 115, 139, 151, 153, 155, 157, 162, 165, 172, 181, 183, 194, 199, 205, 208, 209, 216, 221, 231, 232, 236, 242, 248]	35
DESG	[2, 3, 7, 27, 44, 49, 61, 72, 98, 99, 105, 115, 140, 143, 150, 160, 178, 185, 198, 206, 207, 212, 222, 233, 234, 241, 247]	27
QUAL	[4, 6, 12, 16, 49, 63, 70, 71, 87, 89, 90, 93, 100, 110, 111, 115, 121, 153, 169, 184, 186, 202, 226, 242]	24
METH	[7, 37, 73, 74, 75, 77, 84, 98, 121, 131, 156, 169, 185, 192, 233, 239, 241]	17
MAIN	[5, 23, 48, 82, 83, 107, 140, 157, 165, 190, 192, 194, 206, 208]	14
MNGT	[44, 112, 113, 143, 179, 180, 184, 186, 193, 204, 247]	11
SKIL	[4, 27, 56, 86, 94, 103, 183, 195, 199, 230, 239]	11
PROC	[162, 202, 228]	3
Total		215

A veces, los estudios empíricos son recientes, tales como [154], por lo que parece razonable que sus replications no hayan sido publicadas todavía. Sin embargo, otros estudios no tan recientes como [62], no han sido replicados, lo que es una situación muy común en la mayoría de los experimentos en IS [97].

3.6 ¿QUÉ MÉTODOS EMPÍRICOS SON LOS MÁS UTILIZADOS EN LAS REPLICACIONES?

La tabla §3.9 muestra los estudios primarios agrupados por el método empírico utilizado. Como se puede ver, el *experimento controlado* (EX) se usa en casi el 75% de los estudios.

Hay que señalar que el bajo número de *cuasi-experimentos* (QE), probablemente es debido al hecho de que en algunos estudios se habla de *experimentos controlados* en el resumen e incluso en la introducción y sin embargo en secciones internas son descritos como *cuasi-experimentos*. De hecho, algunos autores consideran los *cuasi-experimentos* como un tipo específico de *experimento controlado* [182, 225].

Tabla 3.9: Métodos empíricos y estudios primarios publicados en 2013-2018 ordenados por número de estudios

Método	Estudios primarios	#Estudios
EX	[2, 3, 5, 6, 7, 12, 13, 14, 18, 23, 24, 27, 30, 33, 44, 45, 46, 48, 49, 70, 71, 72, 73, 74, 75, 77, 82, 83, 84, 87, 89, 90, 94, 98, 99, 102, 104, 105, 106, 107, 111, 113, 114, 115, 121, 131, 139, 140, 142, 150, 151, 155, 156, 157, 159, 160, 165, 167, 169, 170, 172, 173, 175, 178, 179, 180, 181, 183, 185, 189, 190, 191, 192, 195, 198, 199, 202, 203, 204, 205, 206, 207, 208, 209, 210, 212, 213, 221, 222, 223, 226, 228, 229, 233, 234, 236, 239, 242, 247, 248, 249]	101 (73,72%)
CS	[16, 56, 63, 86, 93, 100, 110, 112, 134, 143, 152, 153, 162, 166, 176, 186, 216, 219, 246]	19 (13,87%)
SV	[37, 61, 80, 81, 85, 132, 133, 193, 230, 231, 232, 241]	12 (8,76%)
QE	[4, 67, 103, 184, 194]	5 (3,65%)
Total		137 (100%)

3.7 ¿CÓMO EVOLUCIONA EL NÚMERO DE ESTUDIOS QUE PUBLICAN REPLICACIONES?

El número de estudios de replicación publicados en el período de revisión se muestra en la figura 3.3.

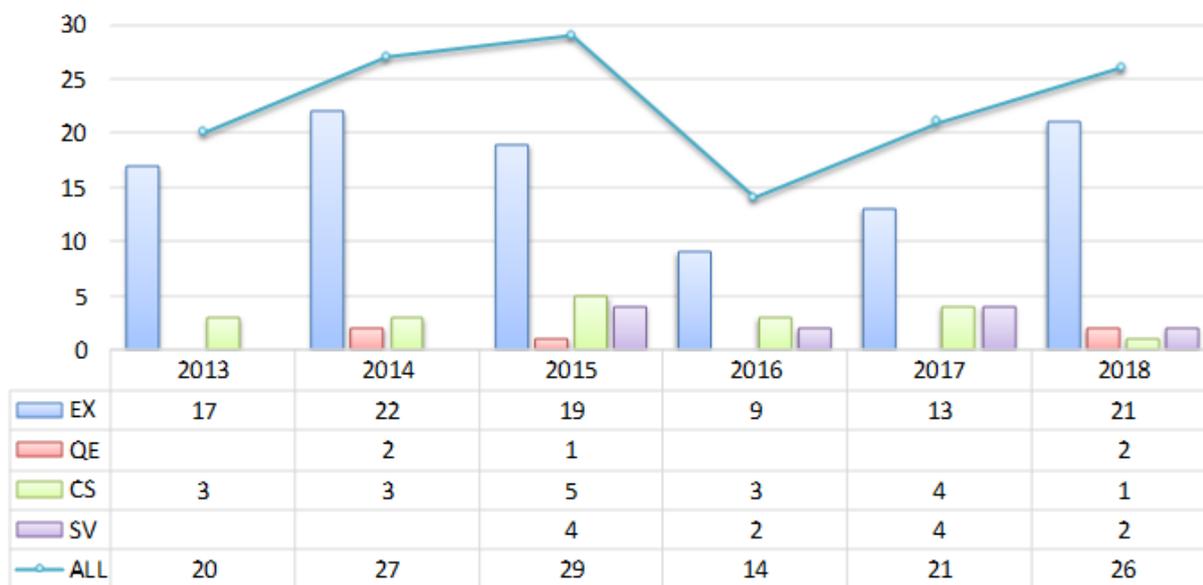


Figura 3.3: Número de estudios de replicación publicados en 2013–2018 agrupados por método empírico

Con un promedio de 22.8 estudios por año, no se observan grandes variaciones debido a que el número de publicaciones es bajo. Con respecto a los métodos empíricos

utilizados, el *experimento controlado* (EX) –con un promedio de 16.8 por año– presenta números similares cada año excepto para 2016. El número de *cuasi-experimentos* (QE) es muy bajo, como se ha comentado en secciones previas. Por otro lado, el número de *estudios de caso* (CS) cada año es muy similar y las *encuestas* (SV) presentan pocas variaciones desde 2015.

3.8 ¿CÓMO SE DISTRIBUYEN LOS ESTUDIOS?

La figura §3.4 muestra el número de estudios según el método empírico y área de conocimiento. Según lo comentado en secciones anteriores, *pruebas* (TEST), *construcción de software* (CONS), *requisitos* (REQS) y *diseño* (DESG) son los temas de investigación con más replicaciones y usando *experimentos controlados* como método empírico (EX) más habitual.

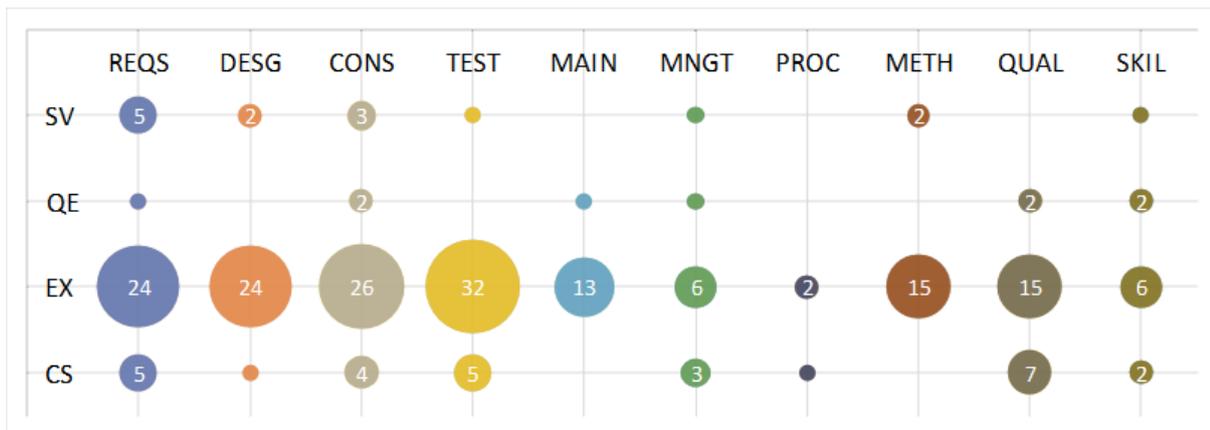


Figura 3.4: Clasificación de estudios de replicación publicados en 2013–2018 según el método empírico y área de conocimiento

Por otro lado, *calidad del software* (QUAL)—con 7 estudios de caso de un total de 24 contribuciones—, REQS y TEST son las áreas temáticas en las que el método empírico *estudio de caso* (CS) se ha aplicado con mayor frecuencia.

3.9 CONCLUSIONES

En esta sección se presentan algunas conclusiones obtenidas como respuesta a las preguntas de investigación:

- **PI_{1.1}**. Se ha encontrado que el número de estudios publicados en revistas (75) es mayor que en las conferencias (62) y que tres foros—las revistas EMSE e IST y la conferencia de ESEM—concentran más del 30% de los estudios de replicación publicados.
- **PI_{1.2}**. Los autores más prolíficos son G. Scanniello (U. de Basilicata), N. Juristo y O. Dieste (U. P. Madrid), D. M. Fernández (T. U. de Munich), S. Abrahão (U. P. de Valencia), F. Ricca (U. de Genoa) y C. Gravino y G. Tortora (U. de Salerno) y las principales instituciones son las universidades de Basilicata, Salerno, Castilla–La Mancha, U. P. de Madrid y U. P. de Valencia.

Utilizando los mapas de coautores (ver figuras §3.1 y §3.2), se han identificado además *clusters* de autores cooperantes como el *cluster* liderado por G. Scanniello y otros investigadores italianos y el grupo de autores de la U. P. de Madrid (N. Juristo, O. Dieste y S. Vegas) que incluye también la cooperación con otros investigadores como J. C. Carver, D. M. Fernández y S. Abrahão entre otros.

La mayoría de las replications en IS han sido realizadas por investigadores e instituciones italianas y españolas, sin subestimar la relevante contribución de Brasil y Alemania.

- **PI_{1.3}**. Los dos estudios más citados son trabajos presentados en la conferencia ICSE sobre técnicas automatizadas para modelos de predicción de defectos [93] y localización de fallos [173]. Sólo tres de los autores más prolíficos (G. Scanniello, D. M. Fernández y S. Abrahão) son autores de cuatro de los 10 estudios más citados, mientras que ninguna de las instituciones que albergan a los autores de [93, 173] está en la lista de instituciones líderes.
- **PI_{1.4}**. Los temas de mayor interés por parte de los investigadores son *pruebas*, *requisitos* y la *construcción de software*.
- **PI_{1.5}**. Los *experimentos controlados* es el método empírico mas usado.
- **PI_{1.6}**. En cuanto a la evolución del número de estudios que reportan replications, se encontraron 137 estudios en 6 años, es decir, una media de 22.8 por año. Presentan pocas variaciones anuales, excepto en el año 2016 que es un poco más bajo. Algunos estudios incluyen más de una replicación, por lo que el número promedio de replications por año podría ser mayor.
- **PI_{1.7}**. En cuanto a la distribución de los estudios, la combinación de *pruebas* y *experimentos controlados* es la que se encuentra con más frecuencia.

3.10 RESUMEN

Para conocer el estado actual de las replications en IS, se han analizado 137 artículos publicados entre 2013 y 2018 que reportan al menos una replicación de un estudio empírico en IS.

Se han identificado *i)* principales foros donde se han publicado estudios sobre replications; *ii)* autores, coautores e instituciones; *iii)* estudios más citados; *iv)* temas de investigación tratados; *v)* métodos empíricos utilizados; *vi)* distribución temporal de los estudios; y *vii)* clasificación de estudios según temas de investigación y métodos empíricos.

Según los resultados, los foros más relevantes son las revistas *Empirical Software Engineering* (EMSE) y *Information and Software Technology* (IST), y el congreso *Empirical Software Engineering and Measurement* (ESEM).

Los estudios que más han atraído la atención y, por tanto, más citados se han publicado principalmente en revistas y en la conferencia *International Conference on Software Engineering*.

Los temas abordados con mayor frecuencia en los estudios de replicación han sido *pruebas, requisitos y construcción de software*, mientras que el método empírico más habitual ha sido el *experimento controlado*. Por otro lado, se han identificado vacíos de investigación en áreas tales como *procesos de la ingeniería de software, gestión de la configuración software* y *economía de la ingeniería del software*.

PROBLEMÁTICA EN LA REALIZACIÓN DE LAS REPLICACIONES

Sorprenderse y maravillarse es comenzar a entender

José Ortega y Gasset (1883 – 1955),

Filósofo y ensayista español

En este capítulo, respondiendo a la pregunta de investigación PI_2 , se analizan algunos de los problemas detectados al abordar la replicación de experimentos y que contribuyen a que sean pocos los experimentos replicados.

En la sección §4.1 se destaca el bajo número de replicaciones. En la sección §4.2 se señala la necesidad de guías específicas para reportar replicaciones. En la sección §4.3 se remarca la carencia de paquetes de laboratorio para facilitar las replicaciones. En la sección §4.4 se señala el problema de la transmisión del conocimiento. En la sección §4.5 se presenta el problema de la falta de acuerdo en la terminología utilizada por los autores. En la sección §4.6 se remarca la dificultad de clarificar el contexto. En la sección §4.7 se advierte sobre la dificultad añadida del factor humano en las replicaciones. En la sección §4.8 se analiza la dificultad en la publicación de replicaciones. En la sección §4.9 se analizan dichos problemas en relación a la plantilla propuesta. En la sección §4.10 se resumen los problemas identificados.

4.1 INTRODUCCIÓN

Los primeros estudios sobre replications en IS son relativamente recientes. El primer artículo que reportaba una replicación de un estudio empírico fue publicado en 1994 por Daly *et al.* [58]. Más tarde, Basili *et al.* [22] en 1999 proponen un marco para generar conocimiento mediante la replicación de experimentos dentro de las *familias de experimentos*.

A pesar de la importancia de las replications y aunque en los últimos años ha aumentado su práctica (Da Silva *et al.* [57]), el número de replications en IS sigue siendo bajo (Solari *et al.* [227]). En [43] se analiza el porcentaje entre el número de replications y experimentos controlados reportados llegando a la conclusión de que la replicación no es una práctica frecuente ya que la mayoría de los experimentos no han sido replicados [97]. La falta de replications se hace más evidente cuando se habla de replications *externas* que son las de mayor poder de confirmación. Sin replications externas los resultados son sólo provisionales [34, 35].

Al bajo número de replications, se unen los problemas presentados en las replications ya publicadas, tales como los problemas relacionados con el uso de la estadística. En [1] se señala que en el campo biomédico se han encontrado estudios que, ya sea debido a métodos experimentales inadecuados o análisis estadísticos incorrectos, no pueden ser repetidos e incluso se han detectado casos de fraude. Según Blanco *et al.* [31], en otras áreas de conocimiento distintas a la IS como Psicología, esta situación se conoce como *crisis de replicabilidad*.

El problema también se presenta en IS. Kitchenham *et al.* [137] revisaron 13 estudios sobre familias de experimentos en IS con participantes humanos, observando un uso indebido de la terminología estadística y mala interpretación de la práctica estadística, añadiendo, que al estar publicados en revistas de alta calidad pueden conducir a la propagación de malas prácticas al asumir la metodología como correcta.

De igual forma, Reyes *et al.* [188] analizaron los experimentos publicados en ICSE (International Conference on Software Engineering) señalando que los conceptos y técnicas estadísticas suelen aplicarse incorrectamente y recomiendan: *i)* mejorar la formación estadística de los investigadores en IS y *ii)* establecer directrices para la presentación de informes y normas de calidad para mejorar la situación.

Por otra parte, los problemas identificados que pueden contribuir al bajo número de replications, se pueden resumir en:

- *Falta de unicidad de criterios o específicamente guías para reportar las replications* (Carver [42]).
- *Carencia y debilidades de los paquetes de laboratorio que facilitan las replications* (Solari *et al.* [227]).
- *Dificultad para transmitir el conocimiento tácito* (Shull *et al.* [217]).
- *Falta de acuerdo sobre la terminología, la tipología, los objetivos y otras cuestiones relativas a las replications* [43, 59, 96].
- *Dificultad para clarificar el contexto de las replications* (Cartaxo *et al.* [39]).
- *Complejidad del factor humano que complica la realización de replications* (Lung *et al.* [146]).
- *Alto grado de esfuerzo y recursos necesarios para llevar a cabo un experimento* (Da Silva *et al.* [57]) *unido a la dificultad de su publicación.*

En las siguientes secciones se analizan, en mayor detalle, cada uno de estos problemas.

4.2 FALTA DE GUÍAS PARA REPORTAR REPLICACIONES

La falta de guías para reportar replications dificulta la comprensión tanto del experimento original y de la replicación a reportar, como de las replications previas. Carver [42] señala la necesidad de directrices para garantizar la información que se publica sobre cada replicación.

En 2005, Sjøberg *et al.* [225] realizaron una búsqueda manual de experimentos controlados en IS publicados entre 1993 y 2002 y ya detectaron la necesidad de directrices para la realización y publicación de experimentos que faciliten su replicación y meta-análisis.

En la misma línea, Almqvist [9] realizó en 2006 una búsqueda manual en revistas y actas de congresos centrada en replications en IS. Agrupó los estudios en 20 series de experimentos, clasificando las replications en internas y externas y analizando si confirmaban los resultados del estudio original. Sus principales recomendaciones fueron mejorar la presentación de informes experimentales, elaborar directrices para llevar a

cabo las replications y disponer de *paquetes de laboratorio* para promover la replicabilidad. La problemática asociada a los *paquetes de laboratorio* se verá en el siguiente apartado §4.3.

Se han propuesto directrices para reportar experimentos controlados en IS (Jedlitschka *et al.* [119]), (Juristo *et al.* [124]), (Wohlin *et al.* [245]). Sin embargo, las directrices de Carver [42] son las únicas enfocadas específicamente a replications y a garantizar que el proceso de experimentación y sus resultados se presentan de forma adecuada. En dichas guías se recomienda incluir: *i)* información sobre el experimento original, describiendo los participantes, preguntas de investigación, diseño, variables de contexto, artefactos utilizados y resultados; *ii)* información sobre la replicación, incluyendo motivo de la replicación, interacción con los experimentadores originales y cambios llevados a cabo; *iii)* comparación de los resultados del original y conclusiones. Cada cambio realizado, ya sea en el diseño, participantes, artefactos, procedimientos, datos recopilados y/o las técnicas de análisis debe ser especificado junto con la motivación de dicho cambio.

Estas directrices son una propuesta inicial y será necesario analizar si el tipo de replicación o la inclusión del experimento original influye, entre otros factores, en la estructura y contenido del informe de replicación. De acuerdo con Carver *et al.* [43], teniendo en cuenta la variedad de tipos de replicación, se necesitaría además, cierta flexibilidad en las directrices.

4.3 CARENCIA DE PAQUETES DE LABORATORIO

Para llevar a cabo una replicación de experimentos, es necesaria la transferencia de conocimientos entre los investigadores [227]. En esta misma línea, Vegas *et al.* [240] señalan que la comunicación con los experimentadores originales favorece que la replicación sea exitosa. Shull *et al.* en [218], analizan el uso de *paquetes de laboratorio* o *paquetes de replicación* como una forma de comunicación entre los equipos que realizan la replicación y los investigadores del estudio original. Basili *et al.* en [22], recomiendan el uso de *paquetes de laboratorio* proporcionados por los experimentadores incluyendo el material experimental junto con detalles de las técnicas utilizadas y Brooks *et al.* en [36] señalan la necesidad de facilitar la reutilización de los *artefactos experimentales* ya que sin un *paquete de laboratorio*, es poco probable que un experimento sea verificado mediante su replicación.

En [227], Solari *et al.* señalan algunas debilidades presentadas por los actuales *paquetes de laboratorio*: *i)* algunos sólo contienen el material experimental, asumiendo que los experimentadores ya habrán accedido al documento completo sobre el experimento; *ii)* no suelen contener las directrices para llevar a cabo la replicación; *iii)* no se proporciona información sobre replicaciones ya realizadas o de la evolución del experimento; y *iv)* no presentan una estructura estándar; la mayoría se basan en una página web y enlaces a otros archivos con distintos formatos.

A lo anterior se une la carencia de *paquetes de laboratorio* o en general de datos disponibles en algunos experimentos. En 2014, Larsson *et al.* [141], realizaron una revisión bibliográfica sobre la *accesibilidad a los datos* de experimentos para su replicación. Se encontró que sólo el 37% de los estudios revisados tenían datos disponibles.

En [227], se propone el contenido y estructura de los *paquetes de laboratorio* para experimentos en IS señalando que deben de incluir: *i)* directrices específicas para replicar el experimento, *ii)* descripción del experimento base separada del informe de cada replicación y *iii)* la evolución del experimento mediante las sucesivas replicaciones. El *paquete de laboratorio* se organiza en módulos, de forma que cada experimentador puede utilizar el módulo que sea de su interés (por ejemplo módulo para adaptar el diseño). Tras realizar una replicación, se debe actualizar el *paquete de laboratorio* añadiendo un nuevo módulo que contenga la descripción detallada de la replicación. De esta forma, queda reflejada la evolución histórica del experimento mediante las replicaciones llevadas a cabo y que entran a formar parte del *paquete de laboratorio*.

Por otro lado, Kitchenham en [135], advierte sobre el uso de los *paquetes de laboratorio* argumentando en contra de las replicaciones exactas. Señala que se pueden encontrar replicaciones que confirman los resultados debido a reutilización de materiales o procesos y contribuyen poco a ganar conocimiento.

Pensamos que la plantilla propuesta en esta memoria para especificar los cambios en las replicaciones, puede incorporarse al *paquete de laboratorio* formando parte de dicho módulo de replicación y facilitando la traza entre sucesivas replicaciones o evolución histórica del experimento.

4.4 EL CONOCIMIENTO TÁCITO

Los *paquetes de laboratorio* no son suficientes para garantizar el intercambio eficaz de conocimientos entre los experimentadores originales y los replicadores para que una

replicación tenga éxito.

Shull *et al.* [217] abordan la transferencia de conocimientos entre los experimentadores y sostienen que hay dos tipos de conocimiento: el *conocimiento tácito* y el conocimiento explícito. El conocimiento explícito forma parte del paquete de laboratorio y es por tanto accesible a los replicadores que necesitan comprender las técnicas que se están estudiando junto con los conocimientos necesarios para llevar a cabo el experimento. El *conocimiento tácito* es conocimiento necesario para llevar a cabo la replicación que tiene el experimentador y no se hace explícito al reportar la replicación, es decir no está incluido en el *paquete de laboratorio* ya sea, entre otras razones, porque no se consideró su relevancia o porque es difícil de transmitir.

Shull *et al.* [217] señalan que la falta de *conocimiento tácito* puede producir que los resultados del experimento original y la replicación no sean comparables debido a las diferencias en la ejecución de algunos pasos del experimento y añaden que en los experimentos con sujetos humanos, estas pequeñas variaciones pueden tener grandes efectos en los resultados. La influencia del factor humano en las replicaciones se analiza en la sección §4.7

La propuesta presentada en esta memoria es un forma de abordar el *conocimiento tácito* ya que contribuye a la transferencia del conocimiento dejando constancia del conocimiento que posee el experimentador y que antes no hacía explícito.

4.5 DIFERENTE TERMINOLOGÍA

Respecto de la terminología utilizada en las replicaciones, los investigadores han utilizado diferentes términos relativos a las replicaciones y específicamente para clasificar y definir los distintos tipos de replicaciones. La problemática relativa a la clasificación de replicaciones se ha analizado en la sección §2.4.

En la revisión de experimentos controlados llevada a cabo en el 2005 por Sjøberg *et al.* [225], además de la necesidad de directrices para la realización y publicación de experimentos comentada anteriormente (ver sección §4.2), se detectó falta de consistencia en la terminológica utilizada que dificulta la comprensión de los experimentos y sus replicaciones.

En 2014–2015, tres revisiones [29, 57, 59] analizaron el estado de las replicaciones en IS. En [29, 57] se seleccionaron estudios publicados hasta el 2012 reportando replicacio-

nes y en [59] se seleccionaron artículos publicados hasta 2013 que aunque no reportan ninguna replicación, abordan temas relativos a las replicaciones. Estos estudios coinciden en la falta de taxonomías, directrices y estandarización para llevar a cabo y publicar las replicaciones de estudios empíricos.

4.6 NECESIDAD DE CLARIFICAR EL CONTEXTO DE LAS REPLICACIONES

En 2015, Cartaxo *et al.* [39] realizaron una revisión bibliográfica para identificar los mecanismos de apoyo a la caracterización del *contexto*. La falta de información sobre el *contexto* en el que se llevan a cabo los estudios dificulta su replicación, la interpretación de resultados y la transferencia de conocimientos entre la academia y la industria [39]. Su investigación concluyó que hay pocos estudios en los que se describa el *contexto*.

Hay otros autores que también han señalado la importancia de describir el *contexto* como Sjøberg *et al.* [224] afirmando que los estudios suelen especificar el *contexto* de forma genérica y sólo al referirse a la *validez externa* del experimento. Dyba *et al.* [66] destacan la importancia del *contexto* en el que se lleva a cabo la investigación analizando su influencia en la investigación empírica.

De igual forma, en estudios llevados a cabo en entornos industriales, Jedlitschka *et al.* [116] destacan la importancia a especifica el *contexto* para facilitar la reutilización de la experiencia entre empresas y Petersen *et al.* [177], señalan la necesidad de describir el *contexto* para obtener conclusiones válidas.

En [214] se reflexiona sobre la similitud que debe existir entre una replicación y el experimento original para se considere que se confirman los resultados. En este sentido, pensamos que caracterizar el *contexto* y específicamente clarificar los cambios llevados a cabo puede contribuir a dicho objetivo. Gómez *et al.* [97] al caracterizar los cambios en las replicaciones, señalan los cambios que se realizan para adaptar la replicación a las nuevas condiciones del *contexto*.

4.7 EL FACTOR HUMANO EN LAS REPLICACIONES

En [211], Seaman analizó la influencia del *factor humano* en los experimentos en IS señalando que se necesitan nuevos métodos de investigación para estudiar los aspectos

tos no técnicos de la IS. Propone la incorporación de métodos cualitativos utilizados en otras disciplinas como las ciencias sociales para analizar la complejidad del *comportamiento humano*. Se añade que el *comportamiento humano* es lo suficientemente complejo como para requerir métodos para su estudio.

En [146], Lung *et al.* analizan las dificultades de replicar los experimentos con sujetos humanos caracterizados por la *variabilidad del comportamiento humano*, variables no controladas y el posible sesgo de los investigadores, añadiendo que para superar estas limitaciones, los estudios deberían ser replicados con un diseño mejorado o incluso con un método de investigación diferente.

En 2015, Falcão *et al.* [69] llevaron a cabo una revisión para analizar experimentos con sujetos humanos publicados entre 2003 y 2013 en *Empirical Software Engineering Journal* (EMSE), en *International Conference on Empirical Software Engineering and Measurement* (ESEM), y en *International Conference on Empirical Assessment & Evaluation in Software Engineering* (EASE). Su análisis incluyó, entre otros, información del tipo de replicación, investigadores e instituciones que publicaron experimentos, replicaciones y tipos de sujetos involucrados en los experimentos. El 78,2% de los participantes eran estudiantes, 10,2% profesionales y 11,1% estudiantes y profesionales. Sorprende que en el 41% de los estudios analizados no se describen las características de los participantes, tales como, sexo, edad o la experiencia y en el 17,2% se describe sólo parcialmente. Según los autores del estudio, esta falta de información, sobre todo el nivel de experiencia en la tarea a desarrollar en el experimento, puede afectar a las conclusiones.

4.8 DIFICULTAD PARA REALIZAR Y PUBLICAR REPLICACIONES

Carver *et al.* [43] señalan que la replicación de estudios empíricos en Ingeniería del Software aún no ha atraído la suficiente atención de los investigadores y en [136], Kitchenham *et al.* destacan la falta de incentivos para llevar a cabo replicaciones y la percepción en los investigadores de la dificultad de su publicación a pesar de su importancia como método científico. Al esfuerzo y los recursos necesarios para llevar a cabo replicaciones se une la dificultad de publicar replicaciones aisladas señalada por Kitchenham [135].

Da Silva *et al.* [57] destacan que el aumento de estudios incluyendo la publicación

del experimento original junto con la replicación puede ser debido a la dificultad de realizar y publicar replications aisladas aunque también puede deberse a la necesidad de validar los resultados mediante su replicación y obtener así resultados fiables y listos para su publicación. Añadiendo que los investigadores no tienden a publicar resultados *no confirmatorios* sobre todo de su propia investigación ya que consideran que no son susceptibles de ser publicados.

4.9 DISCUSIÓN

A continuación, se relaciona cada problema detectado en la realización de replications con la plantilla que se propone en la presente memoria para especificar los cambios de las replications, analizando como dicha plantilla, ayuda, en parte, a paliar el problema.

- Falta de unicidad de criterios o específicamente guías para reportar las replications. La plantilla sirve de guía, asegurando que se reporta la información relevante *sin que se olvide ningún dato de interés*, tales como objetivo de la replicación, descripción de cada cambio junto con su motivación, dimensión afectada y amenaza a la validez.
- Carencia y debilidades de los paquetes de laboratorio. Al cumplimentar la plantilla, se almacena la información relativa a cada replicación en el paquete de laboratorio. De esta forma, se puede conocer la *evolución del experimento* ya que todos los cambios llevados a cabo en cada replicación se actualizan en el paquete de laboratorio siendo su información más completa.
- Dificultad para transmitir el conocimiento tácito. La plantilla permite explicitar el conocimiento tácito, tal como el *motivo del cambio* del que, en otro caso, no se suele dejar constancia.
- Falta de acuerdo sobre la terminología. En la plantilla se estructura la información en forma de tabla y se utilizan patrones lingüísticos que permiten elegir entre varias opciones predefinidas para facilitar su redacción. Permite paliar, en parte, el problema de la diferente terminología.
- Dificultad para clarificar el contexto de las replications. La plantilla facilita la caracterización del contexto ya que cada vez que se produce un *cambio* se des-

cribe la *situación en el experimento base* y la *situación en la replicación* junto con sus implicaciones en las amenazas a la validez y dimensiones afectadas.

- Complejidad del factor humano. La especificación de los cambios tanto en la población (cuando se trata de sujetos humanos) como en los experimentadores permite *analizar la posible influencia* del comportamiento humano en los resultados.
- Alto grado de esfuerzo y recursos necesarios para llevar a cabo un experimento. Al conocer los cambios ya realizados se promueven nuevos cambios y por tanto nuevas replicaciones. Por otro lado, los cambios necesarios para adaptar el experimento al nuevo entorno pueden ser ya conocidos si se ha realizado el experimento en un entorno similar en alguna replicación previa. En ambos casos se facilita el diseño de nuevas replicaciones.

4.10 RESUMEN

Las replicaciones juegan un papel importante en la verificación empírica por lo que se recomienda la replicación de estudios empíricos pero, tal como señalan Lung *et al.* [146], rara vez se practican. En las anteriores secciones se han identificado algunos de los problemas que contribuyen de forma conjunta al bajo número de replicaciones y que se pueden resumir en: *i)* falta de unicidad de criterios para reportar las replicaciones; *ii)* carencia de los paquetes de laboratorio; *iii)* dificultad para transmitir el conocimiento tácito; *iv)* falta de acuerdo sobre la terminología; *v)* dificultad para clarificar el contexto de las replicaciones; *vi)* la complejidad de utilizar sujetos humanos; y todo esto unido a: *vii)* el alto grado de esfuerzo y recursos necesarios.

La especificación de los cambios mediante la plantilla que se propone en la presente memoria, puede contribuir a paliar parte de los problemas identificados. En particular, el *conocimiento tácito* relativo a la descripción del cambio y a su motivación. Es información que el investigador posee y de la que no suele dejar constancia. Sin embargo, este *conocimiento tácito* puede ser revelado por medio de la especificación del cambio mediante la plantilla, ya que la motivación forma parte de la descripción del cambio.

Debido al espacio ocupado por la descripción completa de los cambios en el informe de replicación, sería conveniente que se incluyera en el *paquete de laboratorio*. De esta forma se permite conocer cómo ha evolucionado un experimento original dentro de una familia, por qué se han realizado cambios y de qué tipo son, facilitando así la planificación de nuevos cambios al dejar constancia de los cambios ya realizados.

El primer problema señalado, la falta de guías para reportar repeticiones, es el que ha servido como base para la propuesta de la presente tesis haciendo hincapié en la especificación de los cambios que caracterizan a las repeticiones.

ESPECIFICACIÓN DE LOS CAMBIOS EN REPLICACIONES

Nada es permanente a excepción del cambio

Heráclito de Éfeso (540 a. C.– 480 a. C.),

Filósofo griego

En este capítulo se responde a la pregunta de investigación **PI₃** sobre cómo se están reportando actualmente los cambios y la posible influencia del tipo de replicación (interna, externa) e inclusión del experimento original en el reporte de los cambios.

En la sección §5.1 se comentan algunos factores que pueden influir en la especificación de los cambios. En la sección §5.2 se analiza y comparan las características de los estudios seleccionados para conocer su estructura y contenido en general y específicamente de sus cambios. En la sección §5.3 se clasifican los estudios según incluyen o no el reporte del original y por tipo de replicación. En la sección §5.4 se interpretan los resultados obtenidos. Por último, la sección §5.5 resume el capítulo.

5.1 INTRODUCCIÓN

Antes de presentar una propuesta para especificar los cambios, es necesario revisar los artículos de replications para conocer cómo se están reportando realmente los cambios. Los estudios que vamos a analizar son un subconjunto de los seleccionados en la sección §3 para conocer el estado de las replications.

En la bibliografía consultada hay varias propuestas que reflexionan sobre las peculiaridades de las replications [9, 29, 39, 42, 57, 59, 69, 141, 201, 215, 220, 225] identificando los principales problemas. Sin embargo, la única propuesta para reportar replications expresamente, haciendo hincapié en los cambios y en describir la situación en el original y en la replicación, es la guía de Carver [42]. Por este motivo, revisar los estudios que reportan replications permite analizar si los cambios se suelen reportar en texto convencional o existen estructuras o recomendaciones de cómo hacerlo (por ejemplo con ayuda de tablas) y, en general, conocer la información proporcionada para describir el cambio.

Interesa conocer la estructura del informe de replicación, revisar si los cambios se describen en una sección específica y si hay también una sección aparte para describir el original. Es posible que el experimento original esté ya publicado y en ese caso, posiblemente se describe de forma somera o incluso puede ser que no se describa. En caso de reportar la replicación junto con el original, nos interesa conocer si los cambios o diferencias entre ambos se especifican de forma distinta a cuando únicamente se reporta la replicación.

Es posible que el tipo de replicación también afecte a la descripción de los cambios. En caso de replications externas, el original es menos conocido que en replications internas ya que las externas son llevadas a cabo por experimentadores diferentes. Hay parte del conocimiento que el investigador queda implícito en la publicación [217] y esta dificultad para transmitir el *conocimiento tácito* dificulta la especificación de los cambios.

5.2 REVISIÓN DE LOS ESTUDIOS DE REPLICACIÓN

Para conocer en detalle la información sobre los cambios, si se incluye la descripción del experimento original, la información suministrada sobre las replications, las amenazas a la validez y otras posibles secciones de interés, es necesario acceder, prácti-

camente, al texto completo del estudio, por lo que, de los estudios identificados en la sección §3, se han seleccionado únicamente los estudios publicados en los dos principales foros del área de IS empírica: *Empirical Software Engineering Journal (EMSE)* y *International Conference on Empirical Software Engineering and Measurement (ESEM)*.

En primer lugar, respecto a la terminología utilizada, hay que analizar si en los estudios seleccionados se utilizan otros términos para referirse a los cambios introducidos aparte de *change*. Se ha encontrado que es frecuente utilizar términos tales como *differences* que aparece en [18, 189, 208] o *except* en [102]. Estos términos son equivalentes entre sí y referidos a las modificaciones introducidas en las replications. Aunque no está entre los estudios revisados, otro término para referirse a los cambios es *recipe* utilizado por Brooks *et al.* en [34].

La tabla §5.1 resume el resto de las características analizadas en los 25 estudios de replicación seleccionados:

- *Método empírico*. 19 estudios han utilizado el método empírico *experimento controlado* (EX), 2 son *cuasi-experimentos* (QE), 2 *encuestas* (SV) y 2 *estudios de caso* (CS).
- *Guías seguidas*. En 10 de los estudios (40%) no se cita ninguna guía. Entre los estudios que citan alguna guía, destacan los 11 estudios que siguen la guía de Carver [42] (44% de los estudios).

Hay estudios que dicen seguir más de una guía, uno de los estudios [82] afirma haber seguido las 4 guías principales. Las guías, ordenadas por asiduidad de uso, son:

- Guía de Carver [42]; 11 estudios (9 EX, 1 QE y 1 SV)
- Guía de Wohlin *et al.* [245]; 4 estudios (EX).
- Guía de Jedlitschka *et al.* [118]; 2 estudios (EX).
- Guía de Juristo y Moreno [125]; 1 estudio (EX).
- Guía de Runeson y Höst [196]; 1 estudio (CS).

En la sección §7.7.1, la tabla §7.11 compara la expresividad de las guías aquí identificadas con la plantilla que se propone en esta memoria.

- *Tipo*. Indica el tipo de las replications que se reportan en el estudio. Las replications puede ser *internas* (I) o *externas* (E). Si además se *reporta el experimento original* aparece (O).

Tabla 5.1: Estudios de replicación publicados en 2013–2018 en EMSE y ESEM con sus características

Estudio	Método empírico	Guías seguidas	Tipo	Núm. replic.	Cambios					
					Núm. cambios	Uso de tablas	Motivo cambio	Sección aparte	Amenazas validez	Amenazas cambios
[6]	EX	[42]	I	1	8	No	Sí	Sí	Sí	Sí
[103]	QE	[42]	I	1	2	No	No	Sí	Sí	Sí
[114]	EX	[42]	I	1	3	No	Sí	Sí	Sí	Sí
[139]	EX	[42]	I	1	4	No	No	Sí	Sí	Sí
[189]	EX	[42]	I	3	17	Sí	No	Sí	Sí	Sí
[192]	EX	[118, 245]	I	4	8	No	No	No	Sí	No
[186]	CS	-	I	1	5	Sí	No	Sí	Sí	No
[102]	EX	[245]	O-I	1	2	No	No	No	Sí	No
[208]	EX	[245]	O-I	1	4	No	Sí	Sí	Sí	Sí
[199]	EX	-	O-I	1	4	No	No	Sí	Sí	No
[121]	EX	-	O-I	1	1	No	No	Sí	Sí	No
[155]	EX	-	O-I	1	5	Sí	No	Sí	Sí	No
[12]	EX	[42]	E	1	11	Sí	Sí	Sí	No	-
[90]	EX	[42]	E	1	4	Sí	Sí	Sí	Sí	No
[165]	SV	[42]	E	1	-	-	-	Sí	Sí	-
[87]	EX	[42]	E	1	7	Sí	Sí	Sí	Sí	No
[89]	EX	[42]	E	2	7	Sí	Sí	Sí	Sí	No
[110]	CS	[196]	E	1	-	-	-	Sí	Sí	-
[157]	EX	-	E	5	-	-	-	Sí	Sí	-
[18]	EX	-	E	1	7	Sí	No	Sí	Sí	No
[194]	EX	-	E	2	-	-	-	No	Sí	-
[44]	EX	-	E	1	3	No	No	No	Sí	No
[82]	EX	[42, 118, 125, 245]	O-E	3	-	-	-	No	Sí	-
[81]	SV	-	O-E	9	-	-	-	No	Sí	-
[198]	EX	-	IE	3	12	Sí	No	No	Sí	No
Total 25	20 EX 2 SV 2 CS 1 QE		7 I 5 O-I 10 E 2 O-E 1 IE	1,9 ¹	6 ²	10 No 9 Sí	12 No 7 Sí	7 No 18 Sí	1 No 24 Sí	12 No 6 Sí

¹Número medio de replicaciones.

²Número medio de cambios, considerando sólo los que presentan cambios.

Esto permite agrupar los estudios, de forma similar a [57], y distinguir los siguientes grupos:

- Estudios que reportan sólo replications *internas* (I).
- Estudios que reportan replications *internas* junto con el experimento original (O-I).
- Estudios que reportan sólo replications *externas* (E).
- Estudios que reportan replications *externas* junto con el experimento original (O-E).
- Estudios que reportan sólo replications *externas* e *internas* (IE).

El análisis detallado de estos grupos se muestra en la sección siguiente (sección §5.3).

Referente a las familias de experimentos, los estudios [82, 192] dicen reportar *familias de experimentos* (aparece así definido en dichos estudios) y el estudio [81] reporta una *familia de encuestas*. En el resto de estudios no se habla de *familias*. Por claridad, se ha preferido la clasificación por contenido (tabla §5.2) en lugar de por *familias* ya que hay estudios en que no queda claro que los experimentos reportados constituyan una familia según los criterios de Santos *et al.* [201].

- *Número de replications.* La mayoría de los estudios reportan únicamente una replicación.
 - 12 estudios (48%) presentan una replicación sin el original. De los 12, 5 son replications *internas* y 7 *externas*.
 - 6 estudios (24%) reportan más de una replicación sin el original. 2 son replications *internas* y 3 *externas* y 1 incluye replications *internas* y *externas*.
 - 5 estudios (20%) publican el original junto con una replicación *interna*.
 - Los 2 estudios restantes (8%) publican el original con 3 y 9 replications *externas*.
- *Cambios.* Se analiza: *i) número de cambios* descritos en el estudio, *ii) si la descripción de los cambios se hace mediante el uso de tablas* y *iii) si incluye la especificación del motivo del cambio.*
 - 6 estudios, incluyendo las 2 encuestas (SV), no describen cambios en el experimento original (-). Esto puede deberse a que: *i) no hay cambio*, sólo

cambian los participantes involucrados [82]; y *ii*) los cambios no aparecen en el informe de replicación [201].

- Hay 9 estudios que describen los cambios con la ayuda de tablas (T). De estos, hay 2 estudios [18, 186] que sólo explican los cambios usando tablas, con una fila para describir cada cambio. En el resto, se describen además en lenguaje natural.
 - Los 10 estudios restantes, especifican los cambios únicamente en lenguaje natural con el problema inherente de la ambigüedad.
 - La motivación para el cambio no siempre se describe. En 7 de los 19 estudios (37%), aparece la motivación de cada cambio. En el resto, es decir los 12 restantes, habiendo cambios, no se especifica el motivo de cada cambio
 - 9 de los 11 estudios que siguen la guía de Carver, describen los cambios. Los otros 2 estudios, que no especifican cambios pero siguen las directrices de Carver, son una encuesta (SV) y un (EX) que sigue 4 guías ([42, 118, 125, 245]).
 - Si analizamos los artículos fijándonos en el tipo de replicación, todas las replications *internas* describen sus cambios. Sin embargo, en las replications *externas*, sólo 6 de los 12 estudios describen sus cambios –sin tener en cuenta el estudio que incluye replications *internas* y *externas*–. Esto pone de manifiesto que al cambiar los experimentadores es más difícil identificar los cambios debido al menor conocimiento del original.
 - En los estudios que reportan varias replications, a menudo se explican las diferencias entre las replications, pero no se compara con el original ni se justifica el cambio (por ejemplo [194]).
- *Sección aparte.* Hay estudios en los que el experimento original y las replications se reportan en secciones separadas. Aunque en el estudio no se reporte el experimento original por estar ya publicado, es una práctica común incluir una sección para resumir dicho experimento.
 - 18 estudios separan el original y las replications (denotado Sí).
 - Los 7 estudios restantes, mezclan la descripción del original y las replications (denotado No).
 - 10 de los 11 estudios que siguen la guía de Carver, describen el original y las replications en secciones separadas.

- *Amenazas a la validez.* Generalmente, los estudios analizan la validez *externa, interna, del constructo y de conclusión*. El 96% de los estudios analiza las *amenazas a la validez* (denotado *Sí*), solamente en uno de los estudios no se mencionan.
- *Amenazas cambios.* Por último, analizamos si se relacionan las *amenazas* con los *cambios*. Aparece *Sí* cuando en el estudio se justifica alguno de los cambios como una forma de mitigar algún tipo de *amenaza* [139] o bien se analiza como afecta el cambio a las *amenazas a la validez* [6]. Así por ejemplo, en [139] se presenta una tabla donde se analizan las *amenazas a la validez* en el original y si se han abordado en la replicación llevando a cabo algún cambio o no. En [6] se analiza para algunos de los cambios las *amenazas a la validez* afectadas. En 6 de los estudios se vinculan las *amenazas a la validez* con los *cambios* llevados a cabo.

5.3 INFLUENCIA DEL TIPO DE REPLICACIÓN EN LA ESPECIFICACIÓN DE LOS CAMBIOS

En esta sección se presentan los estudios agrupados según el tipo de replicación (*interna o externa*) y la inclusión o no del original.

La tabla §5.2 muestra para cada grupo: *i*) número de estudios, *ii*) número de estudios que sigue la guía de Carver, *iii*) otras guías seguidas, *iv*) número de estudios que no siguen ninguna guía, *v*) proporción de estudios que especifican los cambios y *vi*) proporción de estudios que describen el experimento original y las replications en secciones específicas.

Tabla 5.2: Estudio de replicación publicados en 2013–2018 en EMSE y ESEM agrupados por tipo

Tipo	Núm. estudios	Núm. estudios que siguen Carver [42]	Otras guías	Núm. estudios sin guías	Proporción estudios con cambios	Proporción estudios con secciones separadas
I	7	5	[118, 245]	1	7/7	6/7
O-I	5	0	[245]	3	5/5	4/5
E	10	5	[196]	4	6/10	8/10
O-E	2	1	[118, 125, 245]	1	0/2	0/2
IE	1	0		1	1/1	0/1
Σ	25	11		10	17/25	18/25

Según el tipo:

- *E.* El mayor número de estudios (10 de 25) [12, 18, 44, 87, 89, 90, 110, 157, 165, 194]

reportan replicaciones *externas* de las que ya está publicado el estudio original. En 5 de ellos, se sigue la guía de Carver. Solo 6 presentan cambios y 8 tienen secciones específicas para la descripción del experimento original.

En los 5 que siguen a Carver, los cambios están especificados en secciones aparte y en 4 de ellos utilizando además tablas. En 4 de estos 5 estudios se reporta una única replicación.

Al ser replicaciones externas el experimento original es menos conocido por lo que sólo en 6 de los 10 estudios se describen los cambios, 5 de ellos mediante la guía de Carver.

- *I.* Hay 7 estudios (de 25) [6, 103, 114, 139, 186, 189, 192] que reportan únicamente replicaciones *internas*. En 5 se sigue la guía de Carver [42]. Los 7 presentan cambios y en 6 de ellos tienen secciones específicas para la descripción del experimento original y la replicación.

En este grupo es donde mayoritariamente se siguen la guía de Carver y, en consecuencia, están bien estructurados con clara definición de los cambios, algunas veces con tablas. Cuentan con secciones específicas para describir el original aunque de forma resumida ya que no se reporta por estar ya publicado.

- *O-I.* Hay 5 estudios (de 25) [102, 121, 155, 199, 208] con replicaciones *internas* y reportando el original. En ninguno se sigue la guía de Carver. Todos presenta cambios y 4 tienen secciones específicas para la descripción del experimento original y la replicación.
- *O-E.* Sólo hay 2 estudios (de 25) [81, 82] de replicaciones *externas* y reportando el original. Es poco frecuente que se reporte el experimento original junto con replicaciones *externas*, ya que los experimentadores son distintos. En estos dos estudios no se especifican los cambios ni se especifican secciones separadas para el original. En uno de ellos no se siguen ninguna guía y en el otro se siguen 4 guías.

En ambos estudios se incluye más de una replicación (3 y 9 replicaciones en cada estudio).

- *IE.* En este estudio (1 de 25) [198] se combinan replicaciones *internas* y *externas* y el original únicamente se referencia. No sigue ninguna guía. Reporta los cambios aunque sin sección específicas para resumir el experimento original.

5.4 DISCUSIÓN Y HALLAZGOS

Para saber cómo se reportan los cambios, se ha analizado: *i)* si los cambios se describen en lenguaje natural o a través de tablas; *ii)* qué información se cuenta del cambio; y *iii)* en qué secciones del estudio se describen:

- *La mayoría de los estudios (19 de 25, 76%) especifican sus cambios ya sea en forma textual o con la ayuda de tablas.*
- *Para cada cambio, se suele especificar la situación en el original y en la replicación.* Hay estudios que describen los cambios en lenguaje natural, sin la ayuda de tablas y con el problema de la ambigüedad del lenguaje natural (por ejemplo [6]). En solo 7 de los 19 estudios que presentan cambios se informa de la *motivación del cambio* (37%) (por ejemplo [103, 139]). Consideramos que el motivo es importante para conocer la causa del cambio y saber si esa misma situación se repite en un nuevo entorno.
- *Hay estudios que describen los cambios usando tablas.* 9 de los 19 estudios que presentan cambios utilizan tablas (47%) para facilitar la descripción de los cambios.
- *Las tablas presentan una estructura común, la propiedad cambiada (el cambio) se define en las filas. En las columnas aparece el valor de dicha propiedad en el original y para cada replicación* (por ejemplo [189]). En [198] se utilizan dos tablas para describir los cambios. En la primera tabla compara los distintos *niveles* que toma cada *factor* (*variable independiente*) en cada una de las replicaciones. La segunda tabla, relativa a las características de los *sujetos*, muestra los valores en cada replicación.

El *motivo* para llevar a cabo el cambio no suele estar incluido en las tablas.

La estructura de la tabla presentada en [88, 155] es diferente de las anteriores. El cambio se describe analizando las *dimensiones* identificadas en [97]. En la plantilla propuesta en la sección §6.3 se utilizan dichas *dimensiones* y los *elementos de la configuración experimental* asociados como parte de la descripción de los cambios.

- *Al describir la replicación, es frecuente incluir una sección de cambios al estudio original* (por ejemplo [114]). En 18 de los 25 estudios (72%) se describen el original y las replicaciones en secciones aparte. Lógicamente, cuando sólo se reportan replicaciones, el original se describe de forma resumida.

- *Hay estudios que describen los cambios en varias secciones ya que los cambios se especifican a medida que se va describiendo el proceso experimental de la replicación (por ejemplo, [198]).*

Referido al contenido de los estudios:

- *Aunque hay igual número de estudios que reportan replicaciones *externas* que *internas*, teniendo en cuenta si se incluye o no el experimento original, el mayor número de estudios (10 de 25) corresponde a replicaciones *externas* de las que ya está publicado el estudio original.*

Los hallazgos encontrados en relación con el tipo de replicación y el uso de guías son:

- *Los cambios se especifican con más frecuencia en las replicaciones *internas* que en las *externas*. Los cambios se describen en el 100% de los estudios con replicaciones *internas* y sólo en el 50% de los estudios con replicaciones *externas*. Puede deberse a que en las replicaciones *externas* es más difícil especificar los cambios debido a que los experimentadores son distintos y al hecho de que hay una falta de conocimiento del experimento original ya que el *paquete de laboratorio* no suele hacerse público [201].*

No obstante, conocer los cambios en las replicaciones *externas* es tan necesario como en las replicaciones *internas* ya que favorece la comprensión del experimento original y cómo ha evolucionado a lo largo del tiempo.

- *La guía de Carver es seguida en el 44% de los estudios, principalmente cuando el tipo de replicación es *interna*, el original no se reporta y especialmente para reportar una sola replicación. El resto de guías utilizadas [118, 125, 196, 245] no están indicadas para replicaciones, por lo tanto, se han adaptado.*
- *En la revisión bibliográfica llevada a cabo por Da Silva *et al.* [57], se encontró que el 74% eran replicaciones *internas* y de éstas el 64% se reportaban junto con el original lo que llevaba a pensar que el objetivo de la replicación era confirmar el que los resultados observados en el experimento original no son fruto de la casualidad [43]. En nuestro caso, el 50% son replicaciones *internas* y de éstas sólo el 42% se reportan junto con el original lo que representa una evolución a publicar las replicaciones de forma independiente que puede ser debido al interés creciente en la práctica de la replicación en sí y dando a la replicación y a su publicación la relevancia que merece.*

5.5 RESUMEN

Con el objetivo de conocer como se especifican los cambios, se han analizado 25 estudios que reportan al menos una replicación.

Se ha encontrado que los cambios suelen reportarse en lenguaje textual y en algunos casos con ayuda de tablas. La guía más utilizada para *experimentos controlados*, *cuasi-experimentos* y *encuestas* es la de Carver [42]. También se utilizan las guías de Wohlin *et al.* [245], Jedlitschka *et al.* [118], Juristo y Moreno [125]. Para *estudios de caso* se utiliza la guía de y Runeson y Höst [196].

El contenido más frecuente en los estudios de replicación analizados es una única replicación *externa*. También es frecuente reportar una única replicación *interna* aunque, en algunos casos, se reportan junto con el experimento original.

Al igual que en la plantilla propuesta en la sección §6.3, es frecuente analizar las *amenazas a la validez* del experimento y hay estudios que describen los cambios identificando la *dimensión afectada* por el cambio [97].

Para describir el cambio se suele especificar la situación en el experimento y en la replicación. El motivo del cambio no siempre se especifica.

Es más frecuente describir los cambios cuando se trata de replications *internas*. En las replications *externas* se conoce menos el original ya que los experimentadores son otros.

La estructura del estudio de replicación, suele incluir secciones específicas para describir el experimento original y para los cambios.

Es de destacar el impacto que ha tenido la guía de Carver [42], con 109 citas en Google Scholar (febrero 2020) lo que desde nuestro punto de vista pone de manifiesto la necesidad de propuestas específicas para el proceso de replicación que complementen las guías generales para experimentación, considerando las particularidades de las replications.

PARTE III

PROPUESTA

ESPECIFICACIÓN DE LOS CAMBIOS MEDIANTE PLANTILLAS Y PATRONES

Experimento es aquella clase de experiencia científica en la cual se provoca deliberadamente algún cambio y se observa e interpreta su resultado con alguna finalidad cognoscitiva

*Mario Augusto Bunge(1919 –),
Físico, filósofo y epistemólogo argentino*

En este capítulo se presenta la contribución principal de esta memoria cuyo objetivo principal es aumentar el beneficio que se obtiene mediante las replicaciones. El enfoque propuesto, respondiendo a la pregunta de investigación PI_4 , sistematiza la concepción de los cambios y la manera de documentarlos facilitando la comprensión de la replicación en cuestión y de las sucesivas replicaciones.

En la sección de introducción §6.1 se destacan las cualidades de la plantilla aquí propuesta. En la sección §6.2 se formaliza la información que constituye el metamodelo. En la sección §6.3 se propone la plantilla como representación visual de dicho metamodelo. En la sección §6.4 se identifican las situaciones y tipos de replicaciones que mejor se adaptan al uso de plantilla. En la sección §6.5, se presenta una herramienta para especificar los cambios basado en la propuesta. Finalmente, en la sección §6.6 se resume el capítulo.

6.1 INTRODUCCIÓN

Al diseñar una replicación, habitualmente surge la necesidad de introducir cambios y tal como se indica en [96], cualquier cambio, por pequeño que sea, puede alterar los resultados del experimento y dificultar su comprensión.

El objetivo de este trabajo es facilitar la especificación de dichos cambios proponiendo: *i*) un metamodelo para formalizar la información involucrada, *ii*) una plantilla que facilite la reusabilidad, la representación visual y evite redundancias y *iii*) patrones lingüísticos que faciliten la redacción de los cambios.

El uso de la plantilla persigue una doble finalidad, por una lado, invita al investigador a hacer explícito el cambio y sus detalles (reduciendo el *conocimiento tácito*) y, por otro, ayuda al lector a comprender mejor la replicación y a seguir la traza de cómo evoluciona el experimento original en la sucesión de experimentos de una familia. Las familias de experimentos permiten generalizar resultados más allá del alcance de experimentos individuales proporcionando evidencias para confirmar o rechazar hipótesis específicas [55].

Sistematizar la concepción de los cambios y la manera de documentarlo facilitará además, nuevas replicaciones.

6.2 METAMODELO DE REPLICACIONES Y CAMBIOS

En esta sección se detalla la información involucrada en la replicación y que servirá como base para la definición del metamodelo.

Para identificar tanto la replicación como el experimento base se utiliza un código ó *acrónimo* relativo al experimento base. Es especialmente útil para la descripción de un conjunto de replicaciones del mismo experimento. El experimento base (o de referencia) puede ser un experimento original o una replicación anterior.

Para describir el objetivo del experimento base, parece razonable registrar el *Goal-Question-Metric (GQM)* [20] o, en su defecto, se describirá el objetivo. También se registrará una breve descripción del experimento.

Respecto del contexto, es relevante dejar constancia del *lugar* y *fecha* de realización tanto del experimento base como de la replicación.

El *tipo* de la replicación puede ser *interna* o *externa*. No se han considerado otras taxonomías de las replicaciones por la falta de acuerdo en la nomenclatura (ver sección §2.4).

El *propósito de la replicación* puede ser: *i) Confirmar resultados, ii) Generalizar resultados y iii) Superar algunas limitaciones del experimento base.*

A continuación, nos centraremos en identificar la información relacionada con la especificación de los *cambios* que caracterizan a la replicación. Es decir, además del nombre del cambio, describir en qué consiste el cambio y la causa, identificar los *elementos de la configuración* afectados y las *amenazas a la validez* abordadas.

Según el *elemento de la configuración experimental* modificada, los cambios se pueden clasificar en *dimensiones*. Las *dimensiones* propuestas, basadas en Gómez *et al.* [97], aparecen descritas en la sección §2.2.7. A continuación se enumeran los *elemento de la configuración experimental* que se pueden modificar en cada *dimensión* dependiendo del cambio realizado:

- *Operacionalización.* Incluye cambios relacionados con los elementos: *i) causa, ii) efecto y iii) procedimiento de medida.*
- *Población.* El elemento afectado por el cambio, es la *propiedad* de los *sujetos o unidades experimentales* que cambia. Será necesario identificar la *propiedad* que se modifica.
- *Protocolo.* Incluye cambios relacionados con los elementos: *i) diseño experimental, ii) material experimental, iii) guías, iv) instrumentos de medición y v) técnicas de análisis de datos.* El *protocolo experimental* es el planteamiento de estos elementos para observar los efectos de los tratamientos [123].
- *Experimentadores.* Referido a los cambios en los roles de los experimentadores. Los elementos son: *i) diseñador, ii) analista, iii) formador, iv) monitor y v) medidor.*
- *Contexto.* El elemento afectado por el cambio, es una *variable de contexto* que será necesario identificar.

En la sección §2.2.7 se han visto ejemplos de cambios que afectan a estas *dimensiones* y específicamente a los elementos de la configuración experimental.

Asimismo, cada *cambio* puede que no afecte a la validez o afectar (incrementar o disminuir) a uno o varios tipos de *amenazas a la validez: interna, externa, de conclusión y*

del constructo. La relación de los cambios con las amenazas a la validez se analiza en la sección §2.3.1.

Una vez conceptualizada la información sobre las replicaciones y sus cambios, se propone un metamodelo para su representación. La figura §6.1 presenta la versión actual del metamodelo utilizando la notación de diagramas de clases UML.

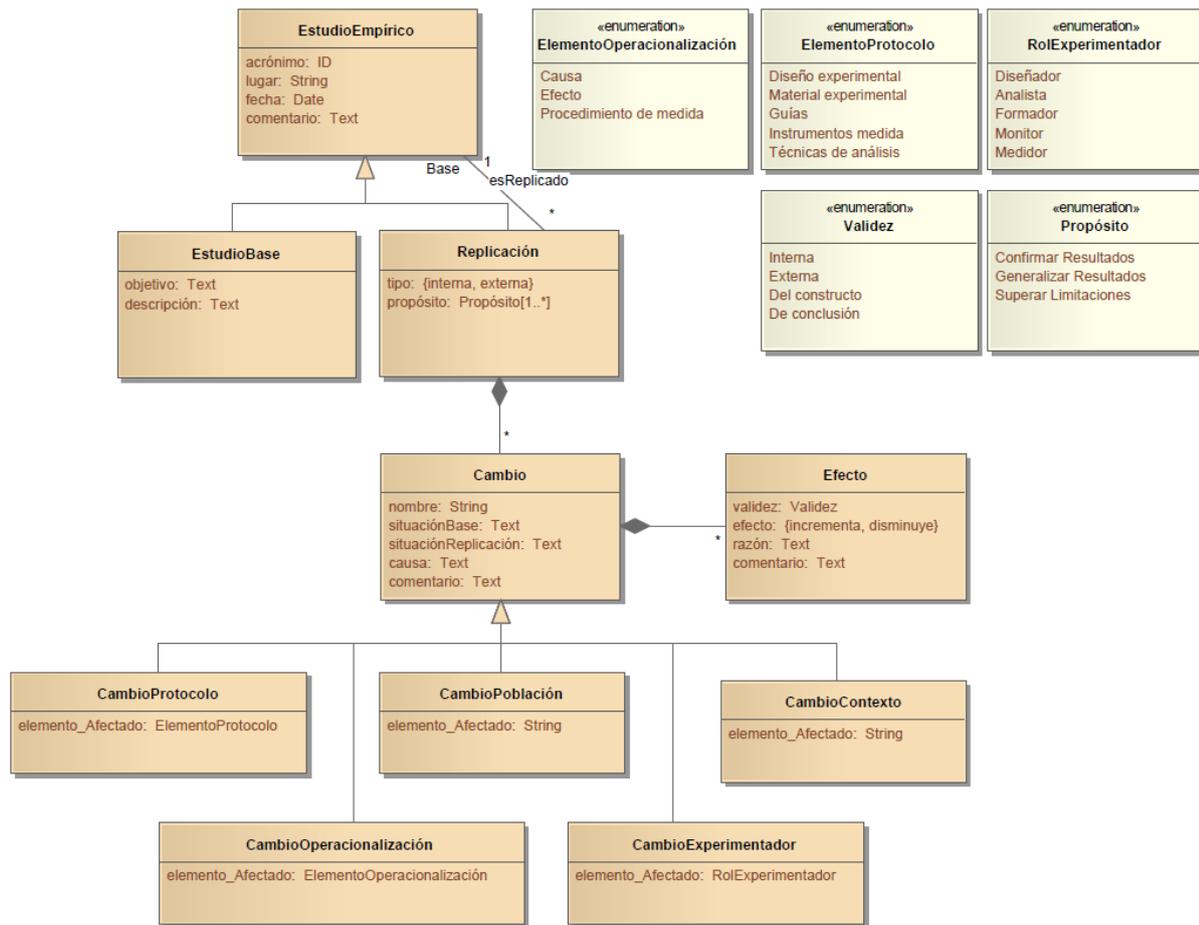


Figura 6.1: Metamodelo de replicaciones y cambios

6.3 PLANTILLA PARA ESPECIFICAR LOS CAMBIOS

Mediante el metamodelo descrito en el apartado anterior, se ha identificado la información involucrada en la descripción de los cambios. El siguiente paso es estructurar dicha información en una plantilla que presente los datos de forma fija y sirva de guía para evitar que falte información relevante.

Para algunos campos de las plantillas se han identificado frases que son habituales y se han parametrizado. Estas frases se denominan patrones lingüísticos o abreviadamente *patrones-L* [64, 65, 235].

En la notación usada para describir los patrones-L:

- Las palabras o frases entre < y > deben ser convenientemente reemplazadas.
- Las palabras o frases que se encuentren entre { y } y separadas por | representan opciones de las que se debe escoger una.
- Las palabras entre [y] son opcionales, es decir, pueden aparecer o no cuando la plantilla es instanciada.
- Las palabras entre [y]⁺ se repiten 1 o más veces; es decir al menos aparece una vez pero puede haber más de una ocurrencia.

Rellenar los espacios en blanco en las oraciones preestablecidas, es decir, en los patrones-L, es más fácil y rápido que escribir un párrafo completo.

La tabla §6.2 muestra la plantilla propuesta.

La figura §6.3 muestra la correspondencia 1:1 que existe entre el metamodelo y cada campo de la plantilla.

La plantilla consta de dos partes claramente diferenciadas: una parte general para la especificación de los aspectos básicos (o generales) de la replicación y una parte específica para cada uno de los *cambios* incluidos en la replicación.

6.3.1 Especificación de los aspectos generales de las replicaciones

El significado de cada uno de los campos junto con el patrón-L utilizado se muestra a continuación.

- *Replicación*. Permite definir el *acrónimo* de la replicación y del experimento base. Se puede seleccionar entre replicación *interna* o *externa* y entre *experimento base* u otra *replicación previa* para definir la replicación. El patrón-L quedaría:

<Acrónimo de la replicación> replicación { Interna | Externa } basada en
{ experimento base | replicación previa } <Acrónimo del experimento base>

Replicación	<Acrónimo de la replicación> replicación { <i>Interna</i> <i>Externa</i> } basada en { <i>el experimento base</i> <i>la replicación previa</i> } <Acrónimo del experimento base>
Objetivo	{ <GQM> <Objetivo del experimento> }
Descripción	<Descripción del experimento>
Lugar y Fecha	{ <i>El experimento base</i> <i>La replicación previa</i> } se llevó a cabo en <lugar> en <fecha> y esta replicación en <lugar> en <fecha>
Propósito de la replicación	[{ <i>Confirmar resultados</i> <i>Generalizar resultados</i> <i>Superar limitaciones del experimento base</i> }] +
[Comentario	<Comentario>]
Cambio#<i>	<Nombre del cambio> Originalmente, <descripción de la situación en el experimento base> En la replicación, <descripción de la situación en la replicación> { <i>Con el fin de</i> <i>Causado por</i> <i>Ya que</i> } <causa del cambio>
Dimensión Modificada	{ Operacionalización , específicamente, { <i>causa</i> <i>efecto</i> <i>procedimiento de medida</i> } Población , específicamente, <nombre de la propiedad> Protocolo , específicamente, { <i>diseño experimental</i> <i>material experimental</i> <i>las guías</i> <i>instrumentos de medida</i> <i>técnicas de análisis de datos</i> } Experimentador , específicamente, { <i>diseñador</i> <i>analista</i> <i>formador</i> <i>monitor</i> <i>medidor</i> } Contexto , específicamente, <variable de contexto> }
Amenazas a la validez	[El cambio { ({ <i>incrementa</i> <i>disminuye</i> } la validez { <i>interna</i> <i>externa</i> <i>del constructo</i> <i>de conclusión</i> }) <i>no afecta a la validez</i> } { <i>debido a</i> <i>ya que</i> } <razón >] +
[Comentario	<Comentario>]

Figura 6.2: Plantilla para la especificación de cambios en replications

- *Objetivo del experimento.* GQM del experimento base o, en su defecto, descripción del objetivo del experimento.

{<GQM> | <Objetivo del experimento>}

- *Descripción del experimento.* Breve descripción del experimento.

<Descripción del experimento>

- *Lugar y Fecha.* Datos relativos al contexto del experimento base o de la replicación previa y de la presente replicación.

{ El experimento base | La replicación previa }
se llevó a cabo en <Lugar> en <Fecha>
y esta replicación en <Lugar> en <Fecha >

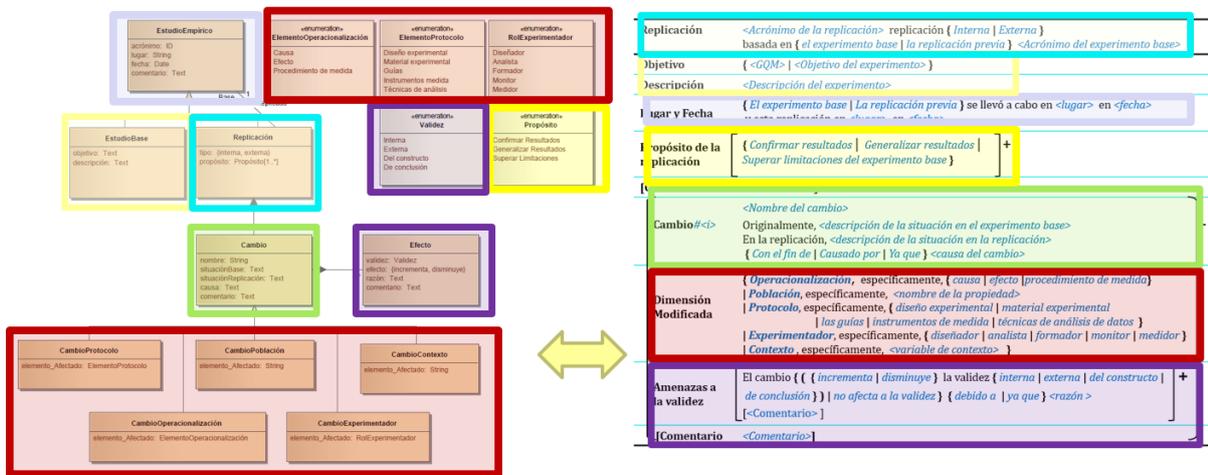


Figura 6.3: Correspondencia metamodelo plantilla

- *Propósito.* Permite seleccionar entre tres valores posibles.

{ Confirmar resultados | Generalizar resultados |
Superar limitaciones del experimento base }

Hay replicaciones que pueden tener más de un propósito por lo que el patrón aparece entre [y]⁺.

- *Comentario.* Información adicional sobre los aspectos generales de la replicación que pueda ser de interés para el lector. Es opcional y de texto libre.

[<Comentario>]

6.3.2 Especificación del propósito de la replicación

En la plantilla que se propone en la presente tesis, se ha concretado el propósito de las replicaciones en: *i) Confirmar resultados, ii) Generalizar resultados y iii) Superar algunas limitaciones del experimento base.*

Esta decisión viene motivada por el estudio de las propuestas de Wohlin *et al.* [245] y de Gómez *et al.* [97]. A saber, Wohlin *et al.* afirman que el propósito de una replicación es mostrar que el resultado del experimento original es válido para una población mayor y por lo tanto, solo se refieren a *generalizar resultados*. Por otro lado, Gómez *et al.* [97] señalan que el hecho de que haya diferentes tipos de replicaciones implica que cada tipo debe tener un propósito que, a su vez, determina los cambios que serán necesarios introducir al llevar a cabo dicha replicación.

Según esto, asociadas a la dimensión experimental afectada por el cambio, Gómez *et al.* [97] identifican 6 funciones en las replicaciones: *i)* Controlar el error de muestreo, *ii)* Controlar la independencia del protocolo, *iii)* Conocer los límites de la operacionalización, *iv)* Conocer los límites de la población, *v)* Controlar la independencia de los experimentadores y *vi)* Validar hipótesis.

A cada una de las funciones se le puede asociar uno o más de los propósitos identificados en la plantilla propuesta. De esta forma:

- *Controlar el error de muestreo.* El cambio consiste en aumentar el tamaño de la muestra. El propósito es *superar limitaciones detectadas en el experimento original*.
- *Controlar la independencia del protocolo.* Si se varía el protocolo y los resultados coinciden, demuestra que son independientes del protocolo seguido. El propósito es por tanto *confirmar resultados* y también *generalizar resultados* ya que demuestra que son válidos, es decir, también coinciden los resultados, para las nuevas condiciones.

Relacionado con el protocolo, es posible que el propósito de la replicación sea *superar limitaciones detectadas en el experimento original*, por ejemplo, cuando se cambia el diseño debido a que la asignación de sujetos al tratamiento en el experimento base no ha sido aleatoria. El *diseño experimental* es un elemento del protocolo experimental.

- *Conocer los límites de la operacionalización.* Trata de comprender hasta dónde se cumplen los resultados cambiando el tratamiento, métricas, procedimientos de medición o momento de realización del experimento. Por tanto, tiene el propósito de *generalizar resultados*.

Relacionado con la operacionalización, es posible, que el propósito de la replicación sea *superar limitaciones detectadas en el experimento original*, por ejemplo cuando se cambia algún *nivel* de la variable independiente al detectar que hay niveles del tratamiento que corresponden a concentraciones de un reactivo químico que es tóxicos para la planta y por tanto no se debe utilizar. Otro ejemplo sería un experimento realizado en un momento del tiempo que puede comprometer los resultados y mediante una replicación se quiere superar dicha limitación.

- *Conocer los límites de la población.* Si se varía la población y los resultados coinciden demuestra que son válidos para la nueva población y el propósito de la replicación es *generalizar resultados* y por tanto también *confirmar resultados*. En el caso de

que la nueva población difiera por ejemplo en la *experiencia*, es posible estudiar si la *experiencia* es una condición que afecta a los resultados.

- *Controlar la independencia de los experimentadores.* Los cambios en los investigadores (o sus roles) permiten comprobar si los resultados pueden estar sesgados por los investigadores. Al cambiar los investigadores se trata de una replicación *externa*. El propósito es, por tanto, *confirmar resultados*.
- *Validar hipótesis.* Solo se mantiene la hipótesis. El propósito suele ser *confirmar resultados*.

En nuestra propuesta, se han agrupado estas 6 funciones para conseguir una terminología más general y por tanto más fácil de utilizar, entendiendo que cada uno de los propósitos identificados puede incluir varias de las funciones señaladas en [97]. A nivel de definición del cambio si se contemplan las 6 funciones pero a nivel de replicación se engloban en los 3 principales propósitos identificados. De esta forma:

Cuando el propósito es *confirmar resultados*, se replica el experimento lo más estrechamente posible verificando que los resultados no son accidentales.

Cuando el propósito es *generalizar resultados*, se suele asociar con la *validez externa* que según Wohlin *et al.* [245] puede verse afectada por el *diseño* elegido además de por los *sujetos y objetos experimentales*, señalando que hay tres riesgos principales para la *validez externa* y por tanto para poder *generalizar resultados*: tener participantes equivocados como sujetos, realizar el experimento en el entorno equivocado y en un momento del tiempo que afecte a los resultados.

Por último, las replications con el propósito de *superar algunas limitaciones encontrada en el experimento base* suele ser consecuencia de alguna *amenaza a la validez experimental* detectada en el experimento base.

6.3.3 Especificación de los cambios de las replications

Cada replicación puede incorporar uno o más *cambios*. Para cada *cambio*, es necesario especificar:

- *Definición de los cambios de la replicación.* Los cambios se numeran secuencialmente, tienen un nombre descriptivo y se especifican mediante un patrón-L que debe completarse. Para describir correctamente cada cambio, es necesario especificar

la situación original en el experimento base, la nueva situación en la replicación y finalmente se justifica la causa o consecuencia del cambio. En esta última parte del patrón, se puede elegir entre *con el fin de*, *causado por* y *ya que* para completar la oración.

<Nombre del cambio>

Originalmente, <descripción de la situación en el experimento base>

En la replicación, <descripción de la situación en la replicación>

{ con el fin de | causado por | ya que } <causa del cambio>

- *Dimensión modificada y elementos*. El patrón-L se completa eligiendo una de las *dimensiones* propuestas y concretar el *elemento modificado* eligiendo alguna de las opciones presentadas o bien texto para las dimensiones *Población* y *Contexto*.

{ **Operacionalización**, específicamente,

{ causa | efecto | procedimiento de medida } |

Población, específicamente, <nombre de la propiedad > |

Protocolo, específicamente,

{ diseño experimental | material experimental | las guías | instrumentos de medida | técnicas de análisis de datos } |

Experimentador, específicamente,

{ diseñador | analista | formador | monitor | medidor } |

Contexto, específicamente, <variable de contexto > }

- *Amenazas a la validez*. El patrón-L se completa con el *tipo de amenaza* y el efecto del cambio (*aumenta*, *disminuye*) en ese tipo de *amenaza a la validez* y la razón. Si la validez no se ve afectada, se especifica *no afecta a la validez*.

El cambio { ({ aumenta | disminuye } la validez

{ interna | externa | del constructo | de conclusión }) |

no afecta a la validez }

{ debido a | ya que } <razón>

Este patrón puede repetirse más de una vez ya que un cambio puede afectar a más de un tipo de *amenaza a la validez*.

- *Comentario*. Información adicional sobre las *amenazas a la validez* que pueda ser de interés para el lector. Es opcional y de texto libre. [<Comentario>]

- *Comentario.* Información adicional sobre el cambio que pueda ser de interés para el lector. Es opcional y de texto libre. [<Comentario>]

La Tabla §6.1 muestra la primera instanciación de la plantilla en la primera replicación de la familia de experimentos Mindfulness (Mind), descrita en la sección §7.4.1).

6.4 APLICABILIDAD DE LA PLANTILLA

En esta sección, se analiza en qué *tipo de replicación* o en general, en qué situación es más recomendable el uso de la plantilla propuesta.

Para ello, se analiza la *aplicabilidad* de la plantilla desde el punto de vista, tanto del experimentador que ha realizado una replicación y quiere dejar constancia de sus cambios, como de un experimentador que quiere abordar una nueva replicación. Teniendo en cuenta por un lado, la clasificación de las replicaciones propuesta por Gómez *et al.* [97], ya comentada en la sección §2.4, en: *literal*, *conceptual* y *operacional* y por otro lado, la clasificación de las replicaciones según sean llevadas a cabo por los mismos experimentadores (*replicación interna*) o distintos (*replicación externa*).

- Cuando la replicación es de tipo *literal*, la aplicabilidad de la plantilla es alta ya que aunque en principio no hay cambios al tratarse de una replicación lo más exacta posible al experimento base, es necesario especificar los cambios como resultado de la adaptación del experimento al nuevo entorno. Se trata de una *repetición*.

Mediante la especificación de los cambios, en caso de que los hubiera, es posible establecer la similitud entre la replicación y el experimento original.

- Cuando la replicación es *conceptual* solo se mantiene la hipótesis. Implica cambios en los experimentadores, nuevos protocolos, contexto, población y/o operacionalización. En este caso, la especificación de cambios mediante la plantilla es de menor utilidad ya que no hay una base sobre la que definir los cambios. Se trata de una *reproducción*.
- Las replicaciones de tipo *operacional* son las que presentan más cambios en cualquiera de las dimensiones identificadas y por tanto la *aplicabilidad* de la plantilla es mayor. De esta forma, es posible:

Tabla 6.1: Instanciación de la plantilla en la replicación Mind#2

Replicación	<i>Mind #2</i> replicación Interna basada en el experimento base <i>Mind #1</i>
Objetivo del experimento	Estudiar si la práctica de mindfulness (causa) mejora la productividad en modelado conceptual (efecto) en estudiantes de Ingeniería del Software (población).
Descripción del experimento	Un grupo de estudiantes del Grado de Ingeniería de Software de la Universidad de Sevilla (muestra) asistió a sesiones de mindfulness de 10 minutos durante 4 semanas, 4 días por semana (tratamiento grupo experimental), mientras que un segundo grupo de estudiantes asistió a un taller de oratoria a modo de placebo (tratamiento grupo de control). Los estudiantes realizaron dos ejercicios de modelado conceptual, uno antes y otro después del tratamiento (diseño experimental), que se puntuaron usando una solución de referencia (procedimiento de medida). Se comparó el rendimiento de ambos grupos en términos de calidad (similitud con la solución de referencia) y productividad (similitud en porcentaje por unidad de tiempo) (métricas).
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en el primer semestre del curso 2013–2014 y esta replicación en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> , en el primer semestre del curso 2014–2015
Propósito	Confirmar resultados Superar limitaciones del experimento base
Cambio #1	Aumento de la duración del tratamiento Originalmente , Durante 4 semanas se practicaba Mindfulness 4 días a la semana en sesiones de 10 minutos En la replicación las sesiones duraron 12 minutos y durante 6 semanas Con el fin de hacer más evidentes los beneficios de Mindfulness.
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente, la causa El cambio incrementa la validez del constructo ya que incrementar la duración del tratamiento refleja mejor el efecto de la práctica de Mindfulness. El cambio incrementa la validez interna ya que incrementar la duración del tratamiento refuerza su efecto sobre el de otros posibles factores.
Cambio #2	Asignación aleatoria de sujetos a grupos Originalmente , la asignación de los sujetos al tratamiento no fue aleatoria En la replicación es aleatorio Con el fin de mitigar la amenaza de sesgo de selección y asignación y evitar las limitaciones del análisis estadístico
Dimensión Modificada Amenazas a la validez	Protocolo , específicamente, el diseño experimental. El cambio incrementa la validez de conclusión, ya que mejora la potencia de los tests estadísticos aplicables. El cambio incrementa la validez interna, ya que elimina el sesgo de asignación.
Cambio #3	Tratamiento nulo del grupo de control Originalmente , se impartió un taller de oratoria al grupo de control como placebo En la replicación El grupo de control recibió un tratamiento nulo, ya que el taller de oratoria se realizó una vez terminado el experimento Con el fin de mitigar el potencial factor perturbador del placebo en los resultados del experimento.
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente, la causa El cambio incrementa la validez interna ya que elimina la posibilidad de que el placebo pudiera tener algún efecto sobre los resultados.

- Encontrar los límites de *operacionalización*. Mediante los cambios se describen, entre otros, las nuevas métricas y tratamientos aplicados y se analizan en consecuencia.
 - Analizar las variables de contexto que puedan afectar a los resultados.
 - Analizar si los resultados se mantienen para otro tipo de *población*. La especificación de los cambios, permite caracterizar a la nueva población marcando las diferencias con la utilizada en el experimento base.
 - Discernir si los resultados son consecuencia de una determinada *configuración experimental*. La especificación de los cambios realizados en el *protocolo experimental* permiten comprobar si los resultados se mantienen a pesar de los cambios.
 - Analizar si los resultados pueden estar sesgados por los *experimentadores*. Al especificar los cambios de investigadores (o sus roles), se puede comprobar si los resultados dependen de los investigadores.
 - Replicaciones para superar algunas limitaciones encontrada en el experimento base que puede representar una *amenaza a la validez*. La especificación de los cambios introducidos, permite dejar constancia del motivo de dichos cambios y cómo afectan a *validez del experimento*.
- Cuando la replicación es *externa*, se presenta el problema de que el experimentador que realiza la replicación y cumplimenta la plantilla, conoce poco el experimento base ya que ha sido llevado a cabo por otros experimentadores; la especificación de los cambios es complicada. En el análisis realizado en la sección §5 solo 6 de los 10 estudios que reportan replicaciones *externas*, sin el experimento base, presentan cambios.

Reportar una replicación *externa* junto con el experimento base no es una práctica frecuente ya que los autores de ambas son distintos. Hay pocas publicaciones de este tipo. En el análisis realizado en la sección §5 solo se encontraron 2 de 25 estudios.

- Pensamos que al reportar una replicación *interna* es donde mejor encaja la plantilla ya que los cambios junto con su motivación son bien conocidos por los autores y es importante que queden documentados. En el análisis realizado en la sección §5 todas las replicaciones *internas* presentaban cambios.

También se puede analizar la *aplicabilidad* de la plantilla desde el punto de vista de un experimentador que va a replicar un experimento de otro experimentador, debe

conocer bien el experimento original, incluso preguntando al experimentador original si algo no queda claro. Al conocerlo bien, debe poder reportar los cambios que introduce usando la plantilla. Además para un posible meta-análisis –agregación de experimentos– es necesario saber qué cambios realiza para poder analizar su impacto y definir correctamente las variables *moderadoras*.

6.5 UNA HERRAMIENTA PARA DEFINIR LOS CAMBIOS

En este apartado se describe la herramienta CAESAR, desarrollada en el grupo ISA y basada en la plantilla propuesta que permite gestionar los cambios de las replications. Ha sido desarrollada en Java con la base de datos H2 incrustada y desplegada en Heroku.

El actual prototipo consta de tres vistas principales que permiten: *i)* Describir el experimento base y sus replications, *ii)* Especificar los cambios de cada replicación y *iii)* Identificar las amenazas a la validez implicadas en cada cambio.

La herramienta está disponible en <https://metamodelo.herokuapp.com/>. Estamos trabajando para que evolucione como asistente para que los experimentadores puedan definir sus replications y cambios, generando de forma automática el código Latex correspondiente a la plantilla.

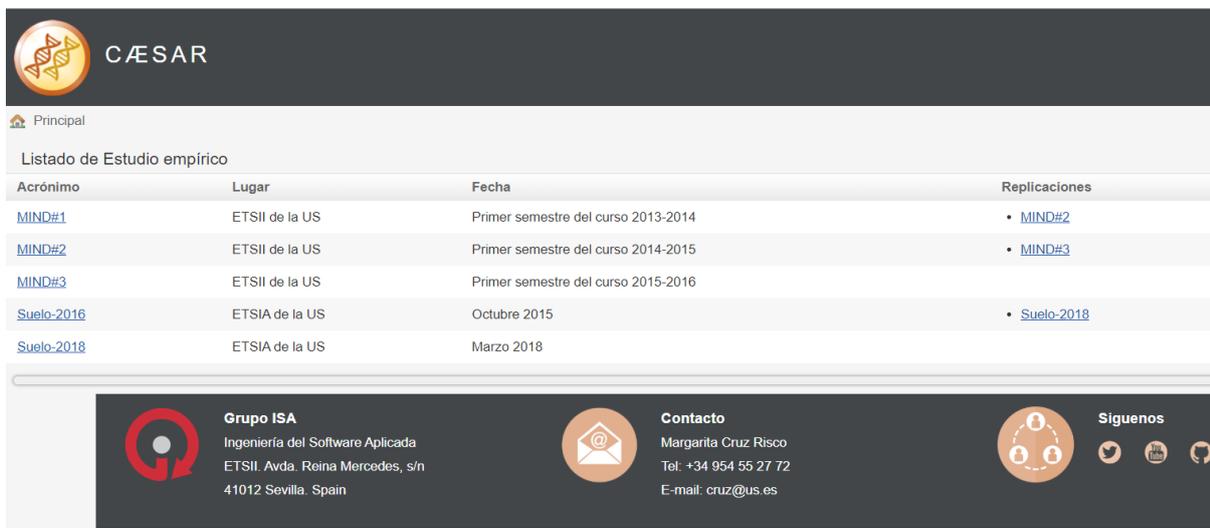
A continuación se describen, de forma somera, las principales características de sus tres vistas.

6.5.1 Vista de estudios empíricos

Desde esta vista se visualizan tanto los experimentos base como las replications. Se ha considerado que para ser una replicación, la hipótesis del experimento base debe mantenerse en la replicación.

La figura §6.4 muestra la lista de estudios empíricos. Es posible filtrar para visualizar solo las replications o bien los estudios base. Permite editar, borrar y crear nuevos estudios empíricos.

Al seleccionar un estudio empírico muestra el detalle de sus cambios.



Acronimo	Lugar	Fecha	Replicaciones
MIND#1	ETSII de la US	Primer semestre del curso 2013-2014	• MIND#2
MIND#2	ETSII de la US	Primer semestre del curso 2014-2015	• MIND#3
MIND#3	ETSII de la US	Primer semestre del curso 2015-2016	
Suelo-2016	ETSIA de la US	Octubre 2015	• Suelo-2018
Suelo-2018	ETSIA de la US	Marzo 2018	

Grupo ISA
Ingeniería del Software Aplicada
ETSII. Avda. Reina Mercedes, s/n
41012 Sevilla. Spain

Contacto
Margarita Cruz Risco
Tel: +34 954 55 27 72
E-mail: cruz@us.es

Síguenos

Figura 6.4: Vista de estudios empíricos

6.5.2 Vista de cambios en las replicaciones

En la descripción de los cambios, para los tipos de cambios, se muestra una lista desplegada con las cinco posibles dimensiones. Dependiendo de la dimensión seleccionada, el siguiente campo, correspondiente al elemento afectado, muestra una lista con los posibles valores o bien un cuadro de texto libre para especificar el elemento.

De esta forma, la clasificación según tipo de cambio es flexible, permitiendo actualizar el tipo de cambio, sin tener que borrar y volver a crear el cambio.

La figura §6.6 muestra el detalle de un cambio. Desde esta pantalla se puede editar, borrar o añadir un nuevo cambio o bien editar las *amenazas a la validez* afectadas.

6.5.3 Vista de amenazas a la validez

La figura §6.6 muestra la lista de *amenazas a la validez*. Su funcionamiento es similar a las vistas anteriores, permitiendo la navegación entre los diferentes elementos de la plantilla.

6.6 RESUMEN

En este capítulo, se ha presentado la principal aportación del trabajo recogido en esta memoria para contribuir a la mejora en el reporte de las replicaciones y específi-

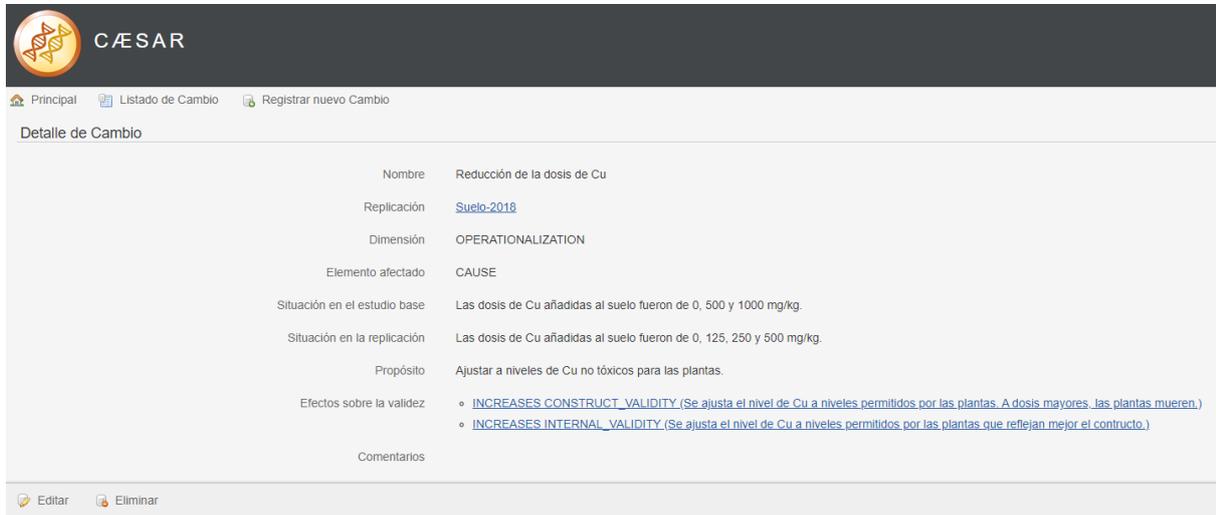


Figura 6.5: Vista de uno de los cambios junto con sus dos amenazas a la validez

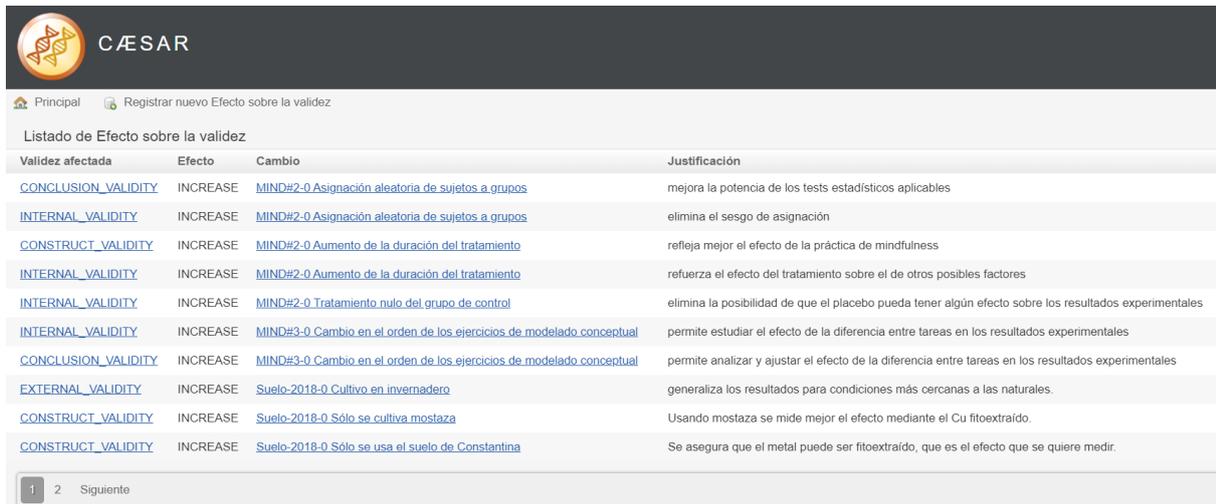


Figura 6.6: Vista de amenazas a la validez

camente de sus cambios. El enfoque propuesto consta de una plantilla junto con el metamodelo subyacente y se completa con *patrones lingüísticos*.

La definición del metamodelo ha permitido identificar la información relacionada con las replicaciones y específicamente con sus cambios. El uso de plantillas además de estructurar dicha información, evita redundancias, ambigüedades y la falta de información relevante. Los *patrones lingüísticos* facilitan su cumplimentación.

Para la definición de cada cambio, se describe la *situación en el experimento original*, en la *replicación* y el *motivo del cambio*. Así mismo, dependiendo del *elemento de la configuración experimental* modificado, se identifica la *dimensión afectada* y se analiza el impacto de cada cambio en la *validez del experimento*. Aunque somos conscientes de la dificultad de identificar tanto la *dimensión* afectada por el cambio como las *amenazas a la validez*, ambos campos contribuyen a mejorar la precisión en la definición de los cambios mediante la plantilla por lo que se han definido como de cumplimentación obligatoria.

Se ha analizado la *aplicabilidad* de la plantilla destacando su uso en replicaciones *internas* para dejar constancia del *conocimiento tácito* sobre la motivación de los cambios junto con las *amenazas a la validez* y *dimensiones* afectadas por dicho cambio. De igual forma, mediante la plantilla ya cumplimentada, un experimentador externo puede conocer la motivación e implicaciones de los cambios ya realizados antes de proponer nuevos cambios.

Basada en la plantilla propuesta, se presenta la herramienta CÆSAR que facilite a los experimentadores la definición de los cambios que pasan a formar parte de un repositorio de información, proporcionando una visión global de la familia de experimentos.

Esperamos que la solución propuesta para facilitar la especificación de los cambios de las replicaciones, redunde en beneficio de las investigaciones en IS.

PARTE IV

VALIDACIÓN

EVALUACIÓN DE LA PROPUESTA EN UN ESTUDIO DE CASO MÚLTIPLE

Todo lo que se puede probar debe ser probado

*Malcolm Gladwell (1963 –),
Escritor, periodista y sociólogo canadiense*

En este capítulo, respondiendo a la pregunta de investigación **PI₅** sobre extender la propuesta a otras áreas y tipos de experimentos, se evalúa la plantilla propuesta mediante un estudio de caso múltiple que incluye tres estudios de caso pertenecientes a las áreas de Ingeniería del Software, Agrobiología y Tecnología de alimentos y un tercer estudio de caso que incluye experimentos de tipo automáticos.

En la sección §7.1 se presenta el estudio de caso múltiple. En la sección §7.2 se presentan algunas singularidades en la experimentación en el área de Ciencias. En la sección §7.3 se detallan las fases del estudio de caso múltiple. En la sección §7.4 se evalúa la plantilla en el área de Ingeniería del software. En la sección §7.5 se evalúa en el área de Agrobiología y Tecnología de alimentos. En la sección §7.6 se instancia la plantilla en experimentos automáticos. En la sección §7.7 se responde a las preguntas de investigación del estudio de caso múltiple. En la sección §7.8 se muestran las conclusiones. Por último, la sección §7.9 resume el capítulo.

7.1 INTRODUCCIÓN

El *estudio de caso* es una metodología de investigación que se lleva a cabo en su contexto natural. Los casos pueden variar desde estudios bien organizados hasta pequeños ejemplos en un laboratorio universitario [196, 197].

El *estudio de caso múltiple* incluye más de un *estudio de caso* con características diferentes, de forma que si la información obtenida es similar, las conclusiones son más sólidas [197]. El presente *estudio de caso múltiple* incluye tres *estudios de caso* con el objetivo principal de evaluar el uso de la plantilla, especificando los cambios en las repeticiones de las familias de experimentos seleccionadas. Debido a los diferentes contextos de cada *estudio de caso*, se propone un *estudio de caso múltiple*.

El método de investigación seguido en este estudio se ha basado en el proceso de investigación de *estudios de caso* propuesto en Runeson y Höst [196] y Runeson et al. [197].

7.2 EXPERIMENTACIÓN EN CIENCIAS

Antes de aplicar la plantilla en el área de Ciencias, concretamente en el área de conocimiento de *Agrobiología y Tecnología de alimentos*, parece necesario comentar algunas diferencias con respecto al área de IS, analizar la terminología utilizada y, en general, exponer los principios básicos de la experimentación en Ciencias.

7.2.1 Principios básicos

Las principales conceptos aquí presentados han sido extraídas de algunos de los textos más utilizados en el área [79, 149, 158].

En este área, un experimento permite, por ejemplo, tomar decisiones sobre la mejor dosis de abonado, concentración de un pesticida, una variedad de planta, etc. Para ello se manipulan las *variables independientes* (causas) y se analizan las consecuencias sobre una o más *variables dependientes* (efectos). El *tratamiento* que se aplica y cuyo efecto se mide, puede ser una dosis de reactivo, una variedad de semilla, una determinada temperatura/humedad, etc. y su efecto se mide por ejemplo analizando las características fisiológicas de las plantas, concentración de nutrientes o metales en el suelo, etc.

Se denomina *observación* a cada valor de la *variables dependientes* recogida en el experimento [125].

La *unidad experimental* (o *parcela experimental*) es la mínima unidad de material a la que se aplica un *tratamiento*, puede ser una parcela, una maceta, una planta, etc. Se recomienda que *unidad experimental* sea lo más uniforme posible.

El *diseño experimental* es la asignación de los *tratamientos* a las *unidades experimentales*.

La *variable independiente*, también denominada *factor* puede tomar distintos valores denominados *niveles*. Cuando hay más de un *factor*, puede producirse una *interacción entre los factores* debido a que uno de los *factores* no produce el mismo efecto según los distintos niveles de otro *factor*. En este caso, se puede utilizar un *diseño factorial* que permite variar los *factores* de forma conjunta y estudiar su interacción.

En el caso de que haya dos *factores* y cada uno con dos *niveles*, se trata de un *diseño factorial 2²*. Así, por ejemplo, se puede analizar el efecto de dos *factores* en el crecimiento de una planta acumuladora de metal en suelos contaminados, tales como: *i*) las distintas concentraciones de cobre (*factor1*) y *ii*) un *biosurfactante* para facilitar la extracción del metal (*factor2*).

El *error experimental* es la variación que resulta en las medidas realizadas en *unidades experimentales* que han recibido el mismo *tratamiento*.

Dicha variación puede deberse, básicamente, a diferencias inherente al material experimental o bien a que la realización del experimento no es uniforme. Es decir, el *error experimental* son los efectos desconocidos sobre la *variable dependiente*. Es necesario controlar dicho *error* para que las diferencias entre los resultados de las *unidades experimentales* sean sólo debidas al *tratamiento* recibido.

El *error experimental* se puede controlar, o al menos puede disminuir, mediante:

- *Diseño de bloques*. Las *unidades experimentales* se colocan en *bloques* de manera que en cada *bloque* las condiciones experimentales sean homogéneas (se bloquea el factor). Es similar a cuando en IS se hacen *bloques* dependiendo de la experiencia de los *sujetos experimentales*. Aquellos con igual nivel de experiencia se colocan en el mismo *bloque*.
- *Análisis de covarianza*. Cuando un *factor* no se puede bloquear (por ejemplo, la humedad relativa) pero se puede medir, es posible realizar un análisis de la *covarianza* para eliminar su efecto.

- *Repeticiones.* Una de las razones para llevar a cabo *repeticiones* es que el suelo no es homogéneo. Cuando un mismo *tratamiento* aparece más de una vez en un experimento se dice que está repetido. Al aumentar el número de *repeticiones* disminuye el error estándar (en la fórmula del error estándar se divide por la raíz cuadrada del tamaño de la muestra).

Para calcular el número de *repeticiones* a llevar a cabo, hay que tener en cuenta el tamaño de las *unidades experimentales*. Las *unidades experimentales* grandes presentan menos variaciones que las pequeñas pero aumentan el coste de las *repeticiones* ya que implica mayor cantidad de material experimental, aumento de la población, aplicación menos homogénea de las técnicas y en definitiva aumento del coste en tiempo y recursos necesarios.

El número de *repeticiones* depende además, de la precisión deseada. Cuanto más pequeña sea la discrepancia con respecto a la hipótesis nula que se ha de comprobar, mayor será el número de *repeticiones* requeridas.

- *Aleatorización.* La asignación de los *tratamientos* a las *unidades experimentales* debe realizarse al azar, de forma que cada *unidad experimental* tenga la misma probabilidad de recibir un determinado *tratamiento*. Se ha demostrado que parcelas adyacentes tienden a tener rendimientos más parecidos que las distantes debido al llamado *error correlacionado* o *residuo*.
- *Control del efecto borde.* En los experimentos, se suelen evaluar solamente las filas interiores y evitar así el *efecto borde*. Se denominan *filas guarda* a las filas correspondientes a *unidades experimentales* situadas en el exterior (rodean a las filas interiores) y cuyas plantas se pueden ver afectadas por condiciones de iluminación, humedad, nutrientes, etc. distintas a las condiciones de las plantas en filas interiores.

Para aclarar los conceptos anteriores, en la figura §7.1 se muestra un ejemplo de diseño con dos *factores*: factor1 con 3 niveles y factor2 con 2 niveles, dando lugar a 6 *tratamientos* (3x2). Para cada *tratamiento* se preparan 4 macetas y se colocan en una bandeja. La *unidad experimental* es la bandeja. Se decide realizar 3 *repeticiones* (3 *bloques*). Cada *bloque* tiene los 6 *tratamientos*. Las bandejas se distribuyen de forma completamente aleatoria en cada *bloque*. En total hay 18 bandejas (*unidades experimentales*) con 4 macetas cada una.

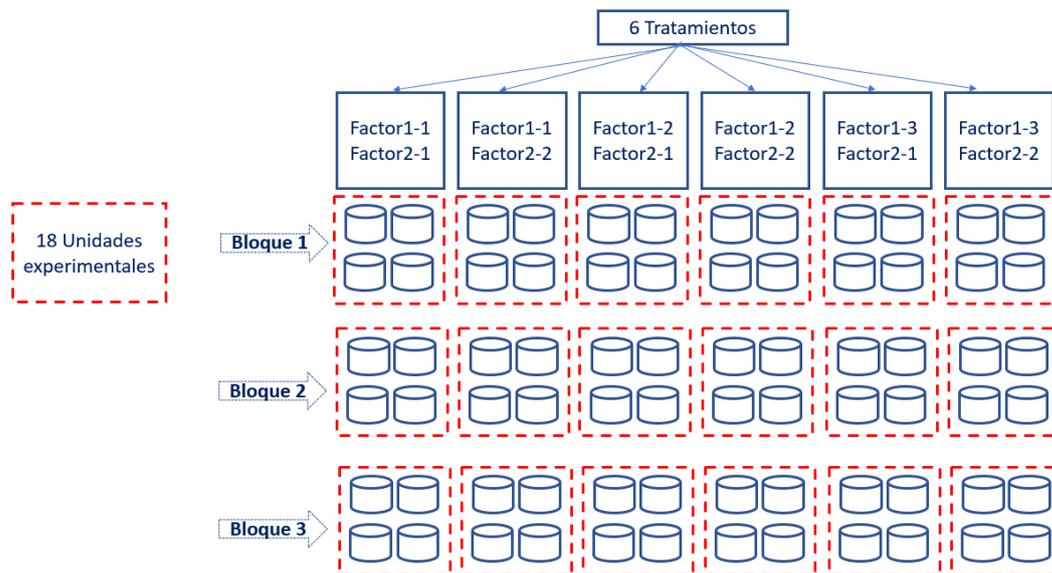


Figura 7.1: Diseño con 6 tratamientos en 3 bloques

7.2.2 Diferencias con la experimentación en IS

Las dos principales diferencias encontradas y que han sido comprobadas revisando algunos de los artículos más citados y hablando con investigadores del área son referentes a: i) las *amenazas a la validez* y ii) el concepto de *replicación*.

Mientras que en IS es una práctica frecuente analizar las *amenazas a la validez*, en Ciencias se habla de *error experimental*. Tal como se muestra a continuación, se puede establecer un paralelismo entre ambos conceptos.

En el área de Ciencias, el error experimental es consecuencia de efectos desconocidos sobre la *variable dependiente*. Es equivalente al concepto de *validez interna* en IS.

Respecto a la *validez externa*, en el área de Ciencias los experimentos se llevan a cabo en primer lugar, en *placas de Petri* o en *cámaras de cultivo* es decir condiciones muy controladas y posteriormente se realizan en *invernaderos*, *jardines* u otros. Está relacionado con el concepto de *validez externa* en IS.

Las *amenazas a la validez del constructo* están relacionadas con la determinación de las *variables independientes y dependientes*. En el área de Ciencias es más probable que las *variables independientes y dependientes* reflejen de manera precisa el constructo que se va a estudiar. Así por ejemplo, para comparar dos tipos de fertilizantes, se utilizan y se miden sus efectos en las plantas. Previamente a la realización del experimento se realizan *pruebas o ensayos preliminares* encaminadas a determinar las condiciones óptimas

para aplicar el *tratamiento* y medir sus efectos. Así por ejemplo, se determinan entre otras, las concentraciones de reactivos convenientes para los objetivos del experimento o los tipos de suelo o de plantas que responderán mejor al *tratamiento*. Las *pruebas preliminares* contribuye a aumentar la *validez del constructo* aunque no se utilice esta terminología de *validez* del experimento. En IS, para comparar dos técnicas es necesario probar su utilización por persona que pueden tener distintos nivel de experiencia y hay que determinar el procedimiento para medir la variable dependiente (resultados) y todo unido a la problemática de los experimentos con personas. Por lo tanto, en IS es más probable que se presente amenazas a la *validez del constructo*.

Referente a la *validez de conclusión*, en el área de Ciencias es más fácil aumentar la muestra a la que se aplica el *tratamiento* que en IS donde los sujetos son humanos. En los experimentos de Ciencias es usual que un mismo *tratamiento* se *repita* más de una vez o bien repetir el experimento completo y agregar los resultados.

La segunda diferencia encontrada es relativa a la terminología. En Ciencias, se utiliza el término *repetición* y no se utiliza *replicación*. En cada experimento se repite el *tratamiento* varias veces (*bloques*) y además se recomienda repetir el experimento en diferentes localizaciones, tipos de suelos y temporadas. Por tanto, el término *replicación* en IS es el equivalente a las *repeticiones* o simplemente *experimento* realizado en una localización, contexto, o en general en una fecha distinta a la del experimento original. El nuevo experimento es consecuencia de los resultados obtenidos en el experimento original. Así por ejemplo se ajusta la dosis de reactivo a valores no tóxico, se analiza la interacción entre dos *factores* etc. y habrá que tener en cuenta los *efectos acumulativos* en el suelo tratado.

Por último, según indican los expertos, tampoco se utiliza el término *familia de experimentos* de uso muy habitual en IS.

7.3 ESTUDIO DE CASO MÚLTIPLE

En el presente *estudio de caso múltiple*, como se muestra en la figura §7.2, se identifican tres fases principales: *i)* Diseño del estudio de caso; *ii)* Planificación, recopilación y análisis; y *iii)* Análisis e informe conjunto.

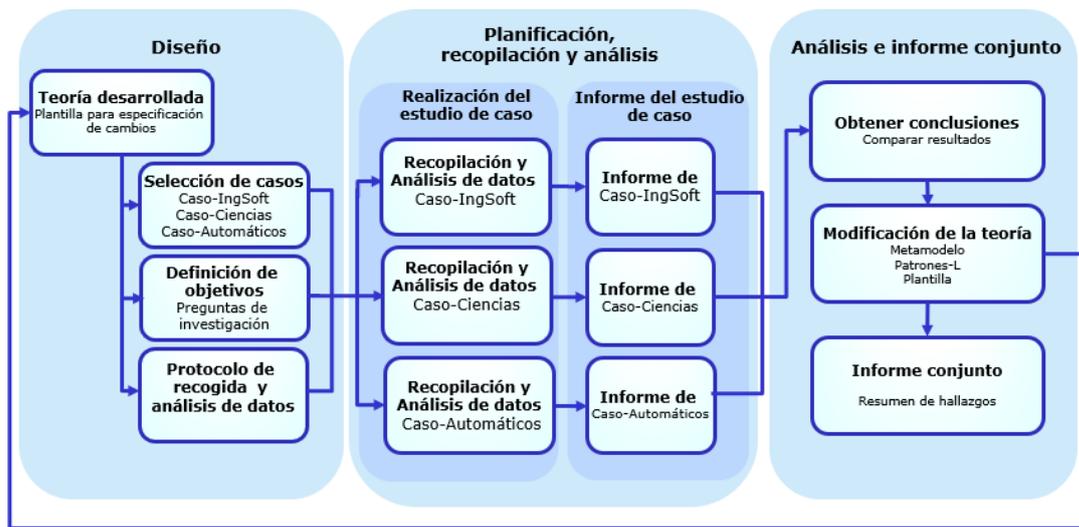


Figura 7.2: Fases del estudio de caso múltiple basado en [197]

7.3.1 Diseño del estudio de caso

En esta primera fase se clarifica el contexto de investigación del estudio.

En secciones anteriores, se ha desarrollado la teoría presentando una plantilla como propuesta para especificar los cambios de las replicaciones de experimentos. Debido a la necesidad de validar dicha plantilla, mediante tres *estudios de caso*, se definen los objetivos que permiten proponer las preguntas de investigación y establecer el protocolo para la recopilación y el análisis de datos.

Los pasos de esta primera fase se detallan a continuación:

1. *Teoría desarrollada*. Se ha presentado un metamodelo que formaliza la información involucrada en las replicaciones (ver sección §6.2). El metamodelo se representa visualmente mediante una plantilla con formato de tabla y se completa con patrones-L (ver sección §6.3).
2. *Selección de casos*. Para instanciar la plantilla, se seleccionan tres *estudios de caso* que cubren varias áreas de conocimiento y tipos de experimentos:
 - *Estudio de caso en Ingeniería del Software (Caso-IngSoft)*. Inicialmente, se seleccionan tres familias de experimentos, pertenecientes al área de IS, que tratan sobre: *i*) Mindfulness (*Mind*), *ii*) Análisis de requisitos (*Req*), y *iii*) Técnicas de evaluación del código (*Code*).

- *Estudio de caso en Agrobiología y Tecnología de alimentos (Caso-Ciencias)*. Para extender el uso de la plantilla a otras áreas del conocimiento, se seleccionan cinco familias de experimentos que versan sobre: *i*) Descontaminación de suelos (*Suelo*), *ii*) Análisis de calidad del aceite de oliva virgen (*Calidad*), *iii*) Biodegradación en suelos agrícolas (*Bio*), *iv*) Extracción de componentes del aceite de oliva (*Oliva*), e *v*) Influencia de la dieta en la acumulación de colesterol (*Dieta*).
 - *Estudio de caso en experimentos automáticos (Caso-Automáticos)*. Para evaluar la plantilla en otros tipos de experimentos, como los experimentos algorítmicos o los de simulación, se seleccionan dos familias de experimentos sobre: *i*) Pruebas automatizadas de software (*Test*), y *ii*) Pruebas en Líneas de Productos software (*SPL*).
3. *Definición de objetivos*. El objetivo principal es evaluar empíricamente la plantilla especificando los cambios en las replicaciones que constituyen este *estudio de caso múltiple*. Previamente, se ha realizado una revisión de la literatura para analizar cómo han abordado otros investigadores el informe de replicación y específicamente sus cambios (ver sección §5.2).

Conocer otras propuestas seguidas, nos lleva a establecer comparaciones con dichas propuestas, analizar los aspectos de: *expresividad, precisión, usabilidad y trazabilidad* de la plantilla propuesta y plantear el conjunto de preguntas de investigación (*PIs*) presentadas a continuación:

- **PI_{4.1}**: *Expresividad*. ¿Qué proporción de cambios ha sido posible definir? ¿La *expresividad* de la plantilla es mayor que la de otras propuestas seguidas?.
 - **PI_{4.2}**: *Precisión*. El uso de los patrones-L, ¿hace más precisa la definición de los cambios que el lenguaje natural y/o en forma de tablas? ¿Permite detectar la información que falta?.
 - **PI_{4.3}**: *Usabilidad*. ¿Cuántos campos de la plantilla han presentado problemas de comprensión? ¿Están los investigadores familiarizados con la terminología?.
 - **PI_{4.4}**: *Trazabilidad*. Al registrar los cambios, ¿se facilita la *trazabilidad* entre replicaciones?.
4. *Protocolo de recogida y análisis de datos*. Se trata de identificar las fuentes para recopilar los datos y definir los procedimientos para llevar a cabo el análisis de los resultados.

En el presente *estudio de caso múltiple*, el protocolo de recolección de datos consiste en: *i)* contactar con los investigadores; *ii)* rellenar la plantilla; *iii)* registrar los detalles de cómo, cuándo y quién llevó a cabo este proceso; y *iv)* mantener entrevistas semiestructuradas (el orden de las preguntas no está establecido con anterioridad) con los investigadores antes y después de completar la plantilla.

7.3.2 Planificación, recopilación y análisis

En esta segunda fase se aplica el protocolo anterior a los tres estudios de caso y se escribe su informe.

Comprende, para cada caso, los procesos de: *i)* *recopilación de datos y análisis de los datos recogidos* y *ii)* *elaboración del informe del estudio de caso*. Los pasos se detallan a continuación:

1. *Proceso de recopilación de datos*. Según Lethbridge *et al.* [144] *las técnicas de recolección de datos* se pueden clasificar en tres grados: *i)* en el primer grado están las técnicas que requieren *acceso directo* a la población participante, *ii)* el segundo grado implica *acceso al entorno de trabajo* de los participantes, y *iii)* el tercer grado requiere *acceso sólo a los artefactos de trabajo*. A continuación, para cada estudio de caso, se especifican las técnicas de recolección de datos utilizadas y las actividades involucradas.
 - En *Caso-IngSoft*, la plantilla es cumplimentada por mí a partir de los estudios de replicaciones publicados.
 - En *Caso-Ciencias*, se mantienen varias reuniones con investigadores pertenecientes a la *E.T.S. Ingeniería Agronómica (ETSIA)*, al *Instituto de Recursos Naturales y Agrobiología (IRNAS-CSIC)*, y al *Instituto de la Grasa (IG-CSIC)*. Los objetivos de estas reuniones son: *i)* presentar y explicar el uso de la plantilla; *ii)* pedir a cada investigador que, aconsejado por mí, seleccione un experimento con al menos una replicación y, si es posible, que haya sido publicado; *iii)* recopilar la opinión de los investigadores una vez que han cumplimentado la plantilla; y *iv)* proponer un cuestionario a dichos investigadores para analizar si la terminología es la misma que la utilizada en IS.
 - En *Caso-Automáticos*, los investigadores pertenecen al área de conocimiento de IS y ya conocen la plantilla. El procedimiento es similar al caso anterior aunque, en este caso, las plantillas son cumplimentadas por mí y validadas

por los autores de las replicaciones. Del mismo modo, al final del proceso, se pasa un cuestionario para recoger su opinión.

Se utilizan *técnicas de recogida de datos* de los tres grados: *i)* se recogen opiniones sobre el uso y utilidad de la plantilla mediante cuestionarios y entrevistas semi-estructuradas en los centros de trabajo de los investigadores (primer y segundo grado); y *ii)* para seleccionar las familias de experimentos se utilizan datos de archivo (*archival data*¹) revisando las principales publicaciones de los investigadores implicados.

2. *Análisis de los datos recogidos.* Una vez que se han recopilado los datos, es necesario su análisis tanto *cualitativo* como *cuantitativo*.

Se analizan los datos recopilados fundamentalmente al cumplimentar las plantillas y los derivados de los cuestionarios que se han pasado a los investigadores.

Así, tras recopilar las plantillas rellenas, se analiza si los campos sin cumplimentar (en blanco) son debidos a la diferente terminología utilizada en otras áreas o por falta de información. Igualmente, se analiza si la plantilla recoge toda la información que el investigador necesita reflejar.

Hay otros detalles de interés sobre la plantilla como: *quién la cumplimenta* (por ejemplo, los autores de la replicación), *cómo* (por ejemplo, a partir del informe de la replicación), *cuándo* (por ejemplo, después de que se lleve a cabo la replicación) o si se ha validado la plantilla, en caso de nos ser cumplimentada por los autores de la replicación.

La instanciación completa de las replicaciones que constituyen el *estudio de caso múltiple* mediante la plantilla, se presentan en el apéndice §B.

3. *Informe del estudio de caso.* Es la redacción del informe de cada estudio de caso. Las secciones §7.4, §7.5 y §7.6 reportan en detalle cada uno de los tres estudios de caso.

7.3.3 Análisis e informe conjunto

Una vez analizado cada *estudio de caso* en su entorno, el siguiente paso es integrar los resultados obtenidos de manera que se puedan comparar y conseguir el objetivo del *estudio de caso múltiple* de obtener conclusiones más sólidas. Los pasos se detallan a continuación:

¹Tipo de datos que se puede recopilar mediante técnicas de tercer grado en un estudio de caso.

1. *Obtener conclusiones.* Se extraen conclusiones cruzadas y se comparan los resultados de la aplicación de la plantilla en los tres estudios de caso. Se presentan en la sección §7.8.
2. *Modificación de la teoría.* Las limitaciones encontradas en la aplicación de la plantilla posibilitan que evolucione desde su versión inicial incorporando las mejoras pertinentes. Los ajustes realizados al aplicar la plantilla en el *Caso-IngSoft* se detallan en la sección §7.4.
3. *Informe conjunto.* Como paso final, se escribe el informe que se presenta en la sección §7.7.

7.4 ESTUDIO DE CASO: CASO-INGSOFT

En este estudio de caso, se valida la plantilla por primera vez en tres familias de experimentos del área de IS, analizando su utilidad y realizando los ajustes pertinentes.

La familia de experimentos de *Mindfulness (Mind)* ha sido seleccionada por ser conocida ya que ha dado origen a la propuesta. Las otras dos familias han sido seleccionadas por haber sido llevadas a cabo por otros investigadores y porque su dominio del problema resulta familiar.

La instanciación con la versión inicial de la plantilla en las distintas replicaciones del presente estudio de caso, aparecen en el apéndice §A. En el apéndice §B se muestra la instanciación con la versión final de la plantilla.

Para cada replicación, se especificaron los aspectos básicos: (*acrónimo, tipo, y objetivo*) y cada uno de sus *cambios*.

En los siguientes apartados se detallan las limitaciones encontradas y los ajustes realizados en ambas categorías: *aspectos básicos* y *cambios*. Dichos problemas han permitido, por un lado, la evolución del metamodelo y de la plantilla desde su versión inicial; se han añadido nuevos campos y se han modificado algunos patrones-L y por otro, se ha demostrado que el uso de la plantilla evidencia la falta de información relevante.

A continuación, se presentan brevemente las tres familias pertenecientes al dominio de la Ingeniería del Software en las que se evalúa empíricamente la plantilla.

7.4.1 Descripción de las familias

Familia Mind.

La familia *Mind* [25, 28], consiste en un experimento y sus dos replicaciones internas realizadas en la *Universidad de Sevilla*. El estudio trata sobre el efecto de la práctica del *Mindfulness* en el rendimiento de los estudiantes a la hora de desarrollar modelos conceptuales. *Mindfulness* es una técnica de meditación destinada a mejorar la claridad de la mente, la atención y la presencia.

El acrónimo usado para hacer referencia al original y sus dos replicaciones son *Mind#1*, *Mind#2* y *Mind#3*. La segunda replicación *Mind#3* se ha definido sobre la base de la primera replicación *Mind#2*.

Esta familia ha inspirado, en parte, la idea de facilitar la definición de los cambios mediante una plantilla y patrones-L. En la presente instanciación de la plantilla, con su estructura inicial, no se ha presentado la necesidad de hacer ningún ajuste a dicha plantilla aunque si se ha hecho patente la exigencia de su validación en experimentos realizados por experimentadores externos.

Familia Req.

La familia de experimentos *Req* estudia la influencia del conocimiento y la experiencia de los analistas en el análisis de requisitos. La familia consta de 9 experimentos en los que participan estudiantes de postgrado de la Universidad Politécnica de Madrid, así como profesionales de diferentes países e instituciones. El estudio se describe en [15].

Según el autor, los acrónimos de los experimentos y las replicaciones son: *Q-2007*, *Q-2009*, *Q-2011*, *Q-2012*, *E-2012A*, *E-2012B*, *E-2013*, *E-2014* y *E-2015*. *Q-2007* es el experimento original. Cada experimento es una replicación del experimento anterior. Es decir, se ha aplicado la plantilla definiendo los cambios en cada replicación con respecto al experimento anterior excepto el experimento *E-2015* que es una replicación del *E-2013*.

La letra Q indica que es un *cuasi-experimentos* y la letra E que es un experimento. El número corresponde al año de ejecución.

Familia Code.

Mediante la familia de experimentos *Code* se evalúa la eficacia de tres técnicas de verificación y validación del código. El experimento original se realiza con estudiantes de la Universidad Politécnica de Madrid. La primera replicación es

también en la Universidad Politécnica de Madrid y las siguientes en la Universidad Politécnica de Valencia, la Universidad de Sevilla y la Universidad ORT de Uruguay. Es decir, las replications son en cuatro sedes diferentes. El estudio se describe en [126, 129, 130].

Los acrónimos utilizados para el experimento original y sus cuatro replications son *VV-UPM*, *VV-UPM1*, *VV-UPV*, *VV-Uds* y *VV-ORT*.

7.4.2 Especificación de los aspectos generales de las replications

En esta sección se detallan los ajustes introducidas en la plantilla como consecuencia de los problemas encontrados al especificar los aspectos básicos de la replicación con la versión inicial de la plantilla.

Los ajustes han sido causados por información que no estaba inicialmente incluida en el metamodelo y luego se reparó en su necesidad. Es decir, la plantilla estaba incompleta en algunos puntos.

A continuación se describen los problemas encontrados junto con el ajuste necesario en el metamodelo y/o plantilla para solucionarlo.

Necesidad de añadir lugar y fecha

- *Descripción del problema:* El lugar y la fecha del experimento base y la replicación no se han registrado en la versión inicial de la plantilla. En [123] se identifica una nueva dimensión para reflejar que la replicación tiene lugar en un lugar diferente al del experimento original. Sin embargo, consideramos *lugar* como un campo de plantilla y no como una nueva *dimensión* ya que el *lugar* es una propiedad de replicación y no de cada *cambio*.

También se ha considerado que es importante conocer la *fecha* ya que hay condiciones que pueden variar en función del momento en que se realiza tanto el experimento como su replicación.

- *Ajuste en el metamodelo:* Se agregaron los campos *lugar* y *fecha* del experimento base (o replicación previa) y la replicación.
- *Ajuste en la plantilla:* Se agregó un patrón-L para recoger el *lugar* y *fecha* del experimento base (o replicación previa) y la replicación.

Necesidad de añadir una descripción del experimento base

- *Descripción del problema:* La versión inicial de la plantilla no contenía la descripción del experimento base. Es necesario registrar el GQM del experimento original o en su defecto, una descripción del objetivo, además de una breve descripción del experimento para contextualizar la replicación.
- *Ajuste en el metamodelo:* Se añaden los campos *objetivo* y *descripción* del experimento base.
- *Ajuste en la plantilla:* Se agrega un patrón-L para recoger el *objetivo del experimento* y otro para la *descripción del experimento*.

Dificultades para especificar el acrónimo

- *Descripción del problema:* En la versión inicial de la plantilla, la replicación se identifica por un código o acrónimo relacionado con el experimento de línea de base y seguido de un número secuencial. Sin embargo, hay replicaciones con una notación diferente asignada por sus autores.
- *Ajuste en el metamodelo:* No es necesario.
- *Ajuste en la plantilla:* El patrón-L se ha simplificado identificando la replicación y el experimento base sólo con un *acrónimo*.

La replicación se define sobre una replicación anterior

- *Descripción del problema:* En la versión inicial de la plantilla, la replicación se define con respecto a un experimento original, pero también hay replicaciones definidas con respecto a una replicación previa.
- *Ajuste en el metamodelo:* No es necesario.
- *Ajuste en la plantilla:* Se ha mejorado el patrón-L permitiendo definir la replicación con respecto a una replicación anterior.

7.4.3 Especificación de los cambios de las replicaciones

En esta sección se detallan las modificaciones introducidas en la plantilla como consecuencia de los problemas encontrados al especificar los cambios con la versión inicial de la plantilla.

El cambio no aumenta la validez experimental (Amenaza)

- *Descripción del problema:* En la plantilla original, la validez sólo podía aumentar. Sin embargo, hay cambios que disminuyen la validez o no afectan a la validez.

Así por ejemplo, al añadir o borrar una variable *dependiente* o *independiente* (*dimensión operacionalización*) puede *aumentar* o *disminuir la validez del constructo* [245].

La tabla §7.1 muestra un ejemplo de un cambio en el que se elimina una variable *dependiente* y por lo tanto la validez del *constructo disminuye*. La plantilla original no permitía especificar esta disminución. Además, hay cambios que pueden afectar a más de un tipo de amenaza.

Tabla 7.1: Especificación de un cambio que disminuye la validez (familia *Req*, replicación Q-2009)

Replicación	Q-2009 replicación Interna basada en el experimento original Q-2007
Cambio #1	<p>Análisis de la efectividad</p> <p>Originalmente, se analiza la efectividad de los analistas en las sesiones de educación (entrevistas) y la efectividad de los analistas en las sesiones de consolidación (pasado un tiempo)</p> <p>En la replicación no se analiza la efectividad en educación.</p> <p>Causado por el alto costo de transcribir las entrevistas.</p>
Dimensión modificada	Operacionalización , específicamente, el efecto
Amenaza a la validez	El cambio disminuye la validez del constructo ya que no se considera la variable dependiente efectividad de educación

- *Ajuste en el metamodelo:* Se añade el campo *tipo* para recoger si el cambio aumenta o disminuye la validez y el campo *razón* para explicar dicha variación.

La figura §7.3 compara la versión original y final del metamodelo en lo referente a *amenazas a la validez*.

- *Ajuste en la plantilla:* El patrón-L ha sido mejorado al permitir expresar que *la validez disminuye* y *no afecta a la validez*. Además, se ha añadido al patrón-L, la *razón* del aumento o la disminución de la validez.

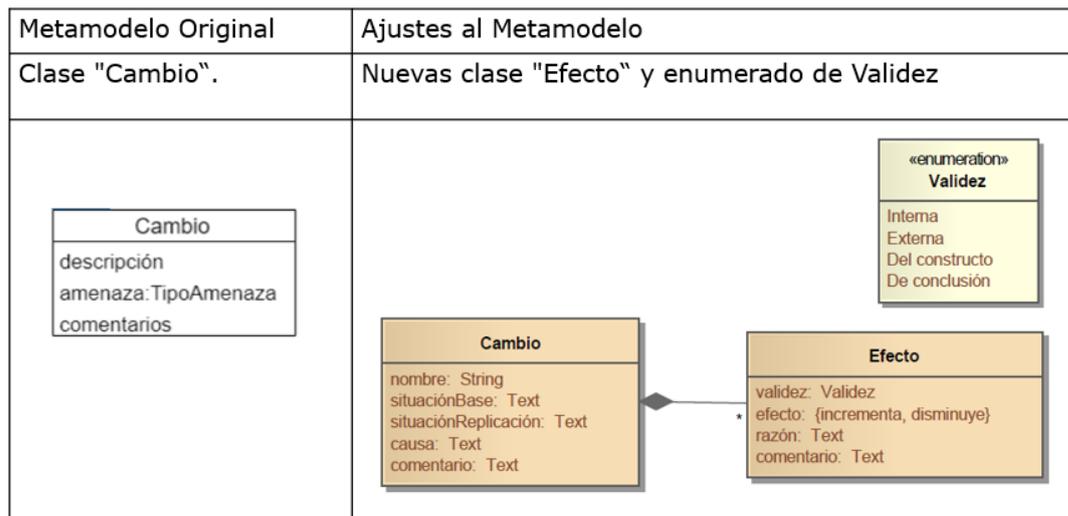


Figura 7.3: Versión inicial y final de la vista de amenaza a la validez

El cambio afecta a una variable de contexto (Contexto)

- *Descripción del problema:* No es posible especificar un cambio que afecte a una *variable de contexto*. En la versión original sólo se consideraron variables *dependientes* y *independientes*. Las *variables de contexto*, también llamadas parámetros, son variables *independientes* que fueron controladas a un nivel fijo durante el experimento [245].
- *Ajuste en el metamodelo:* Se identifica la nueva dimensión *contexto* para incluir las *variables de contexto*.
- *Ajuste en la plantilla:* El patrón-L se ajustó para incluir la nueva dimensión *contexto* con el elemento *variable de contexto*.

La tabla §7.2 muestra la instanciación de la plantilla, tras los ajustes realizados, en una replicación especificando los aspectos básicos y sus cambios. El cambio #1 no afecta a la validez experimental, el cambio #2 afecta a dos tipos de amenazas y cambio #3 afecta a una *variable de contexto*.

No se especifica el motivo del cambio (Razón)

- *Descripción del problema:* El motivo para realizar el cambio se desconoce o no estaba claro en el informe de la replicación. En la versión original de la plantilla la especificación del motivo del cambio es obligatoria.

Tabla 7.2: Especificación de los aspectos básicos y de los cambios (familia *Req*, replicación E-2012B)

Replicación	<i>E-2012A</i> replicación Interna basada en experimento original <i>E-2012B</i>
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Descripción del experimento	..
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2012 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2012
Propósito	Confirmar resultados
Cambio #1	Nuevos problemas Originalmente , se utilizan dos dominios de problemas en el experimento, uno de dominio conocido (DC) y el otro de dominio desconocido (DD). En la replicación se han modificado los dominios de problemas aunque uno sigue siendo de dominio conocido (DC) y el otro de dominio desconocido (DD).
Dimensión modificada	Protocolo , específicamente, material experimental
Amenaza a la validez	El cambio no afecta a la validez
Cambio #2	Orden de los problemas Originalmente , primero se realiza el problema de dominio conocido y luego el de dominio desconocido. En la replicación se permuta el orden de realización de los problemas, primero el problema de dominio desconocido y luego el conocido.
Dimensión modificada	Protocolo , específicamente, las guías
Amenaza a la validez	El cambio aumenta la validez interna ya que permite analizar si el orden afecta a los resultados El cambio aumenta la validez de conclusión ya que permite analizar y ajustar el efecto de la diferencia entre tareas en los resultados experimentales
Cambio #3	Momento de realización del experimento Originalmente , el experimento se ha llevado a cabo al principio del curso. En la replicación , se lleva a cabo al final del curso.
Dimensión modificada	Contexto , específicamente el Momento de realización
Amenaza a la validez	El cambio disminuye la validez interna ya que la replicación se lleva a cabo a final de curso y puede influir en los resultados.

- *Ajuste en el metamodelo*: Ninguno.
- *Ajuste en la plantilla*: Ninguno.

Aunque se pensó en hacer la descripción del motivo del cambio opcional, al analizar lo que se dice tanto [42] como en [97] sobre describir el motivo del cambio, se decidió mantenerlo como de cumplimentación obligatorio. De esta forma, se facilita la comprensión del cambio al estar claramente especificado.

En la definición de los cambios que se muestran en la tabla §7.2 *el motivo no se especifica*.

No se especifica el nombre de la variable (Nombre)

- *Descripción del problema:* Hay cambios que afectan a una variable *dependiente o independiente* y el *nombre de la variable* no estaba especificado en el informe de replicación. En la versión original de la plantilla, era necesario especificar el *nombre de la variable*.
- *Ajuste en el metamodelo:* Ninguno.
- *Ajuste en la plantilla:* Ninguno.

De la misma forma que en el problema anterior, consideramos esencial su especificación al facilitar la tarea para el experimentador que necesita conocer los cambios ya realizados.

7.4.4 Discusión

En esta sección se cuantifica el número de cambios que presentan alguno de los problemas identificados anteriormente.

La tabla §7.3 presenta para cada replicación de cada familia de experimentos, el número de cambios de dicha replicación y los cambios que presentan algún problema junto con el tipo de problema presentado ya sea *Amenaza*, *Contexto*, *Razón* y/o *Nombre*. Los cambios se identifican por Cambio-n, siendo n un número secuencial dentro de cada replicación empezando por el 1.

Estos problemas, son los encontrados en la aplicación de la versión inicial de la plantilla; es decir, antes de hacer cualquier ajuste en el metamodelo y/o plantilla. En la tabla sólo aparecen los problemas que afectan a los cambios, por lo tanto, no se muestran los hallazgos relacionados con los aspectos básicos de la replicación.

Dado que *Mind* es el estudio que da lugar a la propuesta de plantilla, sus cambios se especificaron sin problemas, en su totalidad, con la versión inicial. Han sido *Req* y *Code* las familias que han provocado la evolución de la plantilla desde su versión inicial. La familia *Req* al tener mayor número de cambios, ha presentado más problemas y por tanto su especificación ha sido más laboriosa.

Cabe señalar que:

- Se trata de 3 familias de experimentos, constituidas por 14 replicaciones y 58 cambios.
- En 9 de los 58 cambios, ha habido problemas para especificar las *amenazas a la validez*.

Tabla 7.3: Hallazgos utilizando la versión inicial de la plantilla

Familia	Replicación	Núm. cambios	Amenazas	Contexto	Razón	Nombre
<i>Mind</i>	Mind#2	3				
	Mind#3	1				
<i>Req</i>	Q-2009	5	Cambio-1 Cambio-2 Cambio-5		Cambio-3 Cambio-5	
	Q-2011	6	Cambio-2 Cambio-6		Cambio-2 Cambio-6	
	Q-2012	4				
	E-2012A	10	Cambio-1			
	E-2012B	3		Cambio-3	Cambio-1 Cambio-2 Cambio-3	
	E-2013	2			Cambio-2	Cambio-2
	E-2014	2				Cambio-2
	E-2015	1				Cambio-1
	<i>Code</i>	VV-UPM1	6	Cambio-1		
	VV-UPV	6				
	VV-Uds	5				
	VV-ORT	4	Cambio-1 Cambio-2			
Σ	14	58	9	1	8	4

- 1 cambio no se podía especificar por afectar a una *variable de contexto*. Dichas variables no estaban identificadas en el metamodelo ni por tanto en la plantilla.
- Estos 10 primeros cambios (9+1) son los que no se han podido especificar con la versión inicial de la plantilla y han provocado los ajustes en dicha versión.
- En 8 cambios no se conocía la *razón* para dicho cambio.
- Por último, en 4 cambios no se conocía el *nombre de la variable afectada*.
- Estos 12 cambios, 8 debidos a no especificar el motivo del cambio (*Razón*) y 4 debidos a no especificar el nombre de la variable (*Nombre*) son en realidad 11, ya que uno de los cambios presentaba los dos problemas. Estos cambios estaban incompletos o *subespecificados*.

La tabla §7.4 presenta los resultados anteriores (con la versión inicial de la plantilla) resumidos por familias de experimentos.

Tabla 7.4: Resumen de cambios por familia

	<i>Mind</i>	<i>Req</i>	<i>Code</i>	Σ
Número de replicaciones:	2	8	4	14
Número de cambios:	4	33	21	58
Número medio de cambios:	2	4.1	5.3	4.1
Cambios que han hecho evolucionar de la plantilla:				
- Núm. cambios que presentan el problema <i>Amenaza</i> :	0	6	3	9
- Núm. cambios que presentan el problema <i>Contexto</i> :	0	1	0	1
Σ	0	7	3	10
Número de cambios especificados con la versión inicial de la plantilla:				
	4	26	18	48
	(4-0)	(33-7)	(21-3)	(58-10)
%	(100%)	(79%)	(86%)	(83%)
Cambios subespecificados:				
- Núm. cambios que presentan el problema <i>Razón</i> :	0	8	0	8
- Núm. cambios que presentan el problema <i>Nombre</i> :	0	3	1	4

Teniendo en cuenta que hay 58 cambios en 14 replicaciones, la media es de 4.1 cambios por replicación.

Como ya se comentó anteriormente, de los 58 cambios, 10 cambios no han podido ser especificados ya sea por presentar problemas para especificar las *Amenazas* (9 cambios) o por afectar a las *variables de contexto* (1 cambio). Por tanto, con la versión inicial sólo se podían especificar 48 cambios.

Las limitaciones de la versión inicial de la plantilla, implican ajustes a la plantilla (por ejemplo, modificar el patrón-L de las *amenazas a la validez* o considerar las *variables de contexto*) y también resultan en ajustes al metamodelo.

Por otra parte, entre los 48 cambios inicialmente especificados con la plantilla, hay 11 cambios que no se han completado debido a la falta de información; es decir están subespecificados. 8 de estos cambios presentan el problema de no especificar la causa del cambio (*Razón*) y en 4 no aparece el nombre de la variable afectada (*Nombre*), 1 de los cambios presenta ambos problemas. Apoya la idea de que la plantilla es necesaria para que las replicaciones no estén *subespecificadas*. Es decir, podrían ser cumplimentadas fácilmente por investigadores que realicen la replicación y rellenen la plantilla.

7.5 ESTUDIO DE CASO: CASO-CIENCIAS

En este estudio de caso, se analiza la utilidad de la plantilla en replicaciones de experimentos que no pertenecen al área de IS. Para ello, se han seleccionado cinco familias de experimentos relacionadas con el área de *Agrobiología y Tecnología de alimentos*. En este campo, la experimentación es una práctica frecuente y es habitual volver a realizar los experimentos en varias temporadas para demostrar que los resultados son independientes de las condiciones climáticas. Es por ello que se ha pensado en este área para evaluar la plantilla.

La tabla §7.5 muestra las instituciones a las que pertenecen los investigadores que han evaluado la plantilla, así como las familias de experimentos seleccionadas junto con su *acrónimo*. La elección de las familias de experimentos ha sido un proceso laborioso; se han revisado las principales publicaciones de los investigadores involucrados en el estudio y se ha pedido a cada investigador que seleccione un experimento que presente al menos una replicación y, a ser posible, que haya sido publicado.

A continuación, se presentan brevemente las familias seleccionadas.

Tabla 7.5: Instituciones y familias de experimentos en Caso-Ciencias

Institución	Familia de experimentos	Acrónimo
ETSIA ¹	Descontaminación de suelos	Suelo
ETSIA ¹	Análisis de calidad del aceite de oliva virgen	Calidad
IRNAS-CSIC ²	Biodegradación en suelos agrícolas	Bio
IG-CSIC ³	Extracción de componentes del aceite de oliva	Oliva
IG-CSIC ³	Influencia de la dieta en la acumulación de colesterol	Dieta

¹E.T.S. Ingeniería Agronómica, Universidad de Sevilla

²Instituto de Recursos Naturales y Agrobiología del Consejo Superior de Investigaciones Científicas

³Instituto de la Grasa del Consejo Superior de Investigaciones Científicas

7.5.1 Descripción de las familias

Familia Suelo.

El experimento original y sus dos replicaciones se llevaron a cabo en la *E.T.S. Ingeniería Agronómica (ETSIA)*. El experimento original aparece descrito en [147] y en el trabajo fin de grado (TFG) [41]. Las dos replicaciones corresponden también a TFGs [60, 161].

El objetivo del experimento es evaluar el efecto de un *biosurfactante sobre la fitoremediación asistida de suelos contaminados*. Los acrónimos utilizados para hacer

referencia al experimento original y sus replicasiones son *Suelo-2016*, *Suelo-2018* y *Suelo-2019* respectivamente. Ambas plantillas han sido cumplimentadas por una investigadora que participó en el experimento completo, es decir tanto en el experimento original como en sus dos replicasiones.

Se han podido especificar completamente los 16 cambios que componen las dos replicasiones (*Suelo-2018*, 9 cambios) y (*Suelo-2019*, 7 cambios) basadas, en ambos casos, en el experimento original (*Suelo-2016*).

La tabla §7.6 muestra la especificación de los 5 primeros cambios de la replicación *Suelo-2018*.

Familia Calidad.

El experimento y sus dos replicasiones han sido llevado a cabo en la *ETSIA*. Este estudio es la base de una tesis actualmente en curso. El objetivo del experimento es evaluar el efecto de la *variedad de aceituna* en la calidad del *aceite de oliva virgen* obtenido.

Los acrónimos del experimento original y sus replicasiones son *Calidad-2016A*, *Calidad-2017P* y *Calidad-2017V* respectivamente.

La plantilla ha sido cumplimentada por un investigador que ha participado en el experimento y revisada por el autor del mismo. Se han especificado completamente las tres replicasiones y sus tres cambios.

Familia Bio.

La familia de experimentos *Bio* sobre procesos de *biodegradación en suelos agrícolas* [148], se llevó a cabo en el *Instituto de Recursos Naturales y Agrobiología del Consejo Superior de Investigaciones Científicas (IRNAS-CSIC)*. Los investigadores han analizado la utilidad de la plantilla, aunque debido a la falta de tiempo, no se llegó a cumplimentar. Como en el resto de las familias de experimentos, se han recogido las opiniones de los investigadores para su posterior análisis.

Familia Oliva.

El experimento original y su replicación [92] se han llevado a cabo en el *Instituto de la Grasa del Consejo Superior de Investigaciones Científicas (IG-CSIC)*. El objetivo del estudio es mejorar la *extracción de compuestos fenólicos (EPC)* del aceite de oliva virgen con *disolventes verdes (Deep Eutectic Solvents DES)*. Los acrónimos para el experimento original y su replicación son *Oliva-2015* y *Oliva-Des* respectivamente. El autor del experimento ha cumplimentado la plantilla especificando completamente la replicación y sus 11 cambios.

Tabla 7.6: Especificación de los primeros 5 cambios de la replicación *Suelo-2018* (familia *Suelo*) utilizando la versión final de la plantilla

Replicación	<i>Suelo-2018</i> replicación Interna basada en el experimento original <i>Suelo-2016</i>
Objetivo del experimento Descripción del experimento	<p>Evaluar el efecto de un biosurfactante en la fitorremediación de suelos contaminados.</p> <p>Se analiza si la adición del biosurfactante JBR-425 a dos tipos de suelos contaminados por cobre (Cu) (causa) reduce la concentración Cu en dicho suelo, mediante la determinación del Cu extraído (efecto) por las plantas Cebada (<i>Hordeum vulgare</i> L.) y Mostaza (<i>Brassica juncea</i> L.) germinadas en dichos suelos.</p> <p>A partir de 2 tipos de suelo, se preparan dos grupos de macetas. Un grupo usando la muestra de suelo contaminada artificialmente con Cu y al que se le ha añadido JBR-425 (tratamiento) y otro grupo al que no se le ha añadido JBR-425 ni Cu (grupo de control), En ambos grupos de macetas se plantan semillas de Cebada o de Mostaza, se las deja germinar y crecer un cierto tiempo, y luego se analiza su biomasa (procedimiento de medida) para determinar la concentración de metales (métrica) y analizar la diferencia entre los tratamientos.</p>
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Agronómica, Sevilla</i> en <i>Octubre-2016</i> y esta replicación, en <i>E.T.S. Ingeniería Agronómica, Sevilla</i> en <i>Marzo-2018</i>
Propósito de la replicación	Generalizar resultados Superar limitaciones del experimento base
Cambio #1	<p>Medio de cultivo</p> <p>Originalmente, el experimento se llevó a cabo en una cámara de cultivo.</p> <p>En la replicación, se llevó a cabo en un invernadero</p> <p>Con el fin de simular las condiciones naturales</p>
Dimensión Modificada Amenazas a la validez	<p>Contexto, específicamente el medio</p> <p>El cambio aumenta la validez externa ya que permite generalizar los resultados realizando la replicación en condiciones más cercanas a las naturales</p>
Cambio #2	<p>Tipos de plantas</p> <p>Originalmente, se utilizaron dos plantas: Cebada y Mostaza</p> <p>En la replicación, sólo se utilizó Mostaza</p> <p>Ya que se observó que sólo Mostaza era una planta acumuladora de metal</p>
Dimensión Modificada Amenazas a la validez	<p>Protocolo, específicamente los instrumentos de medida</p> <p>El cambio aumenta la validez del constructo ya que en la Mostaza se puede medir el efecto, es decir el Cu extraído</p>
Comentario	Al utilizar un solo tipo de planta, no afecta a los resultados. No se operacionaliza.
Cambio #3	<p>Tipos de suelos</p> <p>Originalmente, había dos tipos de suelo: Coria (pH=7.8) y Constantina (pH=5.5)</p> <p>En la replicación, sólo se utilizó el suelo de Constantina</p> <p>Ya que en el suelo de Coria el metal estaba fuertemente absorbido y no se extrae.</p>
Dimensión Modificada Amenazas a la validez	<p>Población, específicamente Tipo de suelo</p> <p>El cambio aumenta la validez del constructo ya que se asegura que el metal puede ser extraído del suelo. El Cu extraído es el efecto que se quiere medir</p>
Cambio #4	<p>Reducción de dosis de cobre</p> <p>Originalmente, las dosis de cobre (Cu) añadidas al suelo fueron 0, 500 y 1000 mg kg⁻¹</p> <p>En la replicación, las dosis de Cu se ajustaron a 0, 125, 250 y 500 mg kg⁻¹</p> <p>Ya que la dosis de cobre de 1000 mg kg⁻¹ era tóxica para la planta</p>
Dimensión Modificada Amenazas a la validez	<p>Operacionalización, específicamente la causa</p> <p>El cambio aumenta la validez del constructo ya que la dosis de cobre se ajusta a niveles no tóxicos para la planta</p> <p>El cambio aumenta la validez del interna al analizar el efecto en niveles de Cu válidos</p>
Cambio #5	<p>Forma de aplicación del cobre</p> <p>Originalmente, El cobre se aplicó como Nitrato de cobre</p> <p>En la replicación, El cobre se aplica como Sulfato de cobre</p> <p>Ya que el Sulfato de cobre es más accesible</p>
Dimensión Modificada Amenazas a la validez	<p>Operacionalización, específicamente la causa</p> <p>El cambio no afecta a la validez ya que el reactivo es equivalente</p>

Familia Dieta.

El experimento original y replicación [168], se llevaron a cabo en el IG-CSIC. Su objetivo es analizar el efecto de la *alimentación rica en ácido oleico en sujetos hipertriglicéridémicos*. Los acrónimos utilizados son *Dieta* y *Dieta-Hiper*. El autor del experimento ha cumplimentado la plantilla especificando completamente la replicación y su cambio.

La instanciación de la versión final de la plantilla en las replications de estas familias de experimentos puede verse en el apéndice §B.

7.5.2 Especificación de los aspectos generales de las replications

Al especificar los aspectos básicos de la replicación, los experimentadores han comentado una serie de cuestiones relacionadas con la terminología utilizada y en general con conceptos del área que aparecen enumerados a continuación:

- En esta área, se considera que una *replicación* es realizar un experimento repitiendo exactamente el experimento original, al no ser posible, el término *replicación* no se utiliza.
- Aunque no se llama *replicación*, es habitual realizar un experimento condicionado por limitaciones o lagunas encontradas en experimentos anteriores y no es frecuente que su objetivo sea *confirmar resultados*, se trata más bien de un proceso evolutivo en el que la *replicación* no se propicia. Una excepción es cuando el experimento se repite la próxima temporada para contrarrestar la influencia de las condiciones climáticas. En cualquier caso, los investigadores de Ciencias señalan que siempre hay cambios. Así por ejemplo, al añadir una enmienda orgánica (mantillo) al suelo y volver a realizar el experimento en las mismas parcelas, es ya una segunda dosis y las condiciones del suelo son por tanto distintas al experimento original. En estos casos es frecuente estudiar los efectos acumulativos.
- Se utiliza el término *repetición* para referirse a *bloques* de tratamiento que se repiten dentro del mismo experimento. Es muy frecuente en el tratamiento de suelos en los que se repiten los bloques (por ejemplo, macetas o bandejas) para asegurar que no haya otros factores que influyan en los resultados (por ejemplo, luz o humedad) y aumentar así la *validez interna*.
- Los términos *tipo de replicación* y *familia de experimentos* no se utilizan en esta área.

En la tabla §7.9 se compara la terminología utilizada en el *Caso-IngSoft (Ingeniería del Software)* con la utilizada en *Caso-Ciencias (Agrobiología y Tecnología de alimentos)* y con la utilizada en el *Caso-Automáticos* que se verán en la sección §7.6. El significado de los términos utilizados en IS se ha presentado en la sección §2.2 y los términos utilizados en Caso-Ciencias se han presentado en la sección §7.2.1.

Los datos de la tabla §7.9 se han obtenido pasando un cuestionario durante las reuniones mantenidas con los investigadores para seleccionar las familias de experimentos y presentar la plantilla. Posteriormente, los datos han sido confirmados consultando algunas de las publicaciones del área sobre experimentación en agricultura [79, 149, 158].

7.5.3 Especificación de los cambios de las replicaciones

Todos los cambios llevados a cabo en las replicaciones se han definido utilizando la plantilla. En algunos casos no se han especificado algunos campos. Al intentar cumplimentar las plantillas correspondientes a la familia *Suelo*, la investigadora de la ETSIA requirió nuestra ayuda para poder entender los conceptos de *amenazas a la validez* y *dimensión afectada* y tras discutir las diferentes posibilidades, cumplimentar dichos campos.

La tabla §7.7 presenta el número de cambios de cada replicación de cada familia y un comentario. 16 de los 31 cambios (52%), se han especificado completamente, es decir, especificando las *amenazas a la validez*, *dimensión afectada* y *razón para el cambio*.

Tabla 7.7: Hallazgos al especificar los cambios en Caso-Ciencias

Familia	Replicación	Núm cambios	Comentario
Suelo	Suelo-2018	9	Los cambios han sido completamente especificados
	Suelo-2019	7	Los cambios han sido completamente especificados
Calidad	Calidad-2016P	2	No se especifican las amenazas ni la dimensión modificada
	Calidad-2016V	2	No se especifican las amenazas ni la dimensión modificada
Bio			La plantilla no se ha cumplimentado
Oliva	Oliva-Des	11	No se especifican las amenazas ni la dimensión modificada
Dieta	Dieta-Hiper	1	No se especifican las amenazas ni la dimensión modificada
Σ	6	32	

Referente a las *amenazas a la validez*, cabe señalar que:

- Las *amenazas a la validez* no se especifican en los experimentos en esta área. Sin

embargo, se consideran de forma implícita ya que: *i)* siempre hay una muestra no tratada (por ejemplo, *suelo de control* no contaminado) para garantizar que los resultados son sólo consecuencia del *tratamiento* y *ii)* todo el diseño suele *repetirse* al menos tres veces para evitar efectos locales debido a que las muestras reciban más o menos luz, agua, etc. Ambos procedimientos aumentan la *validez interna*.

- Por otro lado, los experimentos se llevan a cabo en primer lugar, en *placas de Petri* o en *cámaras de cultivo* es decir condiciones muy controladas y posteriormente son llevados a invernaderos, jardines u otros, asegurando así su *validez externa*.

7.5.4 Discusión

En esta discusión o interpretación de los resultados, se analizan los comentarios recibidos por parte de los investigadores en las reuniones mantenidas tras cumplimentar la plantilla. En primer lugar se presentan los comentarios comunes (observaciones generales) y a continuación, los comentarios específicos por cada familia de experimentos.

Observaciones generales:

- *"A pesar del esfuerzo necesario para completar la plantilla, se compensa por el beneficio de registrar cada cambio y su justificación. Anteriormente, en muchos casos, no se registraban los cambios y sólo eran conocidos por el investigador que realizaba el experimento"*.
- *"Algunos campos de la plantilla son difíciles de cumplimentar"*.

Al describir el cambio es importante intentar especificar todos los campos de la plantilla incluida la *motivación* del cambio y su influencia en la *validez del experimento*.

Familia Suelo:

- *"La plantilla es útil ya que aunque se siga el experimento base de la forma más exacta posible, siempre se añaden algunos cambios (por ejemplo, se ajusta la dosis del tratamiento) que es valioso especificar claramente. De igual forma, es interesante dejar registrado el motivo del cambio"*.

- *"Muchos de los cambios son impuestos por el medio, en concreto, por la disponibilidad del material experimental (por ejemplo, se utilizan los productos químicos o los tipos de macetas disponibles) y deben de registrarse".*
- *"Sería útil disponer de un campo para guardar el orden en que se producen los cambios".*

Se les recomienda, por claridad, describir los cambios en orden cronológico, es decir empezar describiendo los cambios relacionados con el *cultivo* y seguir el proceso hasta describir en último lugar los cambios relativos a la *cosecha*.

Familia Calidad:

- *"Sería útil distinguir entre el lugar donde se realiza el experimento y el lugar donde se diseña y se realizan los análisis del experimento".*

Se les recomienda utilizar el campo de *comentarios* disponible en la plantilla, para aclaraciones de este tipo.

Familias Dieta y Oliva:

- *"No es frecuente publicar sólo la repetición del experimento. La práctica común es su publicación junto con el experimento original. Por lo tanto, es importante distinguir las condiciones en ambos experimentos. La plantilla facilita el registro de los cambios introducidos al variar las condiciones experimentales o de otro tipo".*

Es necesario recordar que no se utiliza el término replicación aunque están repitiendo el experimento con alguna variación.

Familia Bio:

- *"Debido al espacio necesario para incluir la plantilla en la publicación de la replicación, la utilizarían como material auxiliar a incluir en el paquete de laboratorio".*

Este comentario, se ha tenido en cuenta para incorporarlo a la propuesta.

7.6 ESTUDIO DE CASO: CASO-AUTOMÁTICOS

En este tercer estudio de caso, la plantilla ha sido probada en *experimentos automáticos*. Este tipo de experimentos, habituales en el área de conocimiento de IS, implican la implementación de *algoritmos* para llevar a cabo el experimento.

Se contactó con dos experimentadores que actualmente están realizando *experimentos automáticos*, se revisaron sus principales publicaciones sobre el tema y se seleccionaron dos familias de *experimentos automáticos*. Una vez que se cumplieron las plantillas, fueron validadas por ambos experimentadores. Como último paso, se les pasó un cuestionario sobre la terminología utilizada en *experimentos automáticos*.

En las siguientes secciones se describen brevemente las familias seleccionadas.

7.6.1 Descripción de las familias

Familia Testing.

El experimento y sus tres replications internas [171] se llevaron a cabo en la E.T.S. Ingeniería Informática (ETSII). Su objetivo es *priorizar los casos de pruebas* de forma que el orden en que se ejecuten intente acelerar la detección de fallos. Los acrónimos utilizados para hacer referencia al experimento original y sus replications son *Test-F*, *Test-NF*, *Test-F&NF* y *Test-FvsNF*. Cada replicación consta de un cambio.

Familia SPL.

Consta de un experimento y una replicación interna [200] realizados en la ETSII. Su objetivo es comparar la *eficacia de cinco criterios de priorización diferentes* para aumentar la tasa de detección temprana de fallos en Línea de Producto Software (SPL).

Los acrónimos utilizados para hacer referencia al experimento original y su replicación son *SPL-Pr* y *SPL-Pr&Com*.

La tabla §7.8 muestra la especificación de la replicación *SPL-Pr&Com* con sus dos cambios.

La instanciación de la versión final de la plantilla en las replications de las dos familias de experimentos puede verse en el apéndice §B.

Tabla 7.8: Replicación SPL-Pr&Com, versión final

Replicación	<i>SPL-Pr&Com</i> replicación Interna basada en el experimento original <i>SPL-Pr</i>
Objetivo del experimento	Comparación de los criterios de priorización de casos de prueba para Líneas de Productos Software (SPL).
Descripción del experimento	..
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2014 y esta replicación, en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2014
Propósito de la replicación	Generalizar resultados
Cambio #1	Originalmente , sólo se generó un conjunto de pruebas En la replicación , para cada modelo, se generó un conjunto de pruebas Con el fin de obtener una lista de productos que cubra todos las posibles características de cada modelo
Dimensión Modificada	Protocolo , específicamente las guías
Amenazas a la validez	El cambio no afecta a la validez
Cambio #2	Originalmente , el conjunto de pruebas se generó aleatoriamente utilizando la herramienta SPLAR En la replicación , el conjunto de pruebas se generó aleatoriamente usando la herramienta SPLCAT Ya que SPLCAT aumenta la tasa de detección de fallas y por lo tanto es considerado como un enfoque de priorización extra en nuestra comparación.
Dimensión Modificada	Protocolo , específicamente el material experimental
Amenazas a la validez	El cambio no afecta a la validez ya que sólo cambia la herramienta utilizada

7.6.2 Especificación de los aspectos generales de las replications

No ha habido nuevos hallazgos en la especificación de los aspectos básicos de la replicación.

En cuanto a la terminología, la tabla §7.9 compara la terminología utilizada específicamente en los *experimentos automáticos* con la utilizada en los experimentos analizados anteriormente. Cabe señalar que aunque la replicación es una práctica común, no se utiliza la expresión *familia de experimentos* y, al ser *experimentos automáticos*, tampoco se utilizan los términos *sujeto* y *objeto experimental*.

Los experimentadores han señalado que "*sería interesante añadir un campo de texto libre y opcional para describir la motivación de la replicación*". En la plantilla propuesta los posibles valores para el objetivo de la replicación están prefijados (*confirmar resultados, generalizar resultados y superar limitaciones del experimento base*) y se justifica específicamente el *motivo* para llevar cada cambio en la descripción de dicho cambio por lo que no parece necesario.

Tabla 7.9: Comparación de la nomenclatura utilizada en el *Caso-IngSoft*, *Caso-Ciencias* y *Caso-Automáticos*

<i>Caso-IngSoft</i>	<i>Caso-Ciencias</i>	<i>Caso-Automáticos</i>
Replicación	Repetición con cambios	Replicación
Repetición	Repetición	Repetición
Reproducción	X	Reproducción
Tipo de replicación	X	Tipo de replicación
Familia experimentos	X	X
Var. dependiente	Var. dependiente	Var. dependiente
Var. independiente ¹	Var. independiente	Var. independiente
Var. de contexto	X	Var. de contexto
Var. de bloqueo	Var. de bloqueo	Var. de bloqueo
Parámetros	Parámetros	Parámetros
Factor	Factor	Factor
Nivel	Nivel	Nivel
Sujeto experimental	Sujeto experimental ²	X
Objeto experimental	X	X
Población	Población	Población
Amenazas a la validez	X	Amenazas a la validez
Validez interna	X	Validez interna
Validez externa	X	Validez externa
Validez del constructo	X	Validez del constructo
Validez de conclusión	X	Validez de conclusión
Diseño experimental	Diseño experimental	Diseño experimental
Dimensión afectada	X	X
Operacionalización	X	X
Protocolo ³	Procedimiento	Protocolo
Tratamiento	Tratamiento	Tratamiento
Guías	Guías	Guías

¹ También se denomina factor

² También se denomina unidad experimental

³ También se denomina procedimiento experimental

7.6.3 Especificación de los cambios de las replicaciones

La tabla §7.10 presenta el número de cambios de cada replicación de cada familia y un comentario. Todos los cambios han sido completamente especificado, incluyendo las *amenazas a la validez, dimensión afectada y razón para el cambio*.

Tabla 7.10: Hallazgos al especificar los cambios en Caso-Automáticos

Familia	Replicación	Núm. cambios	Comentario
Testing	Test-NF	1	El cambio ha sido completamente especificado
	Test-F&NF	1	El cambio ha sido completamente especificado
	Test-FvsNF	1	El cambio ha sido completamente especificado
SPL	SPL- FP&Com	2	Los cambios han sido completamente especificados
Σ	4	5	

El concepto de *dimensión afectada por el cambio* no se utiliza en los *experimentos automáticos* aunque aquí haya sido cumplimentada.

Los autores confirman que cuando se diseñan y reportan experimentos, es una práctica común analizar las *amenazas a la validez*. Es frecuente revisar los cuatro tipos de amenazas que aparecen en la plantilla: *interna, externa, de construcción y de conclusión*. Por lo tanto, tal y como aparece en la plantilla, las amenazas de validez deben formar parte de la especificación del cambio.

7.6.4 Discusión

Entre los comentarios recibidos, los experimentadores señalan "*la utilidad de la plantilla en experimentos automáticos al permitir especificar tanto la situación antes y después de cada cambio como el motivo del cambio*".

También se destaca el interés en analizar las implicaciones del cambio en *las amenazas a la validez*, especialmente la *validez interna y externa* que son de mayor interés en este tipo de experimentos.

7.7 RESULTADOS DEL ESTUDIO DE CASO MÚLTIPLE

En esta sección se abordan cada una de las preguntas de investigación propuestos en la sección §7.3.1 utilizando tablas para visualizar los datos recopilados y proporci-

nando las interpretaciones posteriores.

7.7.1 Expresividad

PI_{4.1}: *Expresividad.* ¿Qué proporción de cambios ha sido posible definir? ¿La *expresividad* de la plantilla es mayor que la de otras propuestas seguidas?.

Al definir las replicaciones del primer estudio de *Caso-IngSoft* con la versión inicial de la plantilla, se detectan limitaciones en la *expresividad* de la plantilla y se mejora en consecuencia.

En concreto, se echan de menos algunos campos para registrar información relevante y se detecta que algunos patrones-L no permiten expresar lo que el experimentador necesitaba reflejar. Las limitaciones encontradas están relacionadas con la incapacidad para expresar:

- Lugar y fecha del experimento base y su replicación.
- Breve descripción del objetivo y del experimento base.
- Acrónimo del experimento base y su replicación.
- Replicación basada en otra replicación anterior.
- *Amenazas a la validez* de algunos cambios.
- Cambios que afectan a variables de contexto.
- Cambios cuya razón es desconocida.
- Cambios en los que se desconoce el nombre de la variable afectada.

Las secciones §7.4.2 y §7.4.3 detallan la descripción de estos problemas en el *Caso-IngSoft* junto con el ajuste realizado en la plantilla y/o metamodelo. Tras los ajustes, la versión actual de la plantilla permite expresar las replicaciones de las familias pertenecientes a los tres estudios de caso y específicamente todos sus cambios.

Por otro lado, es interesante analizar la *expresividad* de la plantilla con respecto a otras propuestas seguidas. En la tabla §7.11 se compara la plantilla propuesta con las guías identificadas en los estudios analizados en la sección §5.2 (ver tabla §5.1).

Tabla 7.11: Comparación de la plantilla con las guías seguidas para reportar replicaciones a nivel de expresividad de las propuestas

Guías	Foco	Motivación experimento	Cambios al original	Motivación del cambio	Amenazas validez	Uso de patrones
Carver [42]	RE	✓	✓	✓	✗	✗
Wohlin et al. [245]	EX	✓	✗	✗	✓	✗
Jedlitschka et al. [118]	EX	✓	✗	✗	✓	✗
Juristo&Moreno[125]	EX	✓	✗ ¹	✗	✓	✗
Runeson&Höst [196]	CS	✓	✗	✗	✓	✗
Esta propuesta	RE	✓	✓	✓	✓	✓

¹Se refiere a los cambios como características modificadas

Es conveniente destacar que las únicas guías específicas para reportar replicaciones son las de Carver [42] el resto se utilizan adaptadas.

Para cada guía se analizan los siguientes aspectos:

- *Foco*. Indica si las guías están destinadas a informar sobre experimentos controlados (EX), estudios de casos (CS) o específicamente sobre replicaciones (RE).
- *Motivación del experimento*. Al analizar la recomendación de la guía, se encuentra que en todas las guías se aconseja reportar la *motivación del experimento*.
- *Cambios al experimento original*. En las guías de Carver [42] se recomienda reportar los cambios al experimento original junto con la causa del cambio. Del mismo modo, Juristo y Moreno [125] al referirse a experimentos anteriores, recomiendan indicar las características modificadas.
- *Amenazas a la validez*. Otra de las cuestiones que deberían de abordarse en el informe del experimento son las *amenazas a la validez*. Se menciona en todas las guías, excepto en las de Carver.
- *Uso de patrones*. Ninguna de las directrices analizadas utiliza patrones. En la plantilla propuesta, el uso de patrones-L facilita la definición del cambio y por lo tanto aumenta la *expresividad*.

Además de estas guías, Gómez *et al.* [97] abordan en profundidad las características de los cambios en las replicaciones pero no se proporciona una guía para su informe por lo que no se incluye en la tabla.

En la propuesta presentada en esta memoria es obligatorio especificar, entre otros, el motivo para llevar a cabo la replicación y la definición del cambio junto con su *motivación específica*.

7.7.2 Precisión

PI_{4.2}: *Precisión.* El uso de los patrones-L, ¿hace más precisa la definición de los cambios que el lenguaje natural y/o en forma de tablas? ¿Permite detectar la información que falta?.

La especificación de los cambios por medio de la plantilla y los patrones-L es más precisa que en lenguaje natural, ya que disminuye la falta de información.

Al especificar los cambios en el primero de los casos de estudio *Caso-IngSoft*, se encontraron cambios *subespecificados* en los que no se especificaba *la razón del cambio* y/o el *nombre de la variable* afectada por el cambio.

Asimismo, gracias a la *precisión* de la plantilla, se constata que hay campos en la plantilla que son difíciles de cumplimentar especialmente por razones de terminología o cuando la plantilla es cumplimentada por alguien que no es el autor de la replicación.

7.7.3 Usabilidad

PI_{4.3}: *Usabilidad.* ¿Cuántos campos de la plantilla han presentado problemas de comprensión? ¿Están los investigadores familiarizados con la terminología?.

La *usabilidad* mide el esfuerzo que debe invertir el usuario para utilizar el sistema y es una de las características establecidas por el estándar ISO 9126 para evaluación de la calidad del software. La *usabilidad* consiste en un conjunto de atributos entre los que destaca la *comprensibilidad* o esfuerzo requerido por los usuarios para conocer las estructuras y conceptos relativos al uso del sistema.

Aplicado a la plantilla propuesta, para verificar su *comprensibilidad* es necesario que la plantilla sea utilizada por experimentadores externos. Los casos *Caso-Ciencias* y *Caso-Automáticos*, han sido diseñados con el objetivo principal de que la plantilla sea usada por los experimentadores que han llevado a cabo las replicaciones. Los comentarios recibidos de los experimentadores se incluyen en las secciones §7.5.4 y §7.6.4.

La tabla §7.12 compara el uso de la plantilla considerando todas las familias de experimentos analizadas en el presente estudio de caso múltiple. Para cada familia de

experimentos, se muestra el número de replicaciones, el número de cambios y si se ha completado *el motivo de cada cambio, la dimensión modificada y las amenazas a la validez*

Tabla 7.12: Uso comparativo de la plantilla

Caso	Familia	Núm. de replicaciones	Núm. de cambios	Motivación del cambio	Dimensión modificada	Amenazas validez
<i>Caso-IngSoft</i>	<i>Mind</i>	2	4	✓	✓	✓
	<i>Req</i>	8	33	✓(75%) ✗(25%)	✓	✓
	<i>Code</i>	4	21	✓	✓	✓
<i>Caso-Ciencias</i>	<i>Suelo</i>	2	16	✓	✓	✓
	<i>Calidad</i>	2	4	✓	✗	✗
	<i>Bio</i>	–	–	–	–	–
	<i>Oliva</i>	1	11	✓	✗	✗
	<i>Dieta</i>	1	1	✓	✗	✗
<i>Caso-Automáticos</i>	<i>Testing</i>	3	3	✓	✓	✓
	<i>SPL</i>	1	2	✓	✓	✓
Σ		24	95			

Cabe señalar que los 95 cambios se han especificado plenamente, incluido el *motivo del cambio*; sin embargo, campos como *dimensión modificada* y *amenazas a la validez*, no se han cumplimentado en las familias de experimentos pertenecientes a *Caso-Ciencias* ya que son conceptos que no se utilizan en dicha área. Por lo tanto, sólo estos dos campos han presentado problemas de comprensión.

En la familia *Suelo*, se ha colaborado con la experimentadora que ha cumplimentado la plantilla para asegurar que se identifican las *amenazas a la validez* y la *dimensión modificada*

La *razón para el cambio* no se ha cumplimentado en algunos cambios de la familia *Req* al no estar especificado en el informe de la replicación.

Las diferencias en la *Comprensibilidad* de la plantilla pueden deberse a:

- *Caso-IngSoft*. La plantilla ha sido propuesta en el área de IS y ha sido cumplimentada a partir del informe de replicación, por lo que lógicamente no ha habido problemas de comprensión.
- *Caso-Ciencias*. Algunos conceptos han sido laborioso o imposible de rellenar.

Por otro lado, el resto de la plantilla no ha presentado problemas de *comprensibilidad*, en parte debido a la ayuda de los patrones-L que facilitan la redacción de los cambios.

- *Caso-Automáticos*. Las familias en la que se ha instanciado la plantilla pertenecen al área de IS, por lo que la terminología es conocida y la plantilla se adapta bien a este tipo de experimentos.

Para resumir la *comprensibilidad* de la plantilla, en la tabla §7.13 se presentan sobre la propia plantilla, los campos que han presentado problemas para su cumplimentación ya que su terminología no se ajusta –o no es válida– a la del área de *Agrobiología y Tecnología de alimentos* (denotado con ✕) o bien presentan alguna singularidad con respecto al área de IS. Con ✓ se denota que el significado es el mismo. En la tercera columna, se añade un comentario sobre la singularidad del campo o como se adaptaría dicho campo de la plantilla al área analizada.

7.7.4 Trazabilidad

PI_{4.4}: *Trazabilidad*. Al registrar los cambios, ¿se facilita la *trazabilidad* entre replications?.

La *trazabilidad* permite conocer la evolución que ha tenido un experimento mediante las replications que constituyen la familia de experimentos. El uso de un *acrónimo* para identificar la replicación favorece dicha *traza* entre replications.

Mediante la *trazabilidad* se facilitan nuevas replications ya que al registrar tanto el *objetivo de cada replicación* como la *razón de cada cambio*, se conocen los cambios ya realizados y se pueden diseñar y proponer nuevos cambios. Cuando se necesita hacer cambios para adaptar el experimento a un nuevo entorno, puede ser útil analizar los cambios ya realizados en entornos de similares condiciones.

Como ya se comentó en la sección §4.3, tras la realización de la replicación, se debe actualizar el *paquete de laboratorio* añadiendo la descripción detallada de la replicación. De esta forma, se establece la *traza* quedando reflejada la evolución histórica del experimento mediante las replications llevadas a cabo y que entran a formar parte del *paquete de laboratorio*.

Espinosa *et al.* [68], abordan el problema de la administración del material experi-

Tabla 7.13: Singularidades de los campos de la plantilla en el Caso-Ciencias

Campos de la plantilla propuesta		Comentarios
Replicación ✗	Replicación Interna ✗ Replicación Externa ✗ Replicación previa ✗ Experimento base ✗	Repiten el experimento en otra temporada o variando las condiciones dependiendo de resultados previos
Objetivo del experimento	GQM ✗	El objetivo se describe mediante texto libre
Lugar y Fecha	✓	
Propósito de la replicación	i) Confirmar resultados ✓ ii) Generalizar resultados ✓ iii) Superar limitaciones del experimento base ✓	
Dimensión Modificada ✗	i) Operacionalización ✗ ii) Población ✓ iii) Protocolo ✗ iv) Experimentadores ✓	i) Se refieren a <i>definición de variables dependientes e independientes</i> iii) Se refieren a <i>materiales y métodos</i>
Amenazas a la validez ✗	i) Validez interna ✗ ii) Validez externa ✗ iii) Validez del constructo ✗ iv) Validez de conclusión ✗	i) Refieren <i>error experimental</i> ¹ ii) Repiten el experimento ² iii) Realizan <i>ensayos preliminares</i> iv) Se aumenta la muestra ³

¹ Se aborda mediante diseño de bloques, repeticiones del tratamiento, etc.

² Se repite el experimento en placas de Petri, invernadero, campo, etc.

³ Mediante la repetición de bloques de tratamiento

mental en IS, señalando que los experimentadores necesitan disponer de información eficiente sobre el estado y *trazabilidad* de los experimentos y materiales para la planificación y ejecución de nuevas repeticiones.

La tabla §7.14 se comparan los aspectos de *expresividad*, *precisión*, *usabilidad* y *trazabilidad* en las propuestas (textual o tablas) identificadas al analizar los estudios de la sección §5.2 con la plantilla aquí propuesta.

Tabla 7.14: Comparación de la plantilla con las propuestas seguidas para reportar los cambios

Propuesta	Expresividad	Precisión	Usabilidad	Trazabilidad
Textual	-	-	-	+
Tablas	+	+	+	+
Esta propuesta	++	++	+	++

Muy alta (++), Alta (+), Baja (-)

La *usabilidad* de la plantilla propuesta, desde el punto de vista de la *comprensibilidad* por parte de los experimentadores, es alta (+). Sin embargo, puede resultar complicado incluir la plantilla en el reporte de la replicación debido al espacio que ocupa y, como ya se ha comentado, sería muy útil su inclusión en el paquete de laboratorio.

Por otro lado, la *expresividad*, *precisión* y *trazabilidad* de la plantilla propuesta es muy alta (++)

7.8 CONCLUSIONES

El estudio de caso múltiple ha permitido la validación y evolución de la plantilla propuesta mediante su aplicación a casos reales e incluyendo la validación en otras áreas distintas a la IS. A continuación, se destaca la principal aportación de cada uno de los estudios de caso.

Caso-IngSoft. Ha permitido detectar la falta de información relevante, así como la necesidad de mejorar la redacción o frases que se proponen como parámetros en algunos de los patrones-L definidos. Una vez incorporadas las mejoras pertinentes, la plantilla ha evolucionado hasta su versión actual.

Caso-Ciencias. Ha revelado la diferente terminología e incluso conceptos distintos utilizados al llevar a cabo un experimento.

Por otra parte, su uso en este área ha reforzado la idea de que la plantilla ayuda en el proceso de definición de los cambios para que queden claramente documentados ya que así lo han reconocido los experimentadores del área tras su utilización.

Caso-Automáticos. Ha servido para confirmar la utilidad de la plantilla en otro tipo de experimentos, tales como *experimentos automáticos* aunque ya en *Caso-Ciencias* se había utilizado la plantilla en replicaciones con sujetos experimentales no humanos.

Cabe señalar, que la plantilla ha demostrado su utilidad al especificar todos los cambios de las replicaciones del estudio de caso múltiple, sin embargo hay campos de la plantilla, tales como la *razón para el cambio*, las *amenazas a la validez*, la *dimensión y elementos* de la configuración experimental afectados por el cambio que aunque sabemos de la dificultad de su cumplimentación ya sea por desconocimiento o bien debido a la distinta terminología que es recomendable especificar para documentar completamente el cambio.

Debido al espacio que ocupa y la limitación del tamaño del reporte de la replicación en artículos tanto de congreso como de revista, sería conveniente la inclusión de la plantilla en el *paquete de laboratorio* del experimento. De esta forma, además, queda reflejada la *evolución del experimento* en las sucesivas replicaciones.

Como conclusión, se recomienda el uso de la plantilla, por un lado: *i)* para reportar replicaciones *internas* ya que los cambios son bien conocidos y quedarían registrados en la plantilla formando parte del *paquete de laboratorio* y por otro lado: *ii)* para experimentadores que necesitan conocer los cambios llevados a cabo por otros experimentadores accediendo a las plantillas que ya han sido cumplimentadas y conocer como ha ido evolucionando el experimento en las sucesivas replicaciones para poder plantear nuevos cambios mediante una nueva replicación, en este caso *externa* que son las de mayor poder de confirmación [34].

7.9 RESUMEN

Para validar la plantilla, se ha instanciado mediante un *estudio de caso múltiple* que abarca tres estudios de caso en: *i)* Ingeniería del Software (*Caso-IngSoft*); *ii)* Agrobiología y Tecnología de alimentos (*Caso-Ciencias*); y *iii)* Experimentos automáticos (*Caso-Automáticos*).

Se han analizado los siguientes aspectos de la plantilla:

Expresividad. Se han modificado algunos patrones-L para expresar lo que el experimentador necesita dejar reflejado. Una vez realizados los ajustes oportunos, se han podido expresar todos los cambios de las replicaciones del *estudio de caso múltiple*.

Por otra parte, se analiza la *expresividad* de la plantilla con respecto a otras propuestas. Destacar que la única propuesta específica para replicaciones son las guías de Carver [42]. Suelen seguirse otras guías enfocadas a experimentos en general en las que no se hace referencia a describir los cambios ni su motivación. En las guías de Carver no se contemplan las *amenazas a la validez*. Ninguna de las propuestas proporciona patrones-L para facilitar la redacción de los cambios.

Precisión. La especificación de los cambios por medio de la plantilla y los patrones-L es más *precisa* que en lenguaje natural, ya que permite detectar la falta de información relevante.

Usabilidad. Los estudio de caso *Caso-Ciencias* y *Caso-Automáticos* han sido diseñados con el objetivo principal de que la plantilla sea usada por otros experimentadores para analizar así la *usabilidad* y específicamente la *comprensibilidad* de plantilla facilitada mediante el aprendizaje y uso de los patrones-L. Se ha comparado la terminología utilizada en los tres *estudios de caso* respecto a los conceptos utilizados en la plantilla.

Trazabilidad. Permite conocer como ha evolucionado el experimento siguiendo la *traza* entre replicaciones sucesivas de la misma familia de experimentos y facilitando así, nuevas replicaciones.

PARTE V

CONCLUSIONES

CONCLUSIONES Y TRABAJOS FUTUROS

Lo importante es no dejar de hacerse preguntas

Albert Einstein (1879 – 1955),

Físico alemán

Este capítulo concluye la memoria, presentando una visión general del trabajo realizado mediante el análisis de los objetivos y las conclusiones.

En la sección §8.1 se analiza el cumplimiento de los objetivos propuestos. En la sección §8.2 se resumen las publicaciones derivadas de este trabajo. Por último, en la sección §8.3 se apuntan las líneas de trabajo futuro.

8.1 CONSECUCIÓN DE OBJETIVOS

Al principio de la memoria (sección §1.3) se planteó el principal objetivo de este trabajo de **facilitar la especificación sistemática de los cambios para apoyar el proceso de replicación de experimentos controlados**. En esta sección se analiza el cumplimiento de cada uno de los objetivos parciales definidos para conformar dicho objetivo principal.

1. *Estudiar el estado de las replications y evaluar su madurez y tendencias*. En el capítulo §3 se han analizado estudios empíricos que reportan replications en IS publicados entre 2013 y 2018. El estudio ha permitido:
 - i) Identificar los principales foros para conocer las revistas y congresos donde publicar futuros estudios de replicación.
 - ii) Conocer los autores e instituciones que albergan el conocimiento sobre replications. Facilita acceder al trabajo de los autores y la realización de estancias de investigación.
 - iii) Detectar vacíos de investigación. Al analizar los temas de investigación y los métodos empíricos más utilizados se ha conocido la distribución de estudios, así como los temas y métodos en los que no se han encontrado replications.
2. *Principales problemas identificados en la práctica de las replications*. Algunos de los problemas detectados en el capítulo §4, se pueden paliar en parte mediante la plantilla propuesta. De esta forma:
 - i) El *conocimiento tácito* puede ser revelado por medio de la especificación del cambio mediante la plantilla ya que la motivación forma parte de la descripción del cambio.
 - ii) La inclusión de la plantilla en el *paquete de laboratorio* permite conocer cómo ha evolucionado un experimento original dentro de una familia, por qué se han realizado cambios y de qué tipo son, facilitando así la planificación de nuevos cambios al tener constancia de los cambios ya realizados.
 - iii) Ante la *falta de guías*, hasta donde nosotros sabemos, las directrices de Carver [42] son las únicas enfocadas específicamente a replications. El resto de guías se aplican a las replications pero no dicen nada sobre la forma de especificar los cambios. La plantilla ayuda a explicitar, de forma sistemática, la información sobre los cambios que caracterizan a las replications y sus implicaciones en la validez del experimento.

3. *Analizar cómo se reportan actualmente los cambios.* En el capítulo §5 se han analizado la estructura y contenido de 25 estudios de replicación publicados en los dos principales foros del área de IS empírica: *Empirical Software Engineering Journal (EMSE)* y *International Conference on Empirical Software Engineering and Measurement (ESEM)*. Los principales hallazgos han sido:
 - i) La mayoría de los estudios analizados (76%) especifican sus cambios. De éstos, el 47% se ayudan de tablas para describir los cambios, además de formato textual.
 - ii) Para cada cambio, se suele especificar la situación en el original y en la replicación. La motivación para llevar a cabo el cambio aparece en el 37% de los estudios que especifican sus cambios.
 - iii) Las tablas suelen presentar una estructura común, la propiedad cambiada (el cambio) se define en las filas. En las columnas aparece el valor de dicha propiedad en el original y para cada replicación. El motivo del cambio no suele aparecer en la tabla.
 - iv) Los cambios se especifican con más frecuencia en las replicaciones *internas* que en las *externas*. El 100% de los estudios con replicaciones *internas* describen sus cambios frente a sólo el 50% de los estudios con replicaciones *externas*.
 - v) La guía de Carver [42] es seguida en el 44% de los estudios, principalmente cuando el tipo de replicación es *interna*, el original no se reporta y especialmente cuando se reporta una sola replicación. En 40% de los estudios no se sigue ninguna guía. El resto de guías seguidas [118, 125, 196, 245] no son específicas para replicaciones.
4. *Sistematizar la definición de los cambios mediante la plantilla propuesta.* En el capítulo §6 se ha propuesto la plantilla basada en el metamodelo que formaliza la información involucrada en las replicaciones y con *patrones-L*. Su objetivo es dejar constancia de la información necesaria para especificar los cambios junto con la *dimensión afectada* y sus implicaciones en la *validez del experimento*. La plantilla facilita la definición de los cambios y detecta la falta de información relevante.

La plantilla propuesta en esta memoria para especificar los cambios en las replicaciones, puede incorporarse al *paquete de laboratorio* facilitando así la traza entre sucesivas replicaciones.
5. *Extender la propuesta a otras áreas.* En el capítulo §7 se ha validado la plantilla mediante un *estudio de caso múltiple* que abarca las áreas de *Agrobiología y Tecnología*

de alimentos y otros tipos de experimentos tales como *experimentos automáticos*

Se ha identificado la diferente terminología e incluso conceptos distintos, relativos a las replicaciones, existentes en es área de *Agrobiología y Tecnología de alimentos*.

La plantilla propuesta se ha validado en dichas áreas, analizando los siguientes aspectos:

i) Expresividad. Se han modificado algunos patrones-L para expresar lo que el experimentador necesita dejar reflejado. Una vez realizados los ajustes oportunos, se han podido expresar todos los cambios de las replicaciones del *estudio de caso múltiple*.

Algunos campos de la plantilla han sido más difíciles de cumplimentar ya que a veces se desconocen, especialmente en el área de *Ciencias* por razones de terminología o cuando la plantilla es cumplimentada por alguien que no es el autor de la replicación.

ii) Precisión. La especificación de los cambios por medio de la plantilla y patrones-L es más *precisa* que en lenguaje natural, ya que permite detectar la falta de información relevante.

iii) Usabilidad. La plantilla ha sido validada por experimentadores que han llevado a cabo replicaciones de experimentos controlados en otras áreas y tipos de experimentos. Mediante el feedback recibido, se han analizado las singularidades del área de *Ciencias* y se ha comparado la terminología con respecto a los conceptos utilizados en la plantilla.

iv) Trazabilidad. La cumplimentación de la plantilla facilita seguir la *traza* de la evolución del experimento en las sucesivas replicaciones de la familia de experimentos.

8.2 PUBLICACIONES

Los resultados parciales obtenidos durante la realización del presente trabajo de investigación han dado lugar a las publicaciones que se enumeran a continuación.

8.2.1 Revistas internacionales

- [51] M. Cruz, B. Bernárdez, A. Durán, J. A. Galindo and A. Ruiz-Cortés. *Replication of Studies in Empirical Software Engineering: A Systematic Mapping Study from 2013 to 2018*. IEEE Access 8: 26773-26791, 2020

Disponible en: <https://ieeexplore.ieee.org/document/8894007>.

DOI: 10.1109/ACCESS.2019.2952191

Índice de impacto: 4,098 (JCR IF: Q1)

En este trabajo se presenta la revisión sistemática llevada a cabo para conocer el estado actual de las replicaciones de experimentos en IS. Se han identificado entre otros, los principales foros donde se publican replicaciones, autores, instituciones, tendencias, métodos empíricos utilizados y temas emergentes en los últimos años.

- [17] C. Arévalo, I. Ramos, J. Gutierrez y M. Cruz. Practical Experiences in the Use of Pattern-Recognition Strategies to Transform Software Project Plans into Software Business Processes of Information Technology Companies. *Scientific Programming*. 2019. Vol. 2019. Núm. 2019. Pag. 1-21.

Disponible en: <https://doi.org/10.1155/2019/7973289>

Índice de impacto: 1,289 (JCR IF: Q3)

En este trabajo se propone un framework para la extracción de procesos focalizando en las reglas temporales ocultas en las bases de datos de sistemas heredados (Legacy Systems). Se aplicó al estudio de caso de AQUAWS que utilizaba MS-Project como Legacy System.

- El estudio de caso múltiple aquí presentado que ha permitido la validación de la plantilla propuesta abarcando las áreas de Ingeniería del Software, *Agrobiología y Tecnología de alimentos* y *Experimentos automáticos*, será próximamente enviado a una revista científica.

8.2.2 Capítulo de libro

- [26] B. Bernárdez, M. Cruz, A. Durán, J. A. Parejo and A. Ruiz-Cortés. A Family of Experiments to Evaluate the Effects of Mindfulness on Software Engineering Students: The MetaMind Dataset. *Research and Evidence in Software Engineering: From Empirical Studies to Open Source Artifacts*. Taylor & Francis Publisher, 2020. [En prensa].

En este capítulo del libro sobre almacenes de datos, se describe el proceso de integración de los conjuntos de datos generados en una familia de experimentos. El conjunto de datos resultante se adaptó para llevar a cabo el meta-análisis en R. Posteriormente, para la ejecución del análisis complementario, se convirtió al tipo y formato de SPSS y se compararon los resultados.

8.2.3 Congresos nacionales

- [54] M. Cruz, B. Bernárdez, M. Resinas y A. Durán. Auditoría de procesos de negocio en la nube: persistencia mediante almacenes no relacionales. Actas de las XVII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2012). Almería, Septiembre 2012.

Se analiza el uso de sistemas de bases de datos no relacionales para dotar de persistencia a los sistemas PRaaS (Process as a Service) persiguiendo el principal objetivo de mejorar su escalabilidad.

- [52] M. Cruz, B. Bernárdez, A. del-Río-Ortega y A. Durán. Una revisión de la notación PPINOT para indicadores de rendimiento mediante su aplicación a un caso real. Actas de las X Jornadas de Ingeniería de Ciencia e Ingeniería de Servicios (JCIS 2014). Cádiz, Septiembre 2014.

Se presenta una revisión de los indicadores de rendimiento de proceso PPIs (Process Performance Indicators) mediante su aplicación al proceso de gestión de correo electrónico del SIC (Servicio de Informática y Comunicaciones) de la US (Universidad de Sevilla) especificado parte de su panel de indicadores.

- [53] M. Cruz, B. Bernárdez, A. Durán. Una Propuesta para Especificar Cambios de Replicaciones de Experimentos en Ingeniería del Software. Actas de las XXIII Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2018). Sevilla, Septiembre 2018.

En este trabajo se propone la plantilla y el metamodelo en su versión inicial y se valida en las dos replicaciones de la familia de experimentos de *Mindfulness* de IS que da origen a la propuesta.

8.3 TRABAJO FUTURO

En un futuro cercano, el trabajo continuará mejorando la herramienta no sólo a nivel de *usabilidad* sino también ampliando su funcionalidad en los siguientes términos:

- Proponer al experimentador, de forma automática, las posibles amenazas a la validez a las que puede dar lugar un cambio concreto. Se profundizará en la *inferencia automática de las amenazas a la validez* a partir de la dimensión afectada por el cambio.
- *Tipificar los elementos* que permiten clasificar los cambios correspondientes a la dimensión *Contexto*. Se analizará la forma de caracterizar el contexto agrupando los posibles cambios en categorías según afecten específicamente al sujeto, al medio donde se desarrolla el experimento o son consecuencia, entre otros, de realizar la replicación en un determinadas condiciones temporales, ambientales o sociales. Para ello será necesario analizar, previamente, otras propuestas para abordar el contexto en los experimentos y específicamente en las replicaciones.
- A partir del repositorio de información experimental, *generar el código Latex* que conforma la plantilla para su inclusión en un artículo o en el *paquete de laboratorio*.
- Integrar los *patrones-L* en la herramienta para facilitar al experimentador la redacción de los cambios.
- *Facilitar las consultas*. Aplicando la plantilla a una familia de experimentos, se proporciona una visión global de la familia de experimentos manteniendo el registro de todos los cambios llevados a cabo, de forma que posteriores consultas permitan conocer las variantes introducidas y proponer nuevos cambios facilitando así nuevas replicaciones.
- Poner la herramienta a disposición de los experimentadores del área de Ciencias que ya han utilizado y validado la plantilla. Conseguir así, un repositorio de información experimental en Ciencias y evaluar si son necesarios nuevos ajustes a la herramienta y por tanto a la plantilla y metamodelo subyacente, debido a las particularidades del área.

Más a largo plazo, nuestra intención es incentivar el uso de la plantilla en nuestra comunidad, para realizar una validación externa de la misma, obteniendo el feedback

de experimentadores expertos. Además, se piensa en ampliar la propuesta para la especificación resumida de la replicación completa y no solo de los cambios.

PARTE VI

ANEXOS

INSTANCIACIÓN DE LA VERSIÓN INICIAL DE LA PLANTILLA

En este anexo se muestra la aplicación de la versión inicial de la plantilla sobre las familias de experimentos Mindfulness, Análisis de requisitos y Técnicas evaluación de código pertenecientes al estudio Caso-IngSoft.

Tabla A.1: Replicación Mind#2

Campo	Valor
<i>Mind#2</i>	Replicación del experimento <i>Mind #1</i>
Tipo de replicación	Interna
Objetivo	Confirmar resultados
Cambio #1	Originalmente , durante 4 semanas se practicó mindfulness 4 días a la semana en sesiones de 10 minutos. En la replicación las sesiones fueron de 12 minutos y durante 6 semanas. Con el fin de hacer más evidentes los beneficios de Mindfulness.
Dimensión modificada Amenaza a la validez	Operacionalización , en concreto, la variable independiente <i>Taller de formación</i> El cambio incrementa la validez del constructo
Cambio #2	Originalmente , la asignación de sujetos a tratamiento no era aleatoria. En la replicación se hace aleatoria. Con el fin de subsanar las amenazas a la validez interna de los quasi experimentos.
Dimensión modificada Amenaza a la validez	Protocolo , en concreto, diseño experimental El cambio incrementa la validez interna
Cambio #3	Originalmente , se impartió un taller de oratoria al grupo de control a modo de placebo. En la replicación el taller de oratoria se llevó a cabo después del experimento. Con el fin de evitar el posible efecto de dicho taller.
Dimensión modificada Amenaza a la validez	Operacionalización , en concreto, la variable independiente <i>Taller de formación</i> El cambio incrementa la validez del constructo

Tabla A.2: Replicación Mind#3

Campo	Valor
<i>Mind#3</i>	Replicación del experimento <i>Mind #2</i>
Tipo de replicación	Interna
Objetivo	Confirmar resultados
Cambio #1	Originalmente , se realizó el problema sobre <i>Eramus</i> y tras el tratamiento, se realizó el problema <i>EoD Projects</i> . En la replicación se cambió el orden de los problemas. Con el fin de estudiar si influye en los resultados.
Dimensión modificada Amenaza a la validez	Protocolo , en concreto, las guías El cambio incrementa la validez interna

Tabla A.3: Replicación Q-2009

Campo	Valor
<i>Q-2009</i>	Replicación del experimento <i>Q-2007</i>
Tipo de replicación	Interna
Objetivo	Confirmar resultados
Cambio #1	Originalmente , se analiza la efectividad de los analistas en las sesiones de educación y la efectividad de los analistas en las sesiones de consolidación. En la replicación no se analiza la efectividad en educación. Debido a al alto costo de transcribir las entrevistas.
Dimensión modificada	Operacionalización , en concreto, la variable dependiente <i>efectividad en la educación</i>
Amenaza a la validez	El cambio *** (<i>Amenaza</i>)
Cambio #2	Originalmente , se analiza la capacidad de retención. En la replicación no se analiza. Debido a al alto costo de transcribir las entrevistas
Dimensión modificada	Operacionalización , en concreto, la variable dependiente <i>capacidad de retención</i>
Amenaza a la validez	El cambio *** (<i>Amenaza</i>)
Cambio #3	Originalmente , no considera experiencia en desarrollo. En la replicación se considera la experiencia en desarrollo para calcular la variable independiente experiencia. Debido a *** (<i>Razón</i>)
Dimensión modificada	Operacionalización , en concreto, la variable independiente <i>experiencia</i>
Amenaza a la validez	El cambio aumenta la validez del constructo
Cambio #4	Originalmente , las entrevistas se hacen en español. En la replicación las entrevistas se hacen en inglés. Debido a el inglés era un requisito del máster.
Dimensión modificada	Protocolo , en concreto, el material experimental
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #5	Originalmente , una persona responde en las entrevistas. En la replicación se cambia la persona encargada de responder las entrevistas. Debido a *** (<i>Razón</i>)
Dimensión modificada	Experimentador , en concreto, el monitor
Amenaza a la validez	El cambio *** (<i>Amenaza</i>)

Tabla A.4: Replicación Q-2011

Campo	Valor
<i>Q-2011</i>	Replicación del experimento <i>Q-2009</i>
Tipo de replicación	Interna
Objetivo	Generalizar resultados
Cambio #1	Originalmente , las entrevistas son individuales. En la replicación las entrevistas son en grupo. Debido al costo de realizar entrevistas individuales y la influencia del cansancio del experimentador en la efectividad de los sujetos.
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #2	Originalmente , la experiencia se calcula a partir de los años de experiencia. En la replicación la experiencia se operacionaliza además en función de las habilidades que el propio sujeto cree tener. Debido a ***(Razón)
Dimensión modificada	Operacionalización , en concreto, las variables independientes <i>Habilidad en requisitos y Habilidad en entrevistas</i>
Amenaza a la validez	El cambio ***(Amenazas) ***(Amenazas Varias)
Cambio #3	Originalmente , el tiempo de educación, es decir la duración de las entrevistas es de 30min. En la replicación el tiempo de educación es de 60min. Debido a que la entrevista es en grupo
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #4	Originalmente , el tiempo transcurrido entre la sesión de educación y la consolidación es decir, cuando el sujeto presenta la información recopilada es de 7 días. En la replicación la consolidación es inmediata a la educación. Con el fin de evitar que se olvide la información
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #5	Originalmente , el tiempo de consolidación, es decir, el tiempo disponible para que el analista presente por escrito la información adquirida en las sesiones de educación no se mide. En la replicación el tiempo de consolidación es de 120min. Debido a que la consolidación es inmediata a educación
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #6	Originalmente , una persona responde en las entrevistas. En la replicación se cambia la persona encargada de responder las entrevistas. Debido a ***(Razón)
Dimensión modificada	Experimentador , en concreto, el monitor
Amenaza a la validez	El cambio ***(Amenaza)

Tabla A.5: Replicación Q-2012

Campo	Valor
<i>Q-2012</i>	Replicación del experimento <i>Q-2011</i>
Tipo de replicación	Externa
Objetivo	Confirmar resultados
Cambio #1	Originalmente , los sujetos son estudiantes de máster. En la replicación los sujetos son profesionales. Debido a que se realiza en el 'International Working Conference on Requirements Engineering'.
Dimensión modificada	Población , en concreto, los sujetos experimentales
Amenaza a la validez	El cambio aumenta la validez externa
Cambio #2	Originalmente , los sujetos no tienen o tienen poca experiencia en desarrollo. En la replicación se analiza la influencia de la <i>habilidad en desarrollo</i> ; los sujetos son profesionales con experiencia en desarrollo. Debido a que se realiza en el 'International Working Conference on Requirements Engineering'.
Dimensión modificada	Operacionalización , en concreto, la variable independiente <i>habilidad en desarrollo</i>
Amenaza a la validez	El cambio aumenta la validez del constructo
Cambio #3	Originalmente , el tiempo de consolidación es de 120min. En la replicación el tiempo de consolidación es de 30min. Debido a restricciones de tiempo.
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio aumenta la validez interna
Cambio #4	Originalmente , el experimento se realiza al final del curso, es decir, del periodo de formación. En la replicación no hay periodo de formación. Debido a que se realiza en el 'International Working Conference on Requirements Engineering'.
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio aumenta la validez interna

Tabla A.6: Replicación E-2012A

Campo	Valor
<i>E-2012A</i>	Replicación del experimento <i>Q-2012</i>
Tipo de replicación	Interna
Objetivo	Generalizar resultados
Cambio #1	Originalmente , el conocimiento se operacionaliza como familiaridad mediante una valoración subjetiva. En la replicación el conocimiento se operacionaliza como variable independiente con dos niveles: problema conocido y desconocido. Debido a que en la población experimental (estudiantes de post-grado) se puede saber si conocen un determinado dominio del problema.
Dimensión modificada	Operacionalización , en concreto, la variable independiente <i>conocimiento</i>
Amenaza a la validez	El cambio *** (<i>Amenaza</i>)
Cambio #2	Originalmente , es un cuasi experimento y la educación de requisitos se realiza en dos días, para evitar el cansancio del experimentador. En la replicación es un diseño de medidas repetidas. Debido a problemas de poder estadístico ya que con este diseño no se necesita un número elevado de sujetos.
Dimensión modificada	Protocolo , en concreto, el diseño experimental
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #3	Originalmente , las entrevistas son en grupo. En la replicación las entrevistas son individuales. Debido a que hay dos analistas e idiomas.
Dimensión modificada	Protocolo , las guías
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #4	Originalmente , no hay variables de bloqueo. En la replicación hay variable de bloqueo por idioma. Debido a que es esperable que los sujetos que utilicen su lengua materna sean más efectivos que los sujetos que utilizan una segunda lengua.
Dimensión modificada	Protocolo , en concreto, diseño experimental
Amenaza a la validez	El cambio incrementa la validez interna

Cambio #5	<p>Originalmente, no hay variables de bloqueo.</p> <p>En la replicación hay variable de bloqueo por analista.</p> <p>Debido a al bloquear por analista, se evitan las interacciones con la variable idioma. Los sujetos realizan la educación en lengua materna.</p> <p>Protocolo, el diseño experimental</p> <p>El cambio incrementa la validez interna</p>
Dimensión modificada Amenaza a la validez	
Cambio #6	<p>Originalmente, hay un analista.</p> <p>En la replicación hay dos analistas.</p> <p>Con el fin de paliar cansancio y aprendizaje del analista.</p> <p>Protocolo, las guías</p> <p>El cambio incrementa la validez interna</p>
Dimensión modificada Amenaza a la validez	
Cambio #7	<p>Originalmente, hay un mismo problema (objeto experimental) para todos los sujetos.</p> <p>En la replicación hay dos problemas.</p> <p>Debido a que se hacen grupos por las variables bloqueo.</p> <p>Protocolo, en concreto, diseño experimental</p> <p>El cambio incrementa la validez interna</p>
Dimensión modificada Amenaza a la validez	
Cambio #8	<p>Originalmente, el tiempo de educación, es decir la duración de las entrevistas es de 60min.</p> <p>En la replicación el tiempo de educación es de 30min.</p> <p>Debido a que la entrevista es individual.</p> <p>Protocolo, en concreto, las guías</p> <p>El cambio incrementa la validez interna</p>
Dimensión modificada Amenaza a la validez	
Cambio #9	<p>Originalmente, el tiempo de consolidación es de 30min.</p> <p>En la replicación el tiempo de consolidación es 90min.</p> <p>Debido a que la duración recomendada de 90 minutos.</p> <p>Protocolo, en concreto, las guías</p> <p>El cambio incrementa la validez interna</p>
Dimensión modificada Amenaza a la validez	
Cambio #10	<p>Originalmente, no se mide la dificultad del problema.</p> <p>En la replicación la variable dificultad indica la dificultad del problema.</p> <p>Debido a que hay dos problemas.</p> <p>Operacionalización, en concreto, la variable independiente <i>dificultad</i></p> <p>El cambio incrementa la validez del constructo</p>
Dimensión modificada Amenaza a la validez	

Tabla A.7: Replicación E-2012B

Campo	Valor
<i>E-2012B</i>	Replicación del experimento <i>E-2012A</i>
Tipo de replicación	Interna
Objetivo	Confirmar resultados
Cambio #1	Originalmente , Se utilizan dos dominios de problemas en el experimento, uno de dominio conocido (DC) y el otro de dominio desconocido (DD). En la replicación Se han modificado los dominios de problemas utilizados en el experimento, pero uno sigue siendo dominio conocido (DC) y el otro dominio desconocido (DD). Debido a *** (<i>Razón</i>).
Dimensión modificada	Protocolo , en concreto, material experimental
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #2	Originalmente , Primero se realiza el problema de dominio conocido y luego el de dominio desconocido. En la replicación Se permuta el orden de realización de los problemas, primero el problema de dominio desconocido y luego el conocido. Debido a *** (<i>Razón</i>).
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #3	Originalmente , el experimento se ha llevado a cabo al principio del curso. En la replicación se ha llevado a cabo después de que los sujetos hayan recibido formación en Ingeniería de Requisitos y específicamente en educación. Debido a *** (<i>Razón</i>).
Dimensión modificada	*** (<i>Contexto</i>).
Amenaza a la validez	El cambio incrementa la validez interna

Tabla A.8: Replicación E-2013

Campo	Valor
<i>E-2013</i>	Replicación del experimento <i>E-2012B</i>
Tipo de replicación	Interna
Objetivo	Generalizar resultados
Cambio #1	Originalmente , el diseño es de medidas repetidas. En la replicación el diseño es entre sujetos (between-subjects). Debido a que el diseño de medidas repetidas da lugar a un sesgo derivado del orden en el que se administran los tratamientos a los sujetos. Esto es, por ejemplo el efecto aprendizaje. En el diseño entre sujetos, cada sujeto sólo se somete a un tratamiento.
Dimensión modificada	Protocolo , en concreto, diseño experimental
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #2	Originalmente , no hay formación breve en actividades relacionadas con requisitos previa a la propiamente dicha del curso. En la replicación la formación breve es de una semana. Debido a *** (<i>Razón</i>).
Dimensión modificada	Operacionalización , *** (<i>Nombre</i>).
Amenaza a la validez	El cambio incrementa la validez del constructo

Tabla A.9: Replicación E-2014

Campo	Valor
<i>E-2014</i>	Replicación del experimento <i>E-2013</i>
Tipo de replicación	Interna
Objetivo	Generalizar resultados
Cambio #1	Originalmente , hay dos analistas. En la replicación hay sólo un analista. Debido a la indisponibilidad de unos de los analistas.
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza a la validez	El cambio incrementa la validez interna
Cambio #2	Originalmente , la formación breve en actividades relacionadas con requisitos previa es de 1 semana. En la replicación la formación breve es de 6 semanas. Debido a que se quiere estudiar el efecto de la formación
Dimensión modificada	Operacionalización , *** (<i>Nombre</i>).
Amenaza a la validez	El cambio incrementa la validez del constructo

Tabla A.10: Replicación E-2015

Campo	Valor
<i>E-2015</i>	Replicación del experimento <i>E-2013</i>
Tipo de replicación	Interna
Objetivo	Generalizar resultados
Cambio #1	Originalmente , la formación breve en actividades relacionadas con requisitos es de 1 semana. En la replicación la formación breve es de 2 semanas. Debido a que se quiere estudiar el efecto de la formación
Dimensión modificada	Operacionalización , *** (<i>Nombre</i>).
Amenaza a la validez	El cambio incrementa la validez del constructo

Tabla A.11: Replicación VV-UPM1

Campo	Valor
VV-UPM1	Replicación del experimento <i>técnicas de verificación y validación (VV-UPM)</i>
Tipo de replicación	Interna
Objetivo	Generalizar resultados
Cambio #1	Originalmente , no se analiza la visibilidad del fallo. En la replicación se analiza la visibilidad del fallo. Debido a querer obtener más conclusiones.
Dimensión modificada	Operacionalización , ***(Nombre)
Amenaza a la validez	El cambio ***(Amenazas) ***(Amenazas Varias)
Comentario	Se utiliza el paquete de laboratorio de Kamsties y Lott
Cambio #2	Originalmente , el programa es un factor (variable independiente) aunque no se estudia su influencia. En la replicación se implementan dos versiones de cada programa que será un nuevo factor. Debido a querer obtener más conclusiones.
Dimensión modificada	Operacionalización , específicamente, la variable independiente <i>versión</i>
Amenaza a la validez	El cambio aumenta la validez del constructo
Cambio #3	Originalmente , tres de los tipos de fallos aparecen una vez mientras que los otros tres tipos aparecen dos veces. En la replicación se duplican todos los tipos de fallos. Debido a que hay dos versiones de cada programa.
Dimensión modificada	Operacionalización , específicamente, la variable dependiente <i>errores</i>
Amenaza abordada	El cambio incrementa la validez del Constructo ya que se refleja menor
Cambio #4	Originalmente , los sujetos generan sus casos de prueba para detectar los fallos del código. En la replicación en primer lugar, los sujetos aplican la técnica para generar los casos de prueba y posteriormente, ejecutarán los casos de prueba que se les proporcionan para detectar los fallos del programa. Con el fin de comprobar si la visibilidad de los fallos influye en su detección.
Dimensión modificada	Protocolo , específicamente, material experimental
Amenaza abordada	El cambio incrementa la validez Interna
Cambio #5	Originalmente , se utilizaron cuatro programas. En la replicación se utilizan tres programas, uno es descartado. Con el fin de equilibrar el diseño.
Dimensión modificada	Protocolo , específicamente, diseño experimental
Amenaza abordada	El cambio incrementa la validez Interna
Cambio #6	Originalmente , cada sujeto aplica una técnica. En la replicación cada sujeto aplica las tres técnicas. Debido a que se cambia el diseño.
Dimensión modificada	Protocolo , específicamente, diseño experimental
Amenaza abordada	El cambio incrementa la validez Interna

Tabla A.12: Replicación VV-UPV

Campo	Valor
VV-UPV	Replicación del experimento <i>técnicas de verificación y validación (VV-UPM)</i>
Tipo de replicación	Externa
Objetivo	Generalizar resultados
Cambio #1	Originalmente , se utilizan las tres técnicas de verificación y validación: lectura de código, partición de equivalencia y branch testing. En la replicación se omite la técnica lectura de código. Debido a restricciones de tiempo.
Dimensión modificada	Operacionalización , la variable independiente <i>técnica</i>
Amenaza a la validez	El cambio aumenta la validez del constructo
Cambio #2	Originalmente , la duración de las tres sesiones es de cuatro horas cada una; es decir el tiempo es ilimitado. En la replicación la duración de cada una de las tres sesiones es de dos horas. Debido a restricciones de tiempo.
Dimensión modificada	Protocolo , específicamente, las guías
Amenaza a la validez	El cambio aumenta la validez interna
Cambio #3	Originalmente , los sujetos reciben tres sesiones de entrenamiento de cuatro horas para aprender a aplicar las técnicas. En la replicación la formación consiste en dos breves tutoriales de dos horas. Debido a que los sujetos ya están familiarizados con las técnicas.
Dimensión modificada	Protocolo , específicamente las guías
Amenaza a la validez	El cambio no afecta a la validez
Comentario	La formación no se operacionaliza.
Cambio #4	Originalmente , el entrenamiento en el uso de las técnicas es antes de que se ejecute el experimento. En la replicación Cada tutorial se lleva a cabo antes de la aplicación de la técnica, en las dos primeras sesiones; es decir, el entrenamiento es intercalado con la operación del experimento. Debido a que los sujetos ya están familiarizados con las técnicas.
Dimensión modificada	Protocolo , específicamente las guías
Amenaza a la validez	El cambio no afecta a la validez
Comentario	La formación no se operacionaliza.
Cambio #5	Originalmente , los sujetos aplican una técnica a un programa en cada sesión. En la replicación aplican la misma técnica a diferentes programas en cada sesión. Debido a restricciones de tiempo.
Dimensión modificada	Protocolo , específicamente, el diseño experimental
Amenaza a la validez	El cambio aumenta la validez interna
Cambio #6	Originalmente , Los sujetos ejecutan casos de prueba con la aplicación de la técnica; es decir en cada sesión. En la replicación Los sujetos ejecutan casos de prueba para uno de los programas que han probado en sesión aparte; es decir en la sesión 3. Debido a restricciones de tiempo.
Dimensión modificada	Protocolo , específicamente, las guías
Amenaza a la validez	El cambio aumenta la validez interna

Tabla A.13: Replicación VV-Uds

Campo	Valor
VV-Uds	Replicación del experimento <i>técnicas de verificación y validación (VV-UPM)</i>
Tipo de replicación	Externa
Objetivo	Generalizar resultados
Cambio #1	Originalmente , la duración de las tres sesiones es de cuatro horas cada una; es decir el tiempo es ilimitado. En la replicación la duración de cada sesión es 2 horas. Debido a restricciones de tiempo.
Dimensión modificada	Protocolo , en concreto, las guías
Amenazas a la validez	El cambio disminuye la validez interna ya que la restricción de tiempo puede afectar a los resultados
Cambio #2	Originalmente , los sujetos ejecutan casos de prueba con la aplicación de la técnica; es decir en cada sesión. En la replicación los casos se ejecutan en sesión posterior; es decir en la sesión cuatro. Debido a restricciones de tiempo.
Dimensión modificada	Protocolo , en concreto, las guías
Amenazas a la validez	El cambio disminuye la validez interna ya que la restricción de tiempo puede afectar a los resultados
Cambio #3	Originalmente , los sujetos trabajan de forma individual. En la replicación los sujetos trabajan en parejas. Debido a no hay suficientes ordenadores.
Dimensión modificada	Protocolo , en concreto, las guías
Amenaza abordada	El cambio incrementa la validez Interna
Cambio #4	Originalmente , los sujetos reciben tres sesiones de entrenamiento de 4h. para aprender a aplicar las técnicas. En la replicación la formación son dos tutoriales de 2h. Debido a ya están familiarizados con las técnicas.
Dimensión modificada	Protocolo , específicamente las guías
Amenazas a la validez	El cambio no afecta a la validez ya que los los sujetos ya están familiarizados con las técnicas y la formación no se operacionaliza
Cambio #5	Originalmente , el entrenamiento en el uso de las técnicas es antes de que se ejecute el experimento. En la replicación cada tutorial se lleva a cabo antes de la aplicación de la técnica, intercalado con el experimento. Debido a ya están familiarizados con las técnicas.
Dimensión modificada	Protocolo , en concreto, las guías
Amenazas a la validez	El cambio no afecta a la validez ya que los los sujetos ya están familiarizados con las técnicas y la formación no se operacionaliza

Tabla A.14: Replicación VV-ORT

Campo	Valor
VV-ORT	Replicación del experimento <i>técnicas de verificación y validación (VV-UIPM)</i>
Tipo de replicación	Externa
Objetivo	Generalizar resultados
Cambio #1	Originalmente , se utilizan las tres técnicas de verificación y validación: lectura de códigos (code reading), partición de equivalencia (equivalence partitioning) y prueba de rama (branch testing). En la replicación se omite la técnica de lectura de códigos (code reading) . Debido a restricciones de tiempo.
Dimensión modificada	Operacionalización , específicamente, la variable independiente <i>técnica</i>
Amenaza abordada	El cambio ***(Amenazas)
Cambio #2	Originalmente , se utilizan tres programas. En la replicación se omite uno de los programas. Debido a restricciones de tiempo.
Dimensión modificada	Operacionalización , específicamente, diseño experimental
Amenaza abordada	El cambio ***(Amenazas)
Cambio #3	Originalmente , el experimento se ejecuta en tres sesiones de cuatro horas cada una; es decir tiempo ilimitado. En la replicación el experimento se ejecuta en una única sesión. Debido a restricciones de tiempo.
Dimensión modificada	Protocolo , específicamente, las guías
Amenaza abordada	El cambio incrementa la validez interna
Cambio #4	Originalmente , los sujetos aplican una técnica a un programa en cada una de las tres sesiones. En la replicación los sujetos aplican las dos técnicas a los dos programas en una única sesión. Debido a que la duración de la sesión es ilimitada.
Dimensión modificada	Protocolo , específicamente, diseño experimental
Amenaza abordada	El cambio incrementa la validez interna

INSTANCIACIÓN DE LA VERSIÓN FINAL DE LA PLANTILLA

En este anexo se muestra la aplicación de la versión final de la plantilla sobre las familias de experimentos del estudio de caso múltiple compuesto por los estudios Caso-IngSoft, Caso-Ciencias y Caso-Automáticos.

Tabla B.1: Replicación Mind#2, versión final

Replicación	<i>Mind #2</i> replicación Interna basada en el experimento base <i>Mind #1</i>
Objetivo del experimento	Estudiar si la práctica de mindfulness (causa) mejora la productividad en modelado conceptual (efecto) en estudiantes de Ingeniería del Software (población).
Descripción del experimento	Un grupo de estudiantes del Grado de Ingeniería de Software de la Universidad de Sevilla (muestra) asistió a sesiones de mindfulness de 10 minutos durante 4 semanas, 4 días por semana (tratamiento grupo experimental), mientras que un segundo grupo de estudiantes asistió a un taller de oratoria a modo de placebo (tratamiento grupo de control). Los estudiantes realizaron dos ejercicios de modelado conceptual, uno antes y otro después del tratamiento (diseño experimental), que se puntuaron usando una solución de referencia (procedimiento de medida). Se comparó el rendimiento de ambos grupos en términos de calidad (similitud con la solución de referencia) y productividad (similitud en porcentaje por unidad de tiempo) (métricas).
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en el primer semestre del curso 2013–2014 y esta replicación en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> , en el primer semestre del curso 2014–2015
Propósito	Confirmar resultados Superar limitaciones del experimento base
Cambio #1	Aumento de la duración del tratamiento Originalmente , Durante 4 semanas se practicaba Mindfulness 4 días a la semana en sesiones de 10 minutos En la replicación las sesiones duraron 12 minutos y durante 6 semanas Con el fin de hacer más evidentes los beneficios de Mindfulness.
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente, la causa El cambio incrementa la validez del constructo ya que incrementar la duración del tratamiento refleja mejor el efecto de la práctica de Mindfulness. El cambio incrementa la validez interna ya que incrementar la duración del tratamiento refuerza su efecto sobre el de otros posibles factores.
Cambio #2	Asignación aleatoria de sujetos a grupos Originalmente , la asignación de los sujetos al tratamiento no fue aleatoria En la replicación es aleatorio Con el fin de mitigar la amenaza de sesgo de selección y asignación y evitar las limitaciones del análisis estadístico
Dimensión Modificada Amenazas a la validez	Protocolo , específicamente, el diseño experimental. El cambio incrementa la validez de conclusión, ya que mejora la potencia de los tests estadísticos aplicables. El cambio incrementa la validez interna, ya que elimina el sesgo de asignación.
Cambio #3	Tratamiento nulo del grupo de control Originalmente , se impartió un taller de oratoria al grupo de control como placebo En la replicación El grupo de control recibió un tratamiento nulo, ya que el taller de oratoria se realizó una vez terminado el experimento Con el fin de mitigar el potencial factor perturbador del placebo en los resultados del experimento.
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente, la causa El cambio incrementa la validez interna ya que elimina la posibilidad de que el placebo pudiera tener algún efecto sobre los resultados.

Tabla B.2: Replicación Mind#3, versión final

Replicación	<i>Mind #3</i> replicación Interna basada en la replicación previa <i>Mind #2</i>
Descripción del experimento	El estudio analiza el efecto de la práctica de Mindfulness en el rendimiento de los estudiantes a la hora de desarrollar modelos conceptuales
Lugar y Fecha	La replicación previa se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en el primer semestre del curso 2014–2015 y esta replicación en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> , en el primer semestre del curso 2015–2016
Propósito	Confirmar resultados
Cambio #1	<p>Cambio en el orden de los ejercicios de modelado conceptual</p> <p>Originalmente, se realizó el problema sobre <i>Eramus</i> y tras el tratamiento, se realizó el problema <i>EoD Projects</i>.</p> <p>En la replicación se cambió el orden de los problemas.</p> <p>Con el fin de estudiar si influye en los resultados.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	<p>El cambio aumenta la validez interna ya que permite estudiar el efecto de la diferencia entre tareas en los resultados experimentales</p> <p>El cambio aumenta la validez de conclusión ya que permite analizar y ajustar el efecto de la diferencia entre tareas en los resultados experimentales</p>

Tabla B.3: Replicación Q-2009, versión final

Replicación	Q-2009 replicación Interna basada en experimento original Q-2007
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Descripción del experimento	Se analiza la efectividad de los analistas, empleando la entrevista como técnica de educación. En las entrevistas, el experimentador actúa como cliente (entrevistado) respondiendo a las preguntas de los sujetos experimentales (analistas) sobre dos posibles problemas, uno de dominio conocido y otro de dominio desconocido. Durante la educación, cada sujeto tiene que identificar la información relevante del dominio del problema (efectividad de educación). Para medir la efectividad de consolidación, pasado un tiempo, el analista presenta por escrito lo que recuerda de la entrevista y se contabiliza la cantidad de elementos del problema mencionados por el analista.
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2007 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2009
Propósito	Confirmar resultados
Cambio #1	Análisis de la efectividad Originalmente , se analiza la efectividad de los analistas en las sesiones de educación (entrevistas) y la efectividad de los analistas en las sesiones de consolidación (pasado un tiempo) En la replicación no se analiza la efectividad en educación. Causado por el alto costo de transcribir las entrevistas.
Dimensión modificada Amenaza a la validez	Operacionalización , específicamente, el efecto El cambio disminuye la validez del constructo ya que no se considera la variable dependiente efectividad de educación
Cambio #2	Análisis de la capacidad de retención Originalmente , se analiza la capacidad de retención de los analistas (sujetos) En la replicación no se analiza. Causado por al alto costo de transcribir las entrevistas
Dimensión modificada Amenaza a la validez	Operacionalización , específicamente, el efecto El cambio disminuye la validez del constructo ya que no se analiza la capacidad de retención
Cambio #3	Considerar la experiencia en desarrollo Originalmente , no considera experiencia en desarrollo. En la replicación se considera la experiencia en desarrollo para calcular la variable independiente experiencia. El motivo no se especifica
Dimensión modificada Amenaza a la validez	Operacionalización , específicamente, la causa El cambio aumenta la validez del constructo ya que hay una variable más
Cambio #4	Idioma de las entrevistas Originalmente , las entrevistas se hacen en español. En la replicación las entrevistas se hacen en inglés. ya que el inglés era un requisito del máster.
Dimensión modificada Amenaza a la validez	Protocolo , específicamente, el material experimental El cambio incrementa la validez interna ya que los resultados son independientes del idioma en que se lleva a cabo la entrevista
Cambio #5	No disponibilidad de la persona que responde en las entrevistas Originalmente , una persona responde en las entrevistas. En la replicación se cambia la persona encargada de responder las entrevistas. ya que la persona que respondía en el experimento original no está disponible
Dimensión modificada Amenaza a la validez	Experimentador , específicamente, el monitor El cambio aumenta la validez interna ya que elimina el sesgo del experimentador

Tabla B.4: Replicación Q-2011, versión final

Replicación	Q-2011 replicación Interna basada en experimento original Q-2009
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2009 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2011
Propósito	Generalizar resultados
Cambio #1	<p>Entrevistas en grupo Originalmente, las entrevistas son individuales. En la replicación las entrevistas son en grupo. Debido al costo de realizar entrevistas individuales y la influencia del cansancio del experimentador en la efectividad de los sujetos.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio incrementa la validez interna ya que todos los sujetos reciben la misma información.
Cambio #2	<p>Determinación de la experiencia Originalmente, la experiencia se calcula a partir de los años de experiencia. En la replicación la experiencia se operacionaliza además en función de las habilidades que el propio sujeto cree tener, en concreto las variables independientes <i>Habilidad en requisitos</i> y <i>Habilidad en entrevistas</i> .</p>
Dimensión modificada	Operacionalización , específicamente, la causa
Amenazas a la validez	El cambio aumenta la validez del constructo ya que se tienen en cuenta más variables El cambio disminuye la validez de conclusión ya que el procedimiento se hace tedioso.
Cambio #3	<p>Duración de las entrevistas Originalmente, el tiempo de educación, es decir la duración de las entrevistas es de 30min. En la replicación el tiempo de educación es de 60min. Debido a que la entrevista es en grupo</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio incrementa la validez interna ya que aumenta la duración de las entrevistas para comprender mejor los requisitos
Cambio #4	<p>Tiempo transcurrido hasta la presentación de la información Originalmente, el tiempo transcurrido entre la sesión de educación y la consolidación es decir, cuando el sujeto presenta la información recopilada es de 7 días. En la replicación la consolidación es inmediata a la educación. Con el fin de evitar que se olvide la información</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio incrementa la validez interna ya que se recoge la información tras celebrar la entrevista evitando que se olvide parte de la información
Cambio #5	<p>Tiempo para presentación de información Originalmente, el tiempo de consolidación, es decir, el tiempo disponible para que el analista presente por escrito la información adquirida en las sesiones de educación no se mide. En la replicación el tiempo de consolidación es de 120min. Debido a que la consolidación es inmediata a educación</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio incrementa la validez interna se analizar si el tiempo influye en la información recopilada
Cambio #6	<p>No disponibilidad de la persona que responde en las entrevistas Originalmente, una persona responde en las entrevistas. En la replicación se cambia la persona encargada de responder las entrevistas. ya que la persona que respondía en el experimento original no está disponible</p>
Dimensión modificada	Experimentador , específicamente, el monitor
Amenazas a la validez	El cambio aumenta la validez interna ya que elimina el sesgo del experimentador

Tabla B.5: Replicación Q-2012, versión final

Replicación	Q-2012 replicación Interna basada en experimento original Q-2011
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2011 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2012
Propósito	Confirmar resultados
Cambio #1	<p>Sujetos profesionales Originalmente, los sujetos son estudiantes de máster. En la replicación los sujetos son profesionales. ya que se realiza en el 'International Working Conference on Requirements Engineering'.</p>
Dimensión modificada	Población , específicamente, experiencia
Amenazas a la validez	El cambio aumenta la validez externa ya que se analiza el efecto en profesionales.
Cambio #2	<p>Influencia de la habilidad en desarrollo Originalmente, los sujetos no tienen o tienen poca experiencia en desarrollo. En la replicación se analiza la influencia de la <i>habilidad en desarrollo</i> ya que se realiza en el 'International Working Conference on Requirements Engineering' y los sujetos son profesionales con experiencia en desarrollo</p>
Dimensión modificada	Operacionalización , específicamente, la causa
Amenazas a la validez	El cambio aumenta la validez del constructo ya que se tiene en cuenta la variable <i>habilidad en desarrollo</i> El cambio aumenta la validez externa ya que se amplía la población.
Cambio #3	<p>Disminución del tiempo de consolidación Originalmente, el tiempo de consolidación (cuando se presenta la información recopilada) es de 120min. En la replicación el tiempo de consolidación es de 30min. Causado por restricciones de tiempo.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio aumenta la validez interna al permitir analizar la posible influencia del tiempo.
Cambio #4	<p>Supresión del período de formación Originalmente, el experimento se realiza al final del curso, es decir, del periodo de formación. En la replicación no hay periodo de formación. ya que se realiza en el 'International Working Conference on Requirements Engineering'.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio aumenta la validez interna ya que se analiza la influencia de la formación y/o experiencia de los profesionales.

Tabla B.6: Replicación E-2012A, versión final

Replicación	<i>E-2012A</i> replicación Interna basada en experimento original <i>Q-2012</i>
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2012 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2012
Propósito	Generalizar resultados
Cambio #1	<p>Conocimiento del dominio del problema Originalmente, el conocimiento se operacionaliza como familiaridad mediante una valoración subjetiva. En la replicación el conocimiento se operacionaliza como variable independiente con dos niveles: problema conocido y desconocido. Debido a que en la población experimental (estudiantes de post-grado) se puede saber si conocen un determinado dominio del problema.</p>
Dimensión modificada Amenaza a la validez	<p>Operacionalización, la causa El cambio aumenta la validez del constructo ya que se añade una variable que mejora la operacionalización</p>
Cambio #2	<p>Diseño de medidas repetidas Originalmente, el diseño del experimento consiste en entrevistas en dos días y tras 7 días se cumplimenta el informe En la replicación es un diseño de medidas repetidas. Ya que el número de sujetos es bajo y con este diseño no se necesita un número elevado de sujetos.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, el diseño experimental El cambio aumenta la validez de conclusión debido a los test aplicables</p>
Cambio #3	<p>Tipo de entrevista Originalmente, las entrevistas son en grupo. En la replicación las entrevistas son individuales. Debido a que hay dos monitores (responden a las entrevistas) y dos idiomas.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, las guías El cambio aumenta la validez interna ya que permite analizar las diferencias entre ambos monitores referente a la información suministrada El cambio disminuye la validez de conclusión ya que, aunque hay dos monitores respondiendo, al aumentar el número de entrevistas por ser individuales, el proceso se puede volver tedioso para dichos monitores</p>
Cambio #4	<p>Bloqueo por idioma Originalmente, no hay variables de bloqueo. En la replicación hay variable de bloqueo por idioma. Ya que se espera que los sujetos que utilicen su lengua materna sean más efectivos que los sujetos que utilizan una segunda lengua.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, diseño experimental El cambio aumenta la validez interna ya que se bloquea el idioma para que no influya en los resultados</p>
Cambio #5	<p>Bloqueo por monitor que responde a las entrevistas Originalmente, no hay variables de bloqueo. En la replicación hay variable de bloqueo por monitor. Ya que se espera que los sujetos que utilicen su lengua materna sean más efectivos que los sujetos que utilizan una segunda lengua.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, el diseño experimental El cambio aumenta la validez interna ya que se bloquea el monitor para que no influya en los resultados</p>

Cambio #6	<p>Número de monitores respondiendo a las entrevistas Originalmente, hay un monitor que responde a las entrevistas. En la replicación hay dos monitores respondiendo. Con el fin de paliar el cansancio del monitor.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, las guías El cambio aumenta la validez interna ya que se evita el cansancio y aprendizaje del monitor que responde en las entrevistas</p>
Cambio #7	<p>Número de problemas Originalmente, hay un mismo problema para todos los sujetos. En la replicación hay dos problemas. Debido a que se hacen grupos por las variables bloqueo.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, diseño experimental El cambio aumenta la validez interna se pueden analizar las diferencias</p>
Cambio #8	<p>Duración de las entrevistas Originalmente, el tiempo de educación, es decir la duración de las entrevistas es de 60min. En la replicación el tiempo de educación es de 30min. Debido a que la entrevista es individual.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, las guías El cambio aumenta la validez interna ya que disminuye el cansancio del monitor</p>
Cambio #9	<p>Tiempo de consolidación Originalmente, el tiempo de consolidación es de 30min. En la replicación el tiempo de consolidación es 90min.</p>
Dimensión modificada Amenaza a la validez	<p>Protocolo, específicamente, las guías El cambio aumenta la validez interna ya que al disponer de más tiempo, la información recogida puede ser más exacta</p>
Cambio #10	<p>Dificultad del problema Originalmente, no se mide la dificultad del problema. En la replicación se mide mediante la variable dificultad. ya que al haber dos problemas, se analiza si la dificultad influye en los resultados.</p>
Dimensión modificada Amenaza a la validez	<p>Operacionalización, específicamente, la causa El cambio aumenta la validez del constructo ya que hay una nueva variable para recoger mejor la operacionalización</p>

Tabla B.7: Replicación E-2012B, versión final

Replicación	<i>E-2012A</i> replicación Interna basada en experimento original <i>E-2012B</i>
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2012 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2012
Propósito	Confirmar resultados
Cambio #1	<p>Nuevos problemas</p> <p>Originalmente, se utilizan dos dominios de problemas en el experimento, uno de dominio conocido (DC) y el otro de dominio desconocido (DD).</p> <p>En la replicación se han modificado los dominios de problemas aunque uno sigue siendo de dominio conocido (DC) y el otro de dominio desconocido (DD).</p>
Dimensión modificada	Protocolo , específicamente, material experimental
Amenaza a la validez	El cambio no afecta a la validez
Cambio #2	<p>Orden de los problemas</p> <p>Originalmente, primero se realiza el problema de dominio conocido y luego el de dominio desconocido.</p> <p>En la replicación se permuta el orden de realización de los problemas, primero el problema de dominio desconocido y luego el conocido.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenaza a la validez	<p>El cambio aumenta la validez interna ya que permite analizar si el orden afecta a los resultados</p> <p>El cambio aumenta la validez de conclusión ya que permite analizar y ajustar el efecto de la diferencia entre tareas en los resultados experimentales</p>
Cambio #3	<p>Momento de realización del experimento</p> <p>Originalmente, el experimento se ha llevado a cabo al principio del curso.</p> <p>En la replicación, se lleva a cabo al final del curso.</p>
Dimensión modificada	Contexto , específicamente el Momento de realización
Amenaza a la validez	El cambio disminuye la validez interna ya que la replicación se lleva a cabo a final de curso y puede influir en los resultados.

Tabla B.8: Replicación E-2013, versión final

Replicación	<i>E-2013</i> replicación Interna basada en experimento original <i>E-2012B</i>
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2012 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2013
Propósito	Generalizar resultados
Cambio #1	<p>Diseño entre sujetos</p> <p>Originalmente, el diseños es de medidas repetidas.</p> <p>En la replicación el diseño es entre sujetos</p> <p>Debido a que el diseño de medidas repetidas produce un sesgo derivado del orden en el que se administran los tratamientos a los sujetos.</p>
Dimensión modificada	Protocolo , específicamente, diseño experimental
Amenaza a la validez	El cambio incrementa la validez interna al evitar el sesgo debido al orden en el que se administran los tratamientos
Comentario	En el diseño entre sujetos, cada sujeto sólo se somete a un tratamiento para evitar el efecto aprendizaje
Cambio #2	<p>Duración de la formación previa</p> <p>Originalmente, no hay formación previa al experimento.</p> <p>En la replicación hay formación previa sobre requisito, durante una semana.</p> <p>Debido a que se quiere estudiar el efecto de la formación</p>
Dimensión modificada	Operacionalización , específicamente la causa
Amenaza a la validez	El cambio incrementa la validez del constructo ya que se refleja mejor el efecto de la formación El cambio incrementa la validez interna ya que se analiza el efecto de la formación frente a otros factores

Tabla B.9: Replicación E-2014, versión final

Replicación	<i>E-2014</i> replicación Interna basada en experimento original <i>E-2013</i>
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2013 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2014
Propósito	Generalizar resultados
Cambio #1	Un solo monitor para responder en las entrevistas Originalmente , hay dos monitores. En la replicación hay sólo un monitor. Debido a la indisponibilidad de unos de los monitores.
Dimensión modificada	Protocolo , específicamente, las guías
Amenaza a la validez	El cambio incrementa la validez interna ya que, al tener un solo monitor, no hay diferencias en la información recibida por los sujetos El cambio disminuye la validez de conclusión ya que, al tener un solo monitor, el proceso se vuelve tedioso para el monitor
Cambio #2	Duración de la formación previa Originalmente , la formación breve en actividades relacionadas con requisitos previa es de 1 semana. En la replicación la formación breve es de 6 semanas. Debido a que se quiere estudiar el efecto de la formación
Dimensión modificada	Operacionalización , específicamente la causa
Amenaza a la validez	El cambio incrementa la validez del constructo ya que se refleja mejor el efecto de la formación El cambio incrementa la validez interna ya que se analiza el efecto de la formación frente a otros factores

Tabla B.10: Replicación E-2015, versión final

Replicación	<i>E-2015</i> replicación Interna basada en experimento original <i>E-2013</i>
Objetivo del experimento	Estudiar la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> en el 2013 y esta replicación en <i>E.T.S. de Ingenieros Informáticos, Universidad Politécnica de Madrid</i> , en el 2015
Propósito	Generalizar resultados
Cambio #1	Duración de la formación previa Originalmente , la formación breve en actividades relacionadas con requisitos es de 1 semana. En la replicación la formación breve es de 2 semanas. Debido a que se quiere estudiar el efecto de la formación
Dimensión modificada	Operacionalización , específicamente la causa
Amenaza a la validez	El cambio incrementa la validez del constructo ya que se refleja mejor el efecto de la formación El cambio incrementa la validez interna ya que se analiza el efecto de la formación frente a otros factores

Tabla B.11: Replicación VV-UPM1, versión final

Replicación	VV-UPM1 replicación Interna basada en experimento original VV-UPM
Objetivo del experimento	Evaluar la eficacia de tres técnicas de verificación y validación del código
Descripción del experimento	Los sujetos evalúan las tres técnicas aplicando cada técnica a cada uno de los programas en C conteniendo los errores a ser detectados. Previamente, los sujetos reciben formación en el uso de cada una de las técnicas de detección de errores.
Lugar y Fecha	El experimento base se llevó a cabo en <i>Universidad Politécnica de Madrid</i> y esta replicación en <i>Universidad Politécnica de Madrid</i>
Propósito	Generalizar resultados
Cambio #1	Visibilidad del fallo Originalmente , no se analiza la visibilidad del fallo En la replicación se analiza la visibilidad del fallo o número de personas que lo detectan Causado por la necesidad de obtener más conclusiones.
Dimensión modificada	Operacionalización , específicamente el efecto
Amenaza a la validez	El cambio aumenta la validez del constructo ya que se refleja mejor el efecto para la evaluación de cada técnica
Comentario	Se utiliza el paquete de laboratorio de Kamsties y Lott
Cambio #2	Dos versiones de cada programa Originalmente , el programa es un factor (variable independiente) aunque no se estudia su influencia. En la replicación se implementan dos versiones por programa que será un nuevo factor. Causado por la necesidad de obtener más conclusiones.
Dimensión modificada	Operacionalización , específicamente, la causa
Amenaza a la validez	El cambio aumenta la validez del constructo ya que se analiza la influencia de la versión del programa
Cambio #3	Aumento del número de fallos Originalmente , hay tres tipos de fallos que aparecen una vez. Los otros tres tipos de fallos aparecen dos veces. En la replicación se duplican todos los tipos de fallos. Ya que hay dos versiones de cada programa que posibilita añadir fallos.
Dimensión modificada	Operacionalización , específicamente, el efecto
Amenaza abordada	El cambio aumenta la validez interna ya que al aumentar el número de fallos a detectar, se facilita la comparación de las técnicas
Cambio #4	Casos de prueba Originalmente , los sujetos generan casos de prueba para detectar los fallos del código. En la replicación en primer lugar, los sujetos aplican la técnica para generar los casos de prueba y posteriormente, ejecutarán los casos de prueba que se les proporcionan para detectar los fallos del programa. Con el fin de comprobar si la visibilidad de los fallos influye en su detección.
Dimensión modificada	Protocolo , específicamente, material experimental
Amenaza abordada	El cambio aumenta la validez interna ya que al generar y ejecutar casos de prueba se refuerza el efecto de la técnica
Cambio #5	Descarte de un programa Originalmente , se utilizaron cuatro programas. En la replicación se utilizan tres programas, uno es descartado. Con el fin de equilibrar el diseño.
Dimensión modificada	Protocolo , específicamente, diseño experimental
Amenaza abordada	El cambio disminuye la validez interna ya que se suprime uno de los programas sobre los que se detectaban los fallos
Cambio #6	Técnicas aplicadas por cada sujeto Originalmente , cada sujeto aplica una técnica. En la replicación cada sujeto aplica las tres técnicas. Ya que se cambia el diseño.
Dimensión modificada	Protocolo , específicamente, diseño experimental
Amenaza abordada	El cambio aumenta la validez interna ya que se aumentan las aplicaciones de cada técnica y se facilita su comparación

Tabla B.12: Replicación VV-UPV, versión final

Replicación	VV-UPV replicación Interna basada en experimento original VV-UPM
Objetivo del experimento	Evaluar la eficacia de tres técnicas de verificación y validación del código
Lugar y Fecha	El experimento base se llevó a cabo en <i>Universidad Politécnica de Madrid</i> y esta replicación en <i>Universidad Politécnica de Valencia</i>
Propósito	Generalizar resultados
Cambio #1	<p>Descarte de una de las técnicas</p> <p>Originalmente, se utilizan las tres técnicas de verificación y validación: lectura de código, partición de equivalencia y branch testing.</p> <p>En la replicación se omite la técnica lectura de código.</p> <p>Causado por restricciones de tiempo</p>
Dimensión modificada	Operacionalización , la causa
Amenazas a la validez	El cambio disminuye la validez del constructo ya que se suprime uno de los niveles del factor técnicas
Cambio #2	<p>Duración de las sesiones</p> <p>Originalmente, la duración de las tres sesiones es de cuatro horas cada una; es decir el tiempo es ilimitado.</p> <p>En la replicación la duración de cada una de las tres sesiones es de dos horas.</p> <p>Causado por restricciones de tiempo.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio disminuye la validez interna ya que la restricción de tiempo puede influir en los resultados
Cambio #3	<p>Reducción del período de formación</p> <p>Originalmente, los sujetos reciben tres sesiones de formación de cuatro horas para aprender el uso de las técnicas.</p> <p>En la replicación la formación consiste en dos breves tutoriales de dos horas.</p> <p>Ya que los sujetos ya están familiarizados con las técnicas.</p>
Dimensión modificada	Protocolo , específicamente las guías
Amenazas a la validez	El cambio no afecta a la validez ya que los sujetos ya están familiarizados con las técnicas y la formación no se operacionaliza
Comentario	La formación no se operacionaliza
Cambio #4	<p>Formación en cada técnica al comienzo de la sesión</p> <p>Originalmente, la formación en el uso de las técnicas es antes de que se ejecute el experimento.</p> <p>En la replicación en cada sesión, antes de la aplicación de la técnica se lleva a cabo, mediante un tutorial, la formación en el uso de dicha técnica .</p> <p>Ya que los sujetos ya están familiarizados con las técnicas.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio no afecta a la validez ya que los sujetos ya conocían las técnicas
Comentario	La formación no se operacionaliza
Cambio #5	<p>Aplicación de las técnicas</p> <p>Originalmente, los sujetos aplican una técnica a un programa en cada sesión.</p> <p>En la replicación aplican la misma técnica a diferentes programas en cada sesión.</p> <p>Causado por restricciones de tiempo.</p>
Dimensión modificada	Protocolo , específicamente, el diseño experimental
Amenazas a la validez	El cambio disminuye la validez interna ya que no se analizan y comparan las tres técnicas
Cambio #6	<p>Casos de prueba en sesión aparte</p> <p>Originalmente, los sujetos ejecutan casos de prueba con la aplicación de la técnica; es decir en cada sesión.</p> <p>En la replicación los sujetos ejecutan casos de prueba para uno de los programas que han probado en sesión aparte; es decir en la sesión 3.</p> <p>Causado por restricciones de tiempo.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio incrementa la validez interna ya que se analiza el efecto de realizar los casos de prueba en sesión aparte

Tabla B.13: Replicación VV-Uds, versión final

Replicación	<i>VV-Uds</i> replicación Interna basada en experimento original <i>VV-UPM</i>
Objetivo del experimento	Evaluar la eficacia de tres técnicas de verificación y validación del código
Lugar y Fecha	El experimento base se llevó a cabo en <i>Universidad Politécnica de Madrid</i> y esta replicación en <i>Universidad de Sevilla</i>
Propósito	Generalizar resultados
Cambio #1	<p>Duración de las sesiones Originalmente, la duración de las tres sesiones es de cuatro horas cada una; es decir el tiempo es ilimitado. En la replicación la duración de cada sesión es 2 horas. Causado por restricciones de tiempo.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio disminuye la validez interna ya que la restricción de tiempo puede afectar a los resultados
Cambio #2	<p>Casos de prueba en sesión aparte Originalmente, los sujetos ejecutan casos de prueba con la aplicación de la técnica; es decir en cada sesión. En la replicación los casos se ejecutan en sesión posterior; es decir en la sesión cuatro. Causado por restricciones de tiempo.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio disminuye la validez interna ya que la restricción de tiempo puede afectar a los resultados
Cambio #3	<p>Los sujetos trabajan en parejas Originalmente, los sujetos trabajan de forma individual. En la replicación los sujetos trabajan en parejas. Ya que no hay suficientes ordenadores</p>
Dimensión modificada	Contexto , específicamente, forma de trabajo
Amenazas a la validez	El cambio incrementa la validez Interna ya que se puede analizar el efecto de trabajar en forma conjunta
Cambio #4	<p>Reducción del período de formación Originalmente, los sujetos reciben tres sesiones de cuatro horas en formación en el uso de las técnicas. En la replicación la formación son dos tutoriales de dos horas Ya que ya están familiarizados con las técnicas.</p>
Dimensión modificada	Protocolo , específicamente las guías
Amenazas a la validez	El cambio no afecta a la validez ya que los sujetos ya están familiarizados con las técnicas y la formación no se operacionaliza
Cambio #5	<p>Formación en cada técnica al comienzo de la sesión Originalmente, el entrenamiento en el uso de las técnicas es antes de que se ejecute el experimento. En la replicación cada tutorial se lleva a cabo antes de la aplicación de la técnica, intercalado con el experimento. Ya que están familiarizados con las técnicas.</p>
Dimensión modificada	Protocolo , específicamente, las guías
Amenazas a la validez	El cambio no afecta a la validez ya que los sujetos ya conocían las técnicas

Tabla B.14: Replicación VV-ORT, versión final

Replicación	VV-ORT replicación Interna basada en experimento original VV-UPM
Objetivo del experimento Lugar y Fecha	Evaluar la eficacia de tres técnicas de verificación y validación del código El experimento base se llevó a cabo en <i>Universidad Politécnica de Madrid</i> y esta replicación en <i>Universidad ORT de Uruguay</i>
Propósito	Generalizar resultados
Cambio #1	Descarte de una de las técnicas Originalmente , se utilizan las tres técnicas de verificación y validación: lectura de código, partición de equivalencia y branch testing. En la replicación se omite la técnica lectura de código. Causado por restricciones de tiempo
Dimensión modificada Amenazas a la validez	Operacionalización , específicamente, la causa El cambio disminuye la validez del constructo ya que se suprime uno de los niveles del factor técnicas
Cambio #2	Descarte de un programa Originalmente , se utilizaron cuatro programas. En la replicación se descarta uno de los programas. Causado por restricciones de tiempo.
Dimensión modificada Amenaza abordada	Protocolo , específicamente, diseño experimental El cambio disminuye la validez interna ya que se suprime uno de los programas sobre los que se detectaban los fallos
Cambio #3	Duración de las sesiones Originalmente , el experimento se ejecuta en tres sesiones de cuatro horas cada una; es decir tiempo ilimitado. En la replicación el experimento se ejecuta en una única sesión. Causado por restricciones de tiempo.
Dimensión modificada Amenazas a la validez	Protocolo , específicamente, las guías El cambio disminuye la validez interna ya que la restricción de tiempo puede afectar a los resultados
Cambio #4	Aplicación de técnicas por programas Originalmente , los sujetos aplican una técnica a un programa en cada una de las tres sesiones. En la replicación los sujetos aplican las dos técnicas a los dos programas en una única sesión. Ya que la duración de la sesión es ilimitada.
Dimensión modificada Amenazas a la validez	Protocolo , específicamente, diseño experimental El cambio disminuye la validez interna ya que no se utiliza una de las técnicas y uno de los programas.

Tabla B.15: Replicación Suelo–2018, versión final

Replicación	<i>Suelo–2018</i> replicación Interna basada en el experimento original <i>Suelo–2016</i>
Objetivo del experimento	Evaluar el efecto de un biosurfactante en la fitorremediación asistida de suelos contaminados.
Descripción del experimento	Se analiza si la adición del biosurfactante JBR-425 a dos tipos de suelos contaminados por cobre (Cu) (causa) reduce la concentración Cu en dicho suelo, mediante la determinación del Cu extraído (efecto) por las plantas Cebada (<i>Hordeum vulgare</i> L.) y Mostaza (<i>Brassica juncea</i> L.) germinadas en dichos suelos. A partir de 2 tipos de suelo, se preparan dos grupos de macetas. Un grupo usando la muestra de suelo contaminada artificialmente con Cu y al que se le ha añadido JBR-425 (tratamiento) y otro grupo al que no se le ha añadido JBR-425 ni Cu (grupo de control), En ambos grupos de macetas se plantan semillas de cebada o de Mostaza, se las deja germinar y crecer un cierto tiempo, y luego se analiza su biomasa (procedimiento de medida) para determinar la concentración de metales (métrica) y analizar la diferencia entre los tratamientos.
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Agronómica, Sevilla</i> en <i>Octubre–2015</i> y esta replicación, en <i>E.T.S. Ingeniería Agronómica, Sevilla</i> en <i>Marzo–2018</i>
Propósito	Generalizar resultados
Cambio #1	Medio de cultivo Originalmente , el experimento se llevó a cabo en una cámara de cultivo. En la replicación , se llevó a cabo en un invernadero Con el fin de simular las condiciones naturales
Dimensión Modificada Amenazas a la validez	Contexto , específicamente el medio El cambio aumenta la validez externa ya que permite generalizar los resultados realizando la replicación en condiciones más cercanas a las naturales
Cambio #2	Tipos de plantas Originalmente , se utilizaron dos plantas: Cebada y Mostaza En la replicación , sólo se utilizó Mostaza Ya que se observó que sólo Mostaza era una planta acumuladora de metal
Dimensión Modificada Amenazas a la validez	Protocolo , específicamente los instrumentos de medida El cambio aumenta la validez del constructo ya que en la Mostaza se puede medir el efecto, es decir el Cu extraído
Comentario	Al utilizar un solo tipo de planta, no afecta a los resultados. No se operacionaliza.
Cambio #3	Tipos de suelos Originalmente , había dos tipos de suelo: Coria (pH=7.8) y Constantina (pH=5.5) En la replicación , sólo se utilizó el suelo de Constantina Ya que en el suelo de Coria el metal estaba fuertemente absorbido y la fitoextracción no afecta a la producción de biomasa
Dimensión Modificada Amenazas a la validez	Población , específicamente Tipo de suelo El cambio aumenta la validez del constructo ya que se asegura que el metal puede ser extraído del suelo. El Cu extraído es el efecto que se quiere medir
Cambio #4	Reducción de dosis de cobre Originalmente , las dosis de cobre (Cu) añadidas al suelo fueron 0, 500 y 1000 mg kg ⁻¹ En la replicación , las dosis de Cu se ajustaron a 0, 125, 250 y 500 mg kg ⁻¹ Ya que la dosis de cobre de 1000 mg kg ⁻¹ era tóxica para la planta
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente la causa El cambio aumenta la validez del constructo ya que la dosis de cobre se ajusta a niveles no tóxicos para la planta El cambio aumenta la validez del interna ya que se analiza el efecto en niveles de Cu válidos

Cambio #5	<p>Forma de aplicación del cobre Originalmente, El cobre se aplicó como Nitrato de cobre En la replicación, El cobre se aplica como Sulfato de cobre Ya que el Sulfato de cobre es más accesible Operacionalización, específicamente la causa El cambio no afecta a la validez ya que el reactivo es equivalente Comentario El cambio afecta a la forma de aplicar el Cu que es el tratamiento</p>
Dimensión Modificada Amenazas a la validez Comentario	
Cambio #6	<p>Tiempo de envejecimiento del suelo Originalmente, el tiempo de envejecimiento del suelo (desde que se aplica Cu hasta que se siembra la planta) es de 45 días En la replicación, el tiempo de envejecimiento del suelo es de 15 días Con el fin que el Cu no se retenga tanto y sea extraído con más facilidad Operacionalización, específicamente la causa El cambio aumenta la validez del constructo ya que la planta extrae el metal con mayor facilidad y se mide mejor el efecto El cambio disminuye la validez del externa ya que reduce la generalización de resultados a suelos contaminados desde hace más tiempo Comentario El cambio afecta a la forma de aplicar el Cu que es el tratamiento</p>
Dimensión Modificada Amenazas a la validez Comentario	
Cambio #7	<p>Aumento de la biomasa obtenida Originalmente, había (3x2x2x2) 24 unidades experimentales (3 niveles de Cu, 2 de JBR, 2 plantas y 2 suelos). Cada unidad experimental es una maceta y se repite 3 veces. En la replicación, hay (4x2) 8 unidades experimentales (4 niveles de Cu y 2 de JBR). Para cada unidad experimental se preparan 4 macetas en una bandeja. Se repite 3 veces Con el fin de aumentar el número de macetas para obtener suficiente biomasa Protocolo, específicamente el diseño experimental El cambio aumenta la validez de conclusión ya que aumentar el número de sujetos (n) mejora la potencia de los tests estadísticos, reduciendo la probabilidad de obtener un falso negativo (error tipo II).</p>
Dimensión Modificada Amenazas a la validez	
Cambio #8	<p>Obtener la biomasa en etapa de fructificación Originalmente, la biomasa se recoge cuando las plantas tienen entre 2 y 3 hojas verdaderas En la replicación, las plantas se aclaran cuando tienen entre 2 y 3 hojas reales y se dejan 4 plantas por maceta. La biomasa se recoge cuando las plantas alcanzan la etapa de fructificación. Con el fin de evitar la competencia entre plantas y dejar que las plantas completen su ciclo vegetativo para así obtener más biomasa Operacionalización, específicamente el procedimiento de medida El cambio aumenta la validez del constructo ya que se obtiene más biomasa para análisis posteriores</p>
Dimensión Modificada Amenazas a la validez	
Cambio #9	<p>Aumento del volumen de suelo Originalmente, las macetas son del tipo de tubo de 300 ml En la replicación, las macetas son del tipo cubo de 500 ml. Ya que al tener mayor volumen de suelo se permite un mayor desarrollo radicular y una mayor producción de biomasa Protocolo, específicamente el material experimental El cambio aumenta la validez del constructo ya que se obtiene más biomasa para análisis posteriores</p>
Dimensión Modificada Amenazas a la validez	

Tabla B.16: Replicación Suelo–2019, versión final

Replicación	<i>Suelo–2019</i> replicación Interna basada en el experimento original <i>Suelo–2016</i>
Descripción del experimento	Evaluar el efecto de un biosurfactante en la fitorremediación asistida de suelos contaminados.
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Agronómica, Sevilla</i> en <i>Octubre–2015</i> y esta replicación, en <i>E.T.S. Ingeniería Agronómica, Sevilla</i> en <i>Marzo–2019</i>
Propósito	Generalizar resultados
Cambio #1	Medio de cultivo Originalmente , el experimento se llevó a cabo en una cámara de cultivo. En la replicación , se llevó a cabo en un invernadero Con el fin de simular las condiciones naturales
Dimensión Modificada Amenazas a la validez	Contexto , específicamente el medio El cambio aumenta la validez externa ya que permite generalizar los resultados realizando la replicación en condiciones más cercanas a las naturales
Cambio #2	Tipos de plantas Originalmente , se utilizaron dos plantas: Cebada y Mostaza En la replicación , sólo se utilizó Mostaza Ya que se observó que sólo Mostaza era una planta acumuladora de metal
Dimensión Modificada Amenazas a la validez	Protocolo , específicamente los instrumentos de medida El cambio aumenta la validez del constructo ya que en la Mostaza se puede medir el efecto, es decir el Cu extraído
Comentario	Al utilizar un solo tipo de planta, no afecta a los resultados. No se operacionaliza.
Cambio #3	Tipos de suelos contaminados de forma natural Originalmente , había dos tipos de suelo: Coria (pH=7.8) y Constantina (pH=5.5) En la replicación , hay tres tipos de suelo: Miraflores-1 (Pb=158 mg kg ⁻¹ , Zn=125 mg kg ⁻¹ y Cu=36 mg kg ⁻¹) y Miraflores-2 (Pb=375 mg kg ⁻¹ , Zn=192 mg kg ⁻¹ y Cu=206 mg kg ⁻¹) y Lebrija (suelo no contaminado por metales) Con el fin de experimentar con suelos contaminados naturalmente.
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente la causa El cambio aumenta la validez interna al comparar suelos de distintas características El cambio aumenta la validez de conclusión ya que se refleja mejor el constructo
Comentario	Los suelos de Miraflores son jardines urbanos con contaminación natural y el suelo de Lebrija fue utilizado como control
Cambio #4	Se anula la dosis de Cu artificial Originalmente , las dosis de cobre (Cu) añadidas al suelo fueron 0, 500 y 1000 mg kg ⁻¹ En la replicación , los suelos no estén contaminados artificialmente con Cu Ya que estos suelos son huertos urbanos con contaminación natural (niveles de Cu 36 y 206 mg kg ⁻¹)
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente la causa El cambio aumenta la validez interna ya que la niveles de Cu son las presentes en el suelo y no son tóxicos para la planta
Cambio #5	Aumento de los tipos de suelos Originalmente , había 6 tratamientos correspondientes a los 3 niveles de Cu y con/sin surfactante. Había 2 suelos y 2 tipos de plantas. Esto representa 24 unidades experimentales (3x2x2x2). Para cada unidad experimental se prepararon 3 macetas. En total había 72 macetas (3x2x2x2x3) En la replicación , hay 2 tratamientos correspondientes a con/sin surfactante. Había 3 suelos y 1 tipo de planta. Esto representa 6 unidades experimentales. Cada unidad experimental se repitió 4 veces (2x3x4 macetas) formando un bloque. Esto se repite 3 veces para tener 3 bloques. En total hay 72 macetas (2x3x4x3). Dentro de cada bloque, las macetas se distribuyen al azar. La unidad experimental es la maceta. Ya que se han ajustado los tipos de plantas y suelos
Dimensión Modificada Amenazas a la validez	Protocolo , específicamente el diseño experimental El cambio aumenta la validez interna ya que permite comparar el efecto sobre distintos tipos de suelos

Cambio #6	<p>Recogida de Biomasa en fructificación Originalmente, la biomasa se recoge cuando las plantas tienen entre 2 y 3 hojas verdaderas En la replicación, las plantas se aclaran cuando tienen entre 2 y 3 hojas reales y sólo queda 1 planta por maceta. La biomasa se recoge cuando las plantas alcanzan la etapa de fructificación. Con el fin evitar la competencia entre plantas y dejar que las plantas completen su ciclo vegetativo para así obtener más biomasa</p>
Dimensión Modificada Amenazas a la validez	<p>Operacionalización, específicamente el procedimiento de medida El cambio aumenta la validez del constructo ya que se obtiene más biomasa para análisis posteriores</p>
Cambio #7	<p>Aumento del volumen de suelo Originalmente, las macetas son del tipo de tubo de 300 ml En la replicación, las macetas son del tipo cubo de 500 ml. Ya que al tener mayor volumen de suelo se permite un mayor desarrollo radicular y una mayor producción de biomasa</p>
Dimensión Modificada Amenazas a la validez	<p>Protocolo, específicamente el material experimental El cambio aumenta la validez del constructo ya que se obtiene más biomasa para análisis posteriores</p>

Tabla B.17: Replicación Calidad–2017P, versión final

Replicación	<i>Calidad–2017P</i> replicación Interna basada en el experimento original <i>Calidad–2016A</i>
Objetivo del experimento	Evaluar el efecto de la variedad de aceituna sobre la calidad del aceite de oliva virgen obtenido
Descripción del experimento	Se analiza y compara la calidad del aceite de oliva obtenido en tres variedades de aceitunas durante dos campañas agrícolas
Lugar y Fecha	El experimento base se llevó a cabo en <i>Instituto de la Grasa, Sevilla</i> en 2016 y esta replicación, en <i>Instituto de la Grasa, Sevilla</i> en 2017
Propósito	Generalizar resultados
Cambio #1	<p>Tipo de aceituna Originalmente, el tipo de aceituna era <i>Arbequina</i> En la replicación, el tipo de aceituna era <i>Picual</i> Con el fin de estudiar la diferencia en la calidad del aceite en las diferentes variedades</p>
Cambio #2	<p>Diferentes condiciones climáticas Originalmente, las condiciones climáticas son las del 2016 En la replicación, las condiciones climáticas son las del 2017 Con el fin de analizar datos correspondientes a dos campañas</p>

Tabla B.18: Replicación Calidad–2017V, versión final

Replicación	<i>Calidad–2017V</i> replicación Interna basada en el experimento original <i>Calidad–2016A</i>
Objetivo del experimento	Evaluar el efecto de la variedad de aceituna sobre la calidad del aceite de oliva virgen obtenido
Lugar y Fecha	El experimento base se llevó a cabo en <i>Instituto de la Grasa, Sevilla</i> en 2016 y esta replicación, en <i>Instituto de la Grasa, Sevilla</i> en 2017
Propósito	Generalizar resultados
Cambio #1	<p>Tipo de aceituna</p> <p>Originalmente, el tipo de aceituna era <i>Arbequina</i></p> <p>En la replicación, el tipo de aceituna era <i>Verdial</i></p> <p>Con el fin de estudiar la diferencia en la calidad del aceite en las diferentes variedades</p>
Cambio #2	<p>Diferentes condiciones climáticas</p> <p>Originalmente, las condiciones climáticas son las del 2016</p> <p>En la replicación, las condiciones climáticas son las del 2017</p> <p>Con el fin de analizar datos correspondientes a dos campañas</p>

Tabla B.19: Replicación Oliva–Des, versión final

Replicación	<i>Oliva–Des</i> replicación Interna basada en el experimento original <i>Oliva–2015</i>
Objetivo del experimento	Extracción de <i>compuestos fenólicos (EPC)</i> del aceite de oliva virgen con <i>disolventes verdes (Deep Eutectic Solvents DES)</i>
Descripción del experimento	Se trata de probar distintos disolventes que no sean tóxicos, para extraer los compuestos fenólicos del aceite de oliva virgen.
Lugar y Fecha	El experimento base se llevó a cabo en <i>Instituto de la Grasa, Sevilla</i> en 2015 y esta replicación, en <i>Instituto de la Grasa, Sevilla</i> en 2015
Propósito	Generalizar resultados
Cambio #1	Extracción con disolvente 1 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y glicerol (1:2) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos
Cambio #2	Extracción con disolvente 2 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y ácido láctico (1:2) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos
Cambio #3	Extracción con disolvente 3 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y Urea (1:2) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos
Cambio #4	Extracción con disolvente 4 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y Sacarosa (1:1) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos
Cambio #5	Extracción con disolvente 5 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y Sacarosa (4:1) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos
Cambio #6	Extracción con disolvente 6 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y 1,4-Butanodiol (1:5) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos
Cambio #7	Extracción con disolvente 7 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y Xylitol (2:1) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos
Cambio #8	Extracción con disolvente 8 Originalmente , la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación , la extracción se realiza con una solución de cloruro de colina y 1,2-Propanediol (1:1) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos

Cambio #9	<p>Extracción con disolvente 9 Originalmente, la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación, la extracción se realiza con una solución de cloruro de colina y Ácido malónico (1:1) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos</p>
Cambio #10	<p>Extracción con disolvente 10 Originalmente, la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación, la extracción se realiza con una solución de cloruro de colina, Urea and Glycerol (1:1:1) Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos</p>
Cambio #11	<p>Extracción con disolvente 11 Originalmente, la extracción se realiza con una solución al 80% (v/v) de metanol y agua En la replicación, la extracción se realiza con una solución <i>D-(-)-Fructosa D-(+)-Glucosa y sacarosa (1:1:1)</i> Con el fin de analizar alternativas no tóxicas para la extracción de compuestos fenólicos</p>

Tabla B.20: Replicación Dieta–Hiper, versión final

Replicación	<i>Dieta–Hiper</i> replicación Interna basada en el experimento original <i>Dieta</i>
Objetivo del experimento	Analizar el efecto de la alimentación rica en ácido oleico en sujetos <i>hipertrigliceridémicos</i>
Descripción del experimento	Los sujetos se alimentan de una dieta rica en ácido oleico y regularmente se realizan análisis para determinar, entre otros, sus niveles de colesterol .
Lugar y Fecha	El experimento base se llevó a cabo en <i>Instituto de la Grasa, Sevilla</i> en 2006 y esta replicación, en <i>Instituto de la Grasa, Sevilla</i> en 2006
Propósito	Generalizar resultados
Cambio #1	<p>Sujetos hipertensos Originalmente, Los sujetos tienen niveles de presión arterial dentro de los límites considerados normales. En la replicación, Los sujetos son <i>hipertensos</i>. Con el fin de estudiar el efecto de la dieta rica en ácido oleico en sujetos <i>hipertrigliceridémicos</i> que también son <i>hipertensos</i></p>

Tabla B.21: Replicación SPL–Pr&Com, versión final

Replicación	<i>SPL–Pr&Com</i> replicación Interna basada en el experimento original <i>SPL–Pr</i>
Objetivo del experimento	Comparación de los criterios de priorización de casos de prueba para Líneas de Productos Software (SPL).
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2014 y esta replicación, en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2014
Propósito	Generalizar resultados
Cambio #1	Conjunto de pruebas Originalmente , sólo se generó un conjunto de pruebas En la replicación , para cada modelo, se generó un conjunto de pruebas Con el fin de obtener una lista de productos que cubra todos las posibles características de cada modelo
Dimensión Modificada Amenazas a la validez	Protocolo , específicamente las guías El cambio no afecta a la validez
Cambio #2	Generación del conjunto de pruebas Originalmente , el conjunto de pruebas se generó aleatoriamente utilizando la herramienta SPLAR En la replicación , el conjunto de pruebas se generó aleatoriamente usando la herramienta SPLCAT Ya que SPLCAT aumenta la tasa de detección de fallas y por lo tanto es considerado como un enfoque de priorización extra en nuestra comparación.
Dimensión Modificada Amenazas a la validez	Protocolo , específicamente el material experimental El cambio no afecta a la validez ya que sólo cambia la herramienta utilizada

Tabla B.22: Replicación Test–NF, versión final

Replicación	<i>Test–NF</i> replicación Interna basada en el experimento original <i>Test–F</i>
Objetivo del experimento	Priorización de la ejecución de casos de prueba para acelerar la detección de fallos en sistemas altamente configurables.
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2015 y esta replicación, en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2015
Propósito	Generalizar resultados
Cambio #1	Objetivos no–funcionales Originalmente , los objetivos son funcionales En la replicación , los objetivos son no–funcionales Con el fin de comparar las diferencias a favor de la priorización multi–objetivo sobre la priorización mono–objetivo utilizando objetivos no–funcionales.
Dimensión Modificada Amenazas a la validez	Operacionalización , específicamente el efecto El cambio no afecta a la validez ya que se sustituye una variable dependiente del experimento original por otra variable dependiente y por tanto la validez no se ve afectada.

Tabla B.23: Replicación Test–F&NF, versión final

Replicación	<i>Test–F&NF</i> replicación Interna basada en el experimento original <i>Test–F</i>
Objetivo del experimento	Priorización de la ejecución de casos de prueba para acelerar la detección de fallos en sistemas altamente configurables.
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2015 y esta replicación, en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2015
Propósito	Generalizar resultados
Cambio #1	<p>Objetivos funcionales y no–funcionales</p> <p>Originalmente, los objetivo son funcionales</p> <p>En la replicación, se combinan los objetivos funcionales y no–funcionales</p> <p>Con el fin de analizar si el establecimiento de prioridades mediante objetivos funcionales y no funcionales supera al establecimiento de prioridades basado en un único objetivo, ya sea funcional o no funcional.</p>
Dimensión Modificada	Operacionalización , específicamente, el efecto
Amenazas a la validez	El cambio no afecta a la validez ya que se sustituye una variable dependiente del experimento original por otra variable dependiente y por tanto la validez no se ve afectada.

Tabla B.24: Replicación Test–FvsNF, versión final

Replicación	<i>Test–FvsNF</i> replicación Interna basada en el experimento original <i>Test–F</i>
Objetivo del experimento	Priorización de la ejecución de casos de prueba para acelerar la detección de fallos en sistemas altamente configurables.
Lugar y Fecha	El experimento base se llevó a cabo en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2015 y esta replicación, en <i>E.T.S. Ingeniería Informática, Universidad de Sevilla</i> en 2015
Propósito	Generalizar resultados
Cambio #1	<p>Comparación de objetivos</p> <p>Originalmente, los objetivo son funcionales</p> <p>En la replicación, se comparan los objetivos funcionales y no–funcionales</p> <p>Con el fin de analizar el dominio de los objetivos no–funcionales por encima de los objetivos funcionales, especialmente cuando éstos se combinan en una perspectiva multi–objetivo.</p>
Dimensión Modificada	Operacionalización , específicamente el efecto
Amenazas a la validez	El cambio no afecta a la validez ya que se sustituye una variable dependiente del experimento original por otra variable dependiente y por tanto la validez no se ve afectada.

BIBLIOGRAFÍA

Al final de cada referencia bibliográfica aparece un enlace a cada página en que se cita dicha referencia.

- [1] ACM Association for Computing Machinery. ACM Publications <https://www.acm.org/publications/policies/artifact-review-badging>, [Accedido Febrero-2020]. (pages 5, 22, 58).
- [2] S. Abrahão, C. Gravino, E. Insfran, G. Scanniello, and G. Tortora. Assessing the effectiveness of sequence diagrams in the comprehension of functional requirements: Results from a family of five experiments. *IEEE Transactions on Software Engineering*, 39(3):327–342, 2013. (pages 49, 51, 52).
- [3] S. Abrahão and E. Insfran. Evaluating software architecture evaluation methods: An internal replication. In *Evaluation and Assessment in Software Engineering (EA-SE)*, pages 144–153, 2017. (pages 51, 52).
- [4] S. T. Acuña, M. N. Gómez, J. E. Hannay, N. Juristo, and D. Pfahl. Are team personality and climate related to satisfaction and software quality? aggregating results from a twice replicated experiment. *Information and Software Technology*, 57(1):141–156, 2015. (pages 51, 52).
- [5] S. Akbarinasaji, B. Caglayan, and A. Bener. Predicting bug-fixing time: A replication study using an open source software project. *Journal of Systems and Software*, 136:173–186, 2018. (pages 51, 52).
- [6] O. Albayrak and J. C. Carver. Investigation of individual factors impacting the effectiveness of requirements inspections: A replicated experiment. *Empirical Software Engineering*, 19(1):241–266, 2014. (pages 51, 52, 72, 75, 76, 77).
- [7] S. Ali, T. Yue, and L.C. Briand. Does aspect-oriented modeling help improve the readability of uml state machines? *Software and Systems Modeling*, 13(3):1189–1221, 2014. (pages 51, 52).

- [8] K. AlKharabsheh, Y. Crespo, E. Manso, and J. A. Taboada. Factores que tienen en cuenta los desarrolladores en la priorización de smells para su corrección: conclusiones después de una réplica. (page 21).
- [9] J. P. F. 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. (pages 31, 33, 34, 59, 70).
- [10] H. G. Amatriain, O. Dieste, E. Fernández, and R. García Martínez. Evaluación de la aplicabilidad del meta-análisis de efectos aleatorios en ingeniería del software. In *XVII Congreso Argentino de Ciencias de la Computación*, 2011. (page 22).
- [11] C. Apa. *Defectos dentro y fuera del alcance de técnicas de verificación: resultados del análisis de una familia de experimentos*. PhD thesis, 2015. (page 21).
- [12] C. Apa, O. Dieste, E.G. Espinosa G., and E.R. Fonseca C. Effectiveness for detecting faults within and outside the scope of testing techniques: An independent replication. *Empirical Software Engineering*, 19(2):378–417, 2014. (pages 51, 52, 72, 75).
- [13] A. Aranda, O. Dieste, and N. Juristo. Evidence of the presence of bias in subjective metrics: Analysis within a family of experiments. In *Evaluation and Assessment in Software Engineering (EASE)*, 2014. (pages 51, 52).
- [14] A. M. Aranda, O. Dieste, and N. Juristo. Effect of domain knowledge on elicitation effectiveness: An internally replicated controlled experiment. *IEEE Transactions on Software Engineering*, 42(5):427–451, 2016. (pages 51, 52).
- [15] A. Aranda López King. *Estudio empírico de la influencia de la experiencia y del conocimiento del dominio del analista en la efectividad de la educación de requisitos*. PhD thesis, ETSI Informatica, 2016. (pages 21, 114).
- [16] O. F. Arar and K. Ayan. Deriving thresholds of software metrics to predict faults on open source software: Replicated case studies. *Expert Systems with Applications*, 61:106–121, 2016. (pages 51, 52).
- [17] C. Arévalo, I. Ramos, J. Gutiérrez, and M. Cruz. Practical experiences in the use of pattern-recognition strategies to transform software project plans into software business processes of information technology companies. *Scientific Programming*. Vol. 2019. Núm. 2019. Pag. 1-21, 2019. (page 149).

- [18] S. Assar, M. Borg, and D. Pfahl. Using text clustering to predict defect resolution time: a conceptual replication and an evaluation of prediction accuracy. *Empirical Software Engineering*, 21(4):1437–1475, 2016. (pages 51, 52, 71, 72, 74, 75).
- [19] M. T. Baldassarre, J. C. Carver, O. Dieste, and N. Juristo. Replication types: Towards a shared taxonomy. In *Proceedings of EASE '14*, pages 18:1–18:4, 2014. (pages 4, 8, 32, 33, 34).
- [20] V. R. Basili, G. Caldiera, and H. D. Rombach. Goal question metrics paradigm. *Encyclopedia of Software Engineering*, pages 528–53, 1994. (page 84).
- [21] V. R. Basili, R. W. Selby, and D. H. Hutchens. Experimentation in software engineering. *IEEE Transactions on software engineering*, (7):733–743, 1986. (page 4).
- [22] V. R. Basili, F. Shull, and F. Lanubile. Building knowledge through families of experiments. *IEEE Transactions on Software Engineering*, 25(4):456–473, 1999. (pages 20, 30, 33, 34, 35, 58, 60).
- [23] G. Bavota, C. Gravino, R. Oliveto, A. De Lucia, G. Tortora, M. Genero, and J. A. Cruz-Lemus. A fine-grained analysis of the support provided by uml class diagrams and er diagrams during data model maintenance. *Software and Systems Modeling*, 14(1):287–306, 2015. (pages 51, 52).
- [24] M. Beck and J. Walden. Using software birthmarks and clustering to identify similar classes and major functionalities. In *ACM Southeast 2018 Conference*, 2018. (pages 51, 52).
- [25] B. Bernárdez, A. Durán, J. A. Parejo, and A. Ruiz-Cortés. A controlled experiment to evaluate the effects of mindfulness in software engineering. In *Proceedings of ESEM*, pages 17–27. ACM, 2014. (page 114).
- [26] B. Bernárdez, M. Cruz, A. Durán, J. A. Parejo, and A. Ruiz-Cortés. A family of experiments to evaluate the effects of mindfulness on software engineering students: The metamind dataset. *Research and Evidence in Software Engineering: From Empirical Studies to Open Source Artifacts*, 2020. En prensa. (page 149).
- [27] B. Bernárdez, A. Durán, J. A. Parejo, and A. 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. (pages 51, 52).

- [28] B. Bernárdez, A. Durán, J. A. Parejo, and A. 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. (page 114).
- [29] R. M. M. Bezerra, F. Q. B da Silva, A. M. Santana, C. V. C. Magalhães, and R. E. S. Santos. Replication of empirical studies in software engineering: An update of a systematic mapping study. In *Empirical Software Engineering and Measurement (ESEM), 2015 ACM/IEEE International Symposium on*, pages 1–4. IEEE, 2015. (pages 40, 62, 70).
- [30] D. Binkley, M. Davis, D. Lawrie, J. I. Maletic, C. Morrell, and B. Sharif. The impact of identifier style on effort and comprehension. *Empirical Software Engineering*, 18(2):219–276, 2013. (pages 49, 51, 52).
- [31] F. Blanco, J. C. Perales López, and M. A. Vadillo. ¿puede la psicología rescatarse a sí misma? incentivos, sesgos y replicabilidad, 2017. (page 58).
- [32] P. Bourque and R. E. Fairley, editors. *Guide to the software engineering body of knowledge, version 3.0*. IEEE Computer Society, 2014. (page 42).
- [33] A. Brooks, J. Chambers, C.N. Lee, and F. Mead. A partial replication with a sample size of one: A smoke test for empirical software engineering. In *International Workshop on Replication in Empirical Software Engineering (RESER)*, pages 56–65, 2013. (pages 51, 52).
- [34] A. Brooks, J. Daly, J. Miller, M. Roper, and M. Wood. Replication’s role in experimental computer science. *EfoCS-5-94 (RR/94/171)*, 1994. (pages 8, 58, 71, 141).
- [35] A. Brooks, J. Daly, J. Miller, M. Roper, and M. Wood. Replication of experimental results in software engineering. *Technical Report ISERN-96-10, University of Strathclyde*, 1996. (pages 30, 31, 33, 34, 58).
- [36] A. Brooks, M. Roper, M. Wood, J. Daly, and J. Miller. *Replication’s Role in Software Engineering*, pages 365–379. Springer London, London, 2008. (page 60).
- [37] V. C. Caires, N. Rios, J. Holvitie, V. Leppänen, M. G. De Mendonça Neto, and R. O. Spínola. Investigating the effects of agile practices and processes on technical debt—the viewpoint of the brazilian software industry. In *International Conference on Software Engineering and Knowledge Engineering (SEKE)*, volume 2018-July, pages 506–511, 2018. (pages 51, 52).

- [38] E. R. Fonseca Carrera. *Conceptualización e Infraestructura para la Investigación Experimental en Ingeniería del Software*. PhD thesis, Universidad Politécnica de Madrid, 2014. (page 21).
- [39] B. Cartaxo, A. Almeida, E. Barreiros, J. Saraiva, W. Ferreira, and S. Soares. Mechanisms to characterize context of empirical studies in software engineering. In *Experimental Software Engineering Latin American Workshop (ESELAW 2015)*, pages 1–14, 2015. (pages 59, 63, 70).
- [40] N. Cartwright. Replicability, reproducibility, and robustness: comments on harry collins. *History of Political Economy*, 23(1):143–155, 1991. (page 22).
- [41] S. Carvajal de la Haza. Efecto de la aplicación de ramnolípido en la fitoextracción de cobre por plantas de hordeum vulgare y brassica juncea en suelo contaminados artificialmente. 2016. (Trabajo Fin de Grado). Universidad de Sevilla, Sevilla. (page 123).
- [42] J. C. Carver. Towards reporting guidelines for experimental replications: a proposal. In *1st International Workshop on Replication in Empirical Software Engineering*, 2010. (pages 4, 8, 59, 60, 70, 71, 72, 74, 75, 76, 79, 119, 135, 142, 146, 147).
- [43] J. C. Carver, N. Juristo, M. T. Baldassarre, and S. Vegas. Replications of software engineering experiments, 2014. (pages 22, 58, 59, 60, 64, 78).
- [44] G. Cavalcanti, P. Accioly, and P. Borba. Assessing semistructured merge in version control systems: A replicated experiment. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 267–276, 2015. (pages 51, 52, 72, 75).
- [45] M. Ceccato, M. Di Penta, P. Falcarin, F. Ricca, M. Torchiano, and P. Tonella. A family of experiments to assess the effectiveness and efficiency of source code obfuscation techniques. *Empirical Software Engineering*, 19(4):1040–1074, 2014. (pages 49, 51, 52).
- [46] M. Ceccato, A. Marchetto, L. Mariani, C. D. Nguyen, and P. Tonella. Do automatically generated test cases make debugging easier? an experimental assessment of debugging effectiveness and efficiency. *ACM Transactions on Software Engineering and Methodology*, 25(1):5:1–5:38, 2015. (pages 51, 52).
- [47] P. Cedillo, E. Insfran, and S. Abrahão. Evaluación de un método de monitorización de calidad de servicios cloud: Una replicación interna. 2011. (page 21).

- [48] D. Chatterji, J. C. Carver, N. A. Kraft, and J. Harder. Effects of cloned code on software maintainability: A replicated developer study. In *Working Conference on Reverse Engineering (WCRE)*, pages 112–121, 2013. (pages 51, 52).
- [49] X. Chen, W. Zhang, P. Liang, and K. He. A replicated experiment on architecture pattern recommendation based on quality requirements. In *IEEE International Conference on Software Engineering and Service Sciences (ICSESS)*, pages 32–36, 2014. (pages 51, 52).
- [50] M. Ciolkowski, F. Shull, and S. Biffi. *A family of experiments to investigate the influence of context on the effect of inspection techniques*. na, 2002. (page 21).
- [51] M. Cruz, B. Bernárdez, A. Durán, J. A Galindo, and A. Ruiz-Cortés. Replication of studies in empirical software engineering: A systematic mapping study from 2013 to 2018. *IEEE Access* 8: 26773–26791, 2020. (pages 40, 149).
- [52] M. Cruz, B. Bernárdez, A. del Río-Ortega, and A. Durán. Una revisión de la notación ppinot para indicadores de rendimiento mediante su aplicación a un caso real. In *Actas de JCIS'2014*, 2014. (page 150).
- [53] M. Cruz, B. Bernárdez, and A. Durán. Una propuesta para especificar cambios de replications de experimentos en ingeniería del software. In *Actas de JISBD'2018*, 2018. (page 150).
- [54] M. Cruz, B. Bernárdez, M. Resinas, and A. Durán. Auditoría de procesos de negocio en la nube: persistencia mediante almacenes no relacionales. In *Actas de JISBD'2012*, 2012. (page 150).
- [55] J. A. Cruz-Lemus, M. Genero, M. E. Manso, S. Morasca, and M. Piattini. Assessing the understandability of uml statechart diagrams with composite states—a family of empirical studies. *Empirical Software Engineering*, 14(6):685–719, 2009. (page 84).
- [56] F. Q. B. Da Silva, A. C. C. França, M. Suassuna, L. M. R. De Sousa Mariz, I. Rosiley, R. C. G. De Miranda, T. B. Gouveia, C. V. F. Monteiro, E. Lucena, E. S. F. Cardozo, and E. Espindola. Team building criteria in software projects: A mix-method replicated study. *Information and Software Technology*, 55(7):1316–1340, 2013. (pages 51, 52).
- [57] F. Q. B. Da Silva, M. Suassuna, A. C. C. França, A. M Grubb, T. B. Gouveia, C. V. F. Monteiro, and I. E. dos Santos. Replication of empirical studies in software en-

- gineering research: a systematic mapping study. *Empirical Software Engineering*, 19(3):501–557, 2014. (pages 40, 48, 50, 58, 59, 62, 64, 70, 73, 78).
- [58] J. W. Daly, A. Brooks, J. Miller, M. Roper, and M. Wood. Verification of results in software maintenance through external replication. In *ICSM*, pages 50–57, 1994. (page 58).
- [59] C. V. C. de Magalhães, F. Q. B. da Silva, R. E. S. Santos, and M. Suassuna. Investigations about replication of empirical studies in software engineering: A systematic mapping study. *Information and Software Technology*, 64:76–101, 2015. (pages 8, 33, 40, 59, 62, 63, 70).
- [60] P. del Río Guijarro. Efecto de ramnolípido jbr-425 en el desarrollo de brassica juncea en suelos de huertos urbanos de sevilla. 2019. (Trabajo Fin de Grado). Universidad de Sevilla, Sevilla. (page 123).
- [61] A. C. Dias-Neto, S. Matalonga, M. Solari, G. Robiolo, and G. H. Travassos. Toward the characterization of software testing practices in south america: looking at brazil and uruguay. *Software Quality Journal*, 25(4):1145–1183, 2017. (pages 51, 52).
- [62] D. Dig, K. Manzoor, Ra. E Johnson, and T. N. Nguyen. Effective software merging in the presence of object-oriented refactorings. *IEEE Transactions on Software Engineering*, 34(3):321–335, 2008. (page 51).
- [63] S. M. Donadelli, Y. C. Zhu, and P. C. Rigby. Organizational volatility and post-release defects: A replication case study using data from google chrome. In *IEEE International Working Conference on Mining Software Repositories (MSR)*, pages 391–395, 2015. (pages 51, 52).
- [64] A. Durán Toro, B. Bernárdez Jiménez, A. Ruiz Cortés, and M. Toro Bonilla. A requirements elicitation approach based in templates and patterns. In *Proceedings of WER'99*, pages 17–29, 1999. (page 87).
- [65] A. Durán Toro, A. Ruiz Cortés, R. Corchuelo Gil, and M. Toro Bonilla. Supporting requirements verification using xslt. 2002. (page 87).
- [66] T. Dybå, D. I. K. Sjøberg, and D. S. Cruzes. What works for whom, where, when, and why?: on the role of context in empirical software engineering. In *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*, pages 19–28. ACM, 2012. (page 63).

- [67] W. Engelsman and R. Wieringa. Understandability of goal concepts by requirements engineering experts. In *International Conference on Conceptual Modeling (ER)*, volume 8823, pages 97–106, 2014. (pages 51, 52).
- [68] E. Espinosa, S. T. Acuña, S. Vegas, and N. Juristo. Adopting configuration management principles for managing experiment materials in families of experiments. *Information and Software Technology*, 113:39–67, 2019. (page 138).
- [69] L. Falcao, W. Ferreira, A. Borges, V. Nepomuceno, S. Soares, and M. T. Baldassare. An analysis of software engineering experiments using human subjects. In *Empirical Software Engineering and Measurement (ESEM), 2015 ACM/IEEE International Symposium on*, pages 1–4. IEEE, 2015. (pages 64, 70).
- [70] S. U. Farooq and S. M. K. Quadri. An externally replicated experiment to evaluate software testing methods. In *Evaluation and Assessment in Software Engineering (EASE)*, pages 72–77, 2013. (pages 51, 52).
- [71] S. U. Farooq, S. M. K. Quadri, and N. Ahmad. A replicated empirical study to evaluate software testing methods. *Journal of Software: Evolution and Process*, 29(9), 2017. (pages 51, 52).
- [72] M. Felderer and A. Herrmann. Manual test case derivation from uml activity diagrams and state machines: A controlled experiment. *Information and Software Technology*, 61:1–15, 2015. (pages 51, 52).
- [73] K. R. Felizardo, E. F. Barbosa, and J. C. Maldonado. A visual approach to validate the selection review of primary studies in systematic reviews: A replication study. In *International Conference on Software Engineering and Knowledge Engineering (SEKE)*, pages 141–146, 2013. (pages 51, 52).
- [74] K. R. Felizardo, E. F. Barbosa, R. M. Martins, P. H. D. Valle, and J. C. Maldonado. Visual text mining: Ensuring the presence of relevant studies in systematic literature reviews. *International Journal of Software Engineering and Knowledge Engineering*, 25(5):909–928, 2015. (pages 51, 52).
- [75] K. R. Felizardo, S. R. S. Souza, and J. C. Maldonado. The use of visual text mining to support the study selection activity in systematic literature reviews: A replication study. In *International Workshop on Replication in Empirical Software Engineering (RESER)*, pages 91–100, 2013. (pages 51, 52).

- [76] N. Fenton and S. Pfleeger. *Software Metrics: A Rigorous and Practical Approach*. PWS Publisher, 1997. (page 24).
- [77] A. Fernandez, S. Abrahão, and E. Insfran. Empirical validation of a usability inspection method for model-driven web development. *Journal of Systems and Software*, 86(1):161–186, 2013. (pages 49, 51, 52).
- [78] D. M. Fernández, Martin Monperrus, Robert Feldt, and Thomas Zimmermann. The open science initiative of the empirical software engineering journal. *Empirical Software Engineering*, 24(3):1057–1060, 2019. (page 6).
- [79] R. Fernández, A. Trapero, and J. Domínguez. Experimentación en agricultura. *Consejería de Agricultura y Pesca, Servicio de Publicaciones y Divulgación*, 2010. (pages 104, 127).
- [80] D. M. Fernández and S. Wagner. Naming the pain in requirements engineering: A design for a global family of surveys and first results from germany. *Information and Software Technology*, 57(1):616–643, 2015. (pages 49, 51, 52).
- [81] D. M. Fernández, S. Wagner, M. Kalinowski, M. Felderer, P. Mafra, A. Vetrò, T. Conte, M. T. Christiansson, D. Greer, C. Lassenius, T. Männistö, M. Nayabi, M. Oivo, B. Penzenstadler, D. Pfahl, R. Prikladnicki, G. Ruhe, A. Schekelmann, S. Sen, R. O. Spínola, A. Tuzcu, J. L. de la Vara, and R. Wieringa. Naming the pain in requirements engineering: Contemporary problems, causes, and effects in practice. *Empirical Software Engineering*, 22(5):2298–2338, 2017. (pages 49, 51, 52, 72, 73, 76).
- [82] A. M. Fernández-Sáez, M. Genero, D. Caivano, and M. R. V. Chaudron. Does the level of detail of uml diagrams affect the maintainability of source code?: a family of experiments. *Empirical Software Engineering*, 21(1):212–259, 2016. (pages 51, 52, 71, 72, 73, 74, 76).
- [83] A. M. Fernández-Sáez, M. Genero, M. R. V. Chaudron, D. Caivano, and I. Ramos. Are forward designed or reverse-engineered uml diagrams more helpful for code maintenance?: A family of experiments. *Information and Software Technology*, 57(1):644–663, 2015. (pages 51, 52).
- [84] F. Fittkau, S. Finke, W. Hasselbring, and J. Waller. Comparing trace visualizations for program comprehension through controlled experiments. In *IEEE International Conference on Program Comprehension (ICPC)*, pages 266–276, 2015. (pages 51, 52).

- [85] X. Franch, D. M. Fernández, M. Oriol, A. Vogelsang, R. Haldal, E. Knauss, G. H. Travassos, J. C. Carver, O. Dieste, and T. Zimmermann. How do practitioners perceive the relevance of requirements engineering research? an ongoing study. In *International Requirements Engineering Conference (RE)*, pages 382–387, 2017. (pages 51, 52).
- [86] A. C. C. França, F. Q. B. Da Silva, A. D. L. C. Felix, and D. E. S. Carneiro. Motivation in software engineering industrial practice: A cross-case analysis of two software organisations. *Information and Software Technology*, 56(1):79–101, 2014. (pages 51, 52).
- [87] D. Fucci, G. Scanniello, S. Romano, M. Shepperd, B. Sigweni, F. Uyaguari, B. Turhan, N. Juristo, and M. Oivo. An external replication on the effects of test-driven development using a multi-site blind analysis approach. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 3:1–3:10, 2016. (pages 51, 52, 72, 75).
- [88] D. Fucci, G. Scanniello, S. Romano, M. Shepperd, B. Sigweni, F. Uyaguari, B. Turhan, N. Juristo, and M. 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. (page 77).
- [89] D. Fucci and B. Turhan. A replicated experiment on the effectiveness of test-first development. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 103–112, 2013. (pages 51, 52, 72, 75).
- [90] D. Fucci and B. Turhan. On the role of tests in test-driven development: A differentiated and partial replication. *Empirical Software Engineering*, 19(2):277–302, 2014. (pages 51, 52, 72, 75).
- [91] E. G. Espinosa Gallardo. *Gestión de configuración y línea de productos para mejorar el proceso experimental en ingeniería del software*. PhD thesis, Universidad Politécnica de Madrid, 2014. (page 21).
- [92] A. García, E. Rodríguez-Juan, G. Rodríguez-Gutiérrez, J. Julian Rios, and J. Fernández-Bolaños. Extraction of phenolic compounds from virgin olive oil by deep eutectic solvents (dcss). *Food chemistry*, 197:554–561, 2016. (page 124).
- [93] B. Ghotra, S. McIntosh, and A. E. Hassan. Revisiting the impact of classification techniques on the performance of defect prediction models. In *International Con-*

- ference on Software Engineering (ICSE), volume 1, pages 789–800, 2015. (pages 49, 51, 52, 54).
- [94] D. Girardi, F. Lanubile, N. Novielli, and D. Fucci. Sensing developers' emotions: The design of a replicated experiment. In *International Conference on Software Engineering (ICSE)*, pages 51–54. IEEE Computer Society, 2018. (pages 51, 52).
- [95] O. S. Gómez, N. Juristo, and S. Vegas. Replication, reproduction and re-analysis: Three ways for verifying experimental findings. In *Proceedings of the 1st international workshop on replication in empirical software engineering research (RESER 2010), Cape Town, South Africa*, 2010. (page 22).
- [96] O. S. Gómez, N. Juristo, and S. Vegas. Replications types in experimental disciplines. In *Proceedings of ESEM'10*. ACM, 2010. (pages 22, 33, 59, 84).
- [97] O. S. Gómez, N. Juristo, and S. Vegas. Understanding replication of experiments in software engineering: A classification. *Information and Software Technology*, 56(8):1033–1048, 2014. (pages 8, 25, 26, 28, 31, 33, 34, 51, 58, 63, 77, 79, 85, 89, 90, 91, 93, 119, 135).
- [98] J. Gonzalez-Huerta, E. Insfran, S. Abrahão, and G. Scanniello. Validating a model-driven software architecture evaluation and improvement method: A family of experiments. *Information and Software Technology*, 57(1):405–429, 2015. (pages 51, 52).
- [99] C. Gravino, G. Scanniello, and G. Tortora. Source-code comprehension tasks supported by uml design models: Results from a controlled experiment and a differentiated replication. *Journal of Visual Languages and Computing*, 28:23–38, 2015. (pages 51, 52).
- [100] T. G. Grbac, P. Runeson, and D. Huljenić. A quantitative analysis of the unit verification perspective on fault distributions in complex software systems: an operational replication. *Software Quality Journal*, 24(4):967–995, 2016. (pages 51, 52).
- [101] X. Guerrón, S. Abrahão, and E. Insfran. Métricas de calidad interna y externa para servicios cloud: Un mapeo sistemático. (page 21).
- [102] L. Guerrouj, M. Di Penta, Y.-G. Guéhéneuc, and G. Antoniol. An experimental investigation on the effects of context on source code identifiers splitting and

- expansion. *Empirical Software Engineering*, 19(6):1706–1753, 2014. (pages 51, 52, 71, 72, 76).
- [103] M. N. Gómez and S.T. Acuña. A replicated quasi-experimental study on the influence of personality and team climate in software development. *Empirical Software Engineering*, 19(2):343–377, 2014. (pages 51, 52, 72, 76, 77).
- [104] O. S. Gómez, K. Cortés-Verdín, and C. J. Pardo. Efficiency of software testing techniques: A controlled experiment replication and network meta-analysis. *E-Informatica Software Engineering Journal*, 11(1):77–102, 2017. (pages 51, 52).
- [105] L. Haaranen, P. Ihanola, J. Sorva, and A. Vihavainen. In search of the emotional design effect in programming. In *International Conference on Software Engineering (ICSE)*, volume 2, pages 428–434, 2015. (pages 51, 52).
- [106] I. Hadar, I. Reinhartz-Berger, T. Kuflik, A. Perini, F. Ricca, and A. Susi. Comparing the comprehensibility of requirements models expressed in use case and tropos: Results from a family of experiments. *Information and Software Technology*, 55(10):1823–1843, 2013. (pages 49, 51, 52).
- [107] J. Harder and N. Göde. Cloned code: Stable code. *Journal of Software: Evolution and Process*, 25(10):1063–1088, 2013. (pages 51, 52).
- [108] L. Hattori, M. D’Ambros, M. Lanza, and M. Lungu. Answering software evolution questions: An empirical evaluation. *Information and software technology*, 55(4):755–775, 2013. (page 50).
- [109] C. Hendrick. Replications, strict replications, and conceptual replications: are they important? *Journal of Social Behavior and Personality*, 5(4):41, 1990. (page 22).
- [110] S. Herbold, A. Trautsch, and J. Grabowski. Global vs. local models for cross-project defect prediction: A replication study. *Empirical Software Engineering*, 22(4):1866–1902, 2017. (pages 33, 51, 52, 72, 75).
- [111] S. Herbold, A. Trautsch, and J. Grabowski. A comparative study to benchmark cross-project defect prediction approaches. *IEEE Transactions on Software Engineering*, 44(9):811–833, 2018. (pages 51, 52).
- [112] H. Huijgens and R. Van Solingen. A replicated study on correlating agile team velocity measured in function and story points. In *International Workshop on Emerging Trends in Software Metrics (WETSoM)*, pages 30–36, 2014. (pages 51, 52).

- [113] A. Idri and A. Zahi. Software cost estimation by classical and fuzzy analogy for web hypermedia applications: A replicated study. In *IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 207–213, 2013. (pages 51, 52).
- [114] J. Itkonen and M. V. Mäntylä. Are test cases needed? replicated comparison between exploratory and test-case-based software testing. *Empirical Software Engineering*, 19(2):303–342, 2014. (pages 51, 52, 72, 76, 77).
- [115] M. A. Javed and U. Zdun. The supportive effect of traceability links in architecture-level software understanding: Two controlled experiments. In *Working IEEE/IFIP Conference on Software Architecture (WICSA)*, pages 215–224, 2014. (pages 51, 52).
- [116] A. Jedlitschka and M. Ciolkowski. Towards evidence in software engineering. In *Proceedings. 2004 International Symposium on Empirical Software Engineering, 2004. ISESE'04.*, pages 261–270. IEEE, 2004. (page 63).
- [117] A. Jedlitschka, M. Ciolkowski, and Pfahl D. *Guide to Advanced Empirical Software Engineering*, chapter Reporting experiments in Software Engineering, pages 201–228. Springer Verlag, 2008. Eds. Forrest Shull and Janice Singer and Dag I.K. Sjøberg. (page 23).
- [118] A. Jedlitschka, M. Ciolkowski, and D. Pfahl. Reporting experiments in software engineering. In *Guide to advanced empirical software engineering*, pages 201–228. Springer, 2008. (pages 4, 71, 72, 74, 75, 78, 79, 135, 147).
- [119] A. Jedlitschka and D. Pfahl. Reporting guidelines for controlled experiments in software engineering. In *Empirical Software Engineering, 2005. 2005 International Symposium on*, pages 10–pp. IEEE, 2005. (page 60).
- [120] L. Jiang, K. M. Carley, and A. Eberlein. Assessing team performance from a socio-technical congruence perspective. In *Proceedings of the International Conference on Software and System Process*, pages 160–169. IEEE Press, 2012. (page 50).
- [121] J. Jung, K. Hoefig, D. Domis, A. Jedlitschka, and M. Hiller. Experimental comparison of two safety analysis methods and its replication. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 223–232, 2013. (pages 51, 52, 72, 76).

- [122] N. Juristo and O. S. Gómez. Replication of software engineering experiments. In *Empirical software engineering and verification*, pages 60–88. Springer, 2010. (page 22).
- [123] N. Juristo and O. 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. (pages 26, 85, 115).
- [124] N. Juristo and A. M. Moreno. *Basics of software engineering experimentation*. Kluwer Academic Publishers, 2001. (pages 23, 24, 25, 60).
- [125] N. Juristo and A. M. Moreno. *Basics of software engineering experimentation*. Springer Science & Business Media, 2013. (pages 4, 71, 72, 74, 75, 78, 79, 105, 135, 147).
- [126] N. Juristo and S. 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. (page 115).
- [127] N. Juristo and S. Vegas. Using differences among replications of software engineering experiments to gain knowledge. In *Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement*, pages 356–366. IEEE Computer Society, 2009. (page 8).
- [128] N. Juristo and S. Vegas. The role of non-exact replications in software engineering experiments. *Empirical Software Engineering*, 16(3):295–324, 2011. (page 8).
- [129] N. Juristo, S. Vegas, M. Solari, S. Abrahao, and I. 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. (page 115).
- [130] N. Juristo, S. Vegas, M. Solari, S. Abrahão, and I. Ramos. A process for managing interaction between experimenters to get useful similar replications. *Information and Software Technology*, 55(2):215–225, 2013. (page 115).
- [131] N. Juristo, S. Vegas, M. Solari, S. Abrahão, and I. Ramos. A process for managing interaction between experimenters to get useful similar replications. *Information and Software Technology*, 55(2):215–225, 2013. (pages 51, 52).

- [132] M. Kalinowski, M. Felderer, T. Conte, R. O. Spínola, R. Prikladnicki, D. Winkler, D. M. Fernández, and S. Wagner. Preventing incomplete/hidden requirements: Reflections on survey data from austria and brazil. In *International Conference on Software Quality. The Future of Systems and Software Development (SWQD)*, pages 63–78, 2016. (pages 51, 52).
- [133] M. Kalinowski, R. O. Spínola, T. Conte, R. Prikladnicki, D. M. Fernández, and S. Wagner. Towards building knowledge on causes of critical requirements engineering problems. In *International Conference on Software Engineering and Knowledge Engineering (SEKE)*, pages 1–6, 2015. (pages 51, 52).
- [134] C. Khatwani, X. Jin, N. Niu, A. Koshoffer, L. Newman, and J. Savolainen. Advancing viewpoint merging in requirements engineering: a theoretical replication and explanatory study. *Requirements Engineering*, 22(3):317–338, 2017. (pages 51, 52).
- [135] B. Kitchenham. The role of replications in empirical software engineering—a word of warning. *Empirical Software Engineering*, 13(2):219–221, 2008. (pages 9, 61, 64).
- [136] B. Kitchenham, T. Dyba, and M. Jorgensen. Evidence-based software engineering. In *Proceedings of the 26th international conference on software engineering*, pages 273–281. IEEE Computer Society, 2004. (page 64).
- [137] B. Kitchenham, L. Madeyski, and P. Brereton. Problems with statistical practice in human-centric software engineering experiments. In *Proceedings of the Evaluation and Assessment on Software Engineering*, pages 134–143. 2019. (page 58).
- [138] B. Kitchenham, S. L. Pfleeger, D. C. Hoaglin., K. E. Emam, and J. Rosenberg. Preliminary Guidelines for Empirical Research in Software Engineering. *IEEE Transactions on Software Engineering*, 28(8):721–734, August 2002. (pages 4, 23).
- [139] T. Kosar, S. Gaberc, J. C. Carver, and M. Mernik. Program comprehension of domain-specific and general-purpose languages: replication of a family of experiments using integrated development environments. *Empirical Software Engineering*, 23(5):2734–2763, 2018. (pages 51, 52, 72, 75, 76, 77).
- [140] J. L. Krein, L. Prechelt, N. Juristo, A. Nanthaamornphong, J. C. Carver, S. Vegas, C. D. Knutson, K. D. Seppi, and D. L. Eggett. A multi-site joint replication of a

- design patterns experiment using moderator variables to generalize across contexts. *IEEE Transactions on Software Engineering*, 42(4):302–321, 2016. (pages 32, 51, 52).
- [141] H. Larsson, E. Lindqvist, and R. Torkar. Outliers and replication in software engineering. In *Software Engineering Conference (APSEC), 2014 21st Asia-Pacific*, volume 1, pages 207–214. IEEE, 2014. (pages 61, 70).
- [142] K. Lauenroth, E. Kamsties, and O. Hehlert. Do words make a difference? an empirical study on the impact of taxonomies on the classification of requirements. In *International Requirements Engineering Conference (RE)*, pages 273–282, 2017. (pages 51, 52).
- [143] V. Lenarduzzi, I. Lunesu, M. Matta, and D. Taibi. Functional size measures and effort estimation in agile development: A replicated study. In *International Conference on Agile Software and Systems Development (XP)*, pages 105–116, 2015. (pages 51, 52).
- [144] T. C. Lethbridge, S. E. Sim, and J. Singer. Studying software engineers: Data collection techniques for software field studies. *Empirical software engineering*, 10(3):311–341, 2005. (page 111).
- [145] R. M. Lindsay and A. S. C. Ehrenberg. The design of replicated studies. *The American Statistician*, 47(3):217–228, 1993. (pages 4, 8, 30, 33, 34).
- [146] J. Lung, J. Aranda, S. Easterbrook, and G. Wilson. On the difficulty of replicating human subjects studies in software engineering. In *Proceedings of ICSE'08*, pages 191–200. IEEE, 2008. (pages 59, 64, 66).
- [147] F. Madrid, S. Carvajal, M. C. Florido, E. Morillo, et al. Fitoextracción asistida por ramnolípidos de suelos contaminados con cu. *Solos e Água: fontes (esgotáveis) de vida e de desenvolvimento*, 2016. (page 123).
- [148] F. Madrid, M. Rubio-Bellido, J. Villaverde, M. Tejada, and E. Morillo. Natural attenuation of fluorene and pyrene in contaminated soils and assisted with hydroxypropyl- β -cyclodextrin. effect of co-contamination. *Science of the Total Environment*, 571:42–49, 2016. (page 124).
- [149] P. B. Mandeville. Tema 28: Diseños experimentales. *Ciencia UANL*, 15(57):151–155, 2012. (pages 104, 127).

- [150] A. Marcolino, E. Oliveira Jr., I. Gimenes, and T. Conte. Towards validating complexity-based metrics for software product line architectures. In *Brazilian Symposium on Software Components, Architectures and Reuse (SBCARs)*, pages 69–79, 2013. (pages 51, 52).
- [151] E. Marrese-Taylor and Y. Matsuo. Replication issues in syntax-based aspect extraction for opinion mining. In *Student Research Workshop at the Conference of the European Chapter of the Association for Computational Linguistics (EACL/SRW)*, pages 23–32, 2017. (pages 51, 52).
- [152] A. K. Massey, P. N. Otto, and A. I. Antón. Evaluating legal implementation readiness decision-making. *IEEE Transactions on Software Engineering*, 41(6):545–564, 2015. (pages 51, 52).
- [153] G. Mauša and T. Galinac Grbac. Co-evolutionary multi-population genetic programming for classification in software defect prediction: An empirical case study. *Applied Soft Computing Journal*, 55:331–351, 2017. (pages 51, 52).
- [154] W. Mehmood, N. Shah, M. J. Khan, M. Memon, and M. Ikramullah. Fine-granular model merge solution for model-based version control system. *Malaysian Journal of Computer Science*, 29(3):225–246, 2016. (page 51).
- [155] R. M. D. Mello, R. F. Oliveira, and A. F. Garcia. On the influence of human factors for identifying code smells: A multi-trial empirical study. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, pages 68–77, 2017. (pages 51, 52, 72, 76, 77).
- [156] A. I. Molina, M. A. Redondo, M. Ortega, and C. Lacave. Evaluating a graphical notation for modeling collaborative learning activities: A family of experiments. *Science of Computer Programming*, 88:54–81, 2014. (pages 51, 52).
- [157] M. Mondal, M. S. Rahman, C. K. Roy, and K. A. Schneider. Is cloned code really stable? *Empirical Software Engineering*, 23(2):693–770, 2018. (pages 51, 52, 72, 75).
- [158] D. C. Montgomery. *Diseño y análisis de experimentos*. Limusa Wiley, 2005. (pages 104, 127).
- [159] A. Moraes, W. L. Andrade, and P. D. L. Machado. A family of test selection criteria for timed input-output symbolic transition system models. *Science of Computer Programming*, 126:52–72, 2016. (pages 51, 52).

- [160] J. M. Morales, E. Navarro, P. Sánchez, and D. Alonso. A family of experiments to evaluate the understandability of TRiStar and i* for modeling teleo-reactive systems. *Journal of Systems and Software*, 114:82–100, 2016. (pages 51, 52).
- [161] M. Márquez Esteban. Extracción de cobre por plantas de mostaza parda (brassica juncea) durante su desarrollo vegetativo, en un suelo contaminado artificialmente y efecto de la aplicación de ramnolípido. 2018. (Trabajo Fin de Grado). Universidad de Sevilla, Sevilla. (page 123).
- [162] M. Nassif and M. P. Robillard. Revisiting turnover-induced knowledge loss in software projects. In *IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 261–272, 2017. (pages 51, 52).
- [163] D. Navarro. Learning statistics with R: A tutorial for psychology students and other beginners (version 0.6), 2018. (page 28).
- [164] D. Navarro and D. Foxcroft. Learning statistics with jamovi: A tutorial for psychology students and other beginners (version 0.70), 2018. (page 28).
- [165] S. Nielebock, D. Krolkowski, J. Krüger, T. Leich, and F. Ortmeier. Commenting source code: is it worth it for small programming tasks? *Empirical Software Engineering*, 2018. (pages 51, 52, 72, 75).
- [166] N. Niu, A. Koshoffer, L. Newman, C. Khatwani, C. Samarasinghe, and J. Savolainen. Advancing repeated research in requirements engineering: A theoretical replication of viewpoint merging. In *International Requirements Engineering Conference (RE)*, pages 186–195, 2016. (pages 51, 52).
- [167] J. F. S. Ouriques, E. G. Cartaxo, and P. D. L. Machado. Test case prioritization techniques for model-based testing: a replicated study. *Software Quality Journal*, 26(4):1451–1482, 2018. (pages 51, 52).
- [168] Y. M. Pacheco, S. López, B. Bermúdez, R. Abia, J. Villar, and F. J. G. Muriana. A meal rich in oleic acid beneficially modulates postprandial sicam-1 and svcam-1 in normotensive and hypertensive hypertriglyceridemic subjects. *The Journal of nutritional biochemistry*, 19(3):200–205, 2008. (page 126).
- [169] J. I. Panach Navarrete, O. Dieste, B. Marin, S. Espana, S. Vegas, O. Pastor, and N. Juristo. Evaluating model-driven development claims with respect to quality: A family of experiments. *IEEE Transactions on Software Engineering*, 2018. (pages 51, 52).

- [170] A. Panichella, F. M. Kifetew, and P. Tonella. A large scale empirical comparison of state-of-the-art search-based test case generators. *Information and Software Technology*, 104:236–256, 2018. (pages 51, 52).
- [171] J. A. Parejo, A. B. Sánchez, S. Segura, A. Ruiz-Cortés, R. E. Lopez-Herrejon, and A. Egyed. Multi-objective test case prioritization in highly configurable systems: A case study. *Journal of Systems and Software*, 122:287–310, 2016. (page 130).
- [172] P. Peachock, N. Iovino, and B. Sharif. Investigating eye movements in natural language and c++ source code - a replication experiment. In *International Conference on Augmented Cognition. Neurocognition and Machine Learning. (AC)*, pages 206–218, 2017. (pages 51, 52).
- [173] S. Pearson, J. Campos, R. Just, G. Fraser, R. Abreu, M.D. Ernst, D. Pang, and B. Keller. Evaluating and improving fault localization. In *International Conference on Software Engineering (ICSE)*, pages 609–620, 2017. (pages 49, 51, 52, 54).
- [174] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007. (page 11).
- [175] B. Penzenstadler. Sustainability analysis and ease of learning in artifact-based requirements engineering: The newest member of the family of studies (it’s a girl!). *Information and Software Technology*, 95:130–146, 2018. (pages 51, 52).
- [176] B. Penzenstadler, J. Eckhardt, and D. M. Fernández. Two replication studies for evaluating artefact models in re: Results and lessons learnt. In *International Workshop on Replication in Empirical Software Engineering (RESER)*, pages 66–75, 2013. (pages 51, 52).
- [177] K. Petersen and C. Wohlin. Context in industrial software engineering research. In *2009 3rd International Symposium on Empirical Software Engineering and Measurement*, pages 401–404. IEEE, 2009. (page 63).
- [178] J. A. Pow-Sang. Evaluating and comparing perceptions between undergraduate students and practitioners in controlled experiments for requirements prioritization. *Advances in Intelligent Systems and Computing*, 405:189–199, 2016. (pages 51, 52).
- [179] C. Quesada-López and M. Jenkins. An empirical validation of function point structure and applicability: A replication study. In *Ibero-American Conference on Software Engineering (CIBSE)*, pages 418–431, 2015. (pages 51, 52).

- [180] C. Quesada-López, D. Madrigal, and M. Jenkins. An empirical evaluation of automated function points. In *Ibero-American Conference on Software Engineering (CIBSE)*, pages 214–228, 2016. (pages 51, 52).
- [181] K. Quille and S. Bergin. Programming: Predicting student success early in cs1. a re-validation and replication study. In *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, pages 15–20, 2018. (pages 51, 52).
- [182] R. Ramac, B. Turhan, I. Karac, N. Juristo, and V. Mandic. Lessons learned from a partial replication of an experiment in the context of a software engineering course. In *International Scientific Conference on Industrial Systems (IS)*, pages 198–203, 2017. (pages 22, 51).
- [183] G. Raura, E. R. Fonseca C, J. W. Castro, T. Gualotuña, C. Rebeca Mejía, T. Mónica Santillán, C. Pons, and O. Dieste. Gender gap in computing: A preliminary empirical study. In *Avances en Ingeniería de Software a Nivel Iberoamericano, CIBSE 2018*, pages 57–70, 2018. (pages 51, 52).
- [184] R. Ré, R. M. Meloca, D. N. R. Junior, M. A. da Cruz Ismael, and G. C. Silva. An empirical study for evaluating the performance of multi-cloud apis. *Future Generation Computer Systems*, 79:726–738, 2018. (pages 51, 52).
- [185] G. Reggio, F. Ricca, G. Scanniello, F.D. Cerbo, and G. Doderio. On the comprehension of workflows modeled with a precise style: results from a family of controlled experiments. *Software and Systems Modeling*, 14(4):1481–1504, 2015. (pages 51, 52).
- [186] D. Reimanis, C. Izurieta, R. Luhr, L. Xiao, Y. Cai, and G. Rudy. A replication case study to measure the architectural quality of a commercial system. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2014. (pages 51, 52, 72, 74, 76).
- [187] R. P. Reyes Chicango. *Evidencias de la existencia de errores estadísticos en experimentos de ingeniería del software: cuatro estudios empíricos*. PhD thesis, ETSI Informatica, 2018. (page 21).
- [188] R. P. Reyes Chicango, O. Dieste, N. Juristo, et al. Statistical errors in software engineering experiments: A preliminary literature review. In *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE)*, pages 1195–1206. IEEE, 2018. (page 58).

- [189] M. Riaz, J. King, J. Slankas, L. Williams, F. Massacci, C. Quesada-López, and M. Jenkins. Identifying the implied: Findings from three differentiated replications on the use of security requirements templates. *Empirical Software Engineering*, 22(4):2127–2178, 2017. (pages 51, 52, 71, 72, 76, 77).
- [190] M. Ribeiro, P. Borba, and C. Kästner. Feature maintenance with emergent interfaces. In *International Conference on Software Engineering (ICSE)*, pages 989–1000, 2014. (pages 49, 51, 52).
- [191] F. Ricca, G. Scanniello, M. Torchiano, G. Reggio, and E. Astesiano. Assessing the effect of screen mockups on the comprehension of functional requirements. *ACM Transactions on Software Engineering and Methodology*, 24(1):1:1–1:38, 2014. (pages 51, 52).
- [192] F. Ricca, M. Torchiano, M. Leotta, A. Tiso, G. Guerrini, and G. Reggio. On the impact of state-based model-driven development on maintainability: a family of experiments using unimod. *Empirical Software Engineering*, 23(3):1743–1790, 2018. (pages 51, 52, 72, 73, 76).
- [193] N. Rios, R. O. Spínola, M. Mendonça, and C. Seaman. The most common causes and effects of technical debt: First results from a global family of industrial surveys. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2018. (pages 51, 52).
- [194] A. Rodrigues Santos, I. do Carmo Machado, E. Santana de Almeida, J. Siegmund, and S. Apel. Comparing the influence of using feature-oriented programming and conditional compilation on comprehending feature-oriented software. *Empirical Software Engineering*, 2018. (pages 51, 52, 72, 74, 75).
- [195] G. Rong, H. Zhang, B. Liu, Q. Shan, and D. Shao. A replicated experiment for evaluating the effectiveness of pairing practice in psp education. *Journal of Systems and Software*, 136:139–152, 2018. (pages 51, 52).
- [196] P. Runeson and M. Höst. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*, 14(2):131, 2009. (pages 4, 23, 71, 72, 75, 78, 79, 104, 135, 147).
- [197] P. Runeson, M. Höst, A. Rainer, and B. Regnell. Case study research in software engineering. In *Guidelines and examples*. Wiley Online Library, 2012. (pages xv, 104, 109).

- [198] P. Runeson, A. Stefik, and A. Andrews. Variation factors in the design and analysis of replicated controlled experiments: Three (dis)similar studies on inspections versus unit testing. *Empirical Software Engineering*, 19(6):1781–1808, 2014. (pages 51, 52, 72, 76, 77, 78).
- [199] N. Salleh, E. Mendes, and J. Grundy. Investigating the effects of personality traits on pair programming in a higher education setting through a family of experiments. *Empirical Software Engineering*, 19(3):714–752, 2014. (pages 51, 52, 72, 76).
- [200] A. B. Sánchez, S. Segura, and A. Ruiz-Cortés. A comparison of test case prioritization criteria for software product lines. In *2014 IEEE Seventh International Conference on Software Testing, Verification and Validation*, pages 41–50. IEEE, 2014. (page 130).
- [201] A. Santos, O. S. Gómez, and N. Juristo. Analyzing families of experiments in se: a systematic mapping study. *IEEE Transactions on Software Engineering*, 2018. (pages 20, 25, 70, 73, 74, 78).
- [202] G. Santos, T. Conte, G. Travassos, R. Prikładnicki, A.R. Rocha, N. Franco, and K.C. Weber. Towards successful software process improvement initiatives: Experiences from the battlefield. In *Americas Conference on Information Systems (AMCIS)*, 2015. (pages 51, 52).
- [203] J. A. M. Santos and M. G. De Mendonça. Exploring decision drivers on god class detection in three controlled experiments. In *ACM Symposium on Applied Computing (SAC)*, pages 1472–1479, 2015. (pages 51, 52).
- [204] G. Scanniello and U. Erra. Distributed modeling of use case diagrams with a method based on think-pair-square: Results from two controlled experiments. *Journal of Visual Languages and Computing*, 25(4):494–517, 2014. (pages 51, 52).
- [205] G. Scanniello, C. Gravino, M. Genero, J. A. Cruz-Lemus, and G. Tortora. On the impact of uml analysis models on source-code comprehensibility and modifiability. *ACM Transactions on Software Engineering and Methodology*, 23(2):13:1–13:26, 2014. (pages 49, 51, 52).
- [206] G. Scanniello, C. Gravino, M. Risi, G. Tortora, and G. Doderio. Documenting design-pattern instances: A family of experiments on source-code comprehensibility. *ACM Transactions on Software Engineering and Methodology*, 24(3):14:1–14:35, 2015. (pages 51, 52).

- [207] G. Scanniello, C. Gravino, G. Tortora, M. Genero, M. Risi, J. Cruz-Lemus, and G. Dodero. Studying the effect of uml-based models on source-code comprehensibility: Results from a long-term investigation. In *International Conference on Product-Focused Software Process Improvement*, volume 9459, pages 311–327, 2015. (pages 51, 52).
- [208] G. Scanniello, A. Marcus, and D. Pascale. Link analysis algorithms for static concept location: an empirical assessment. *Empirical Software Engineering*, 20(6):1666–1720, 2015. (pages 51, 52, 71, 72, 76).
- [209] G. Scanniello, M. Risi, P. Tramontana, and S. Romano. Fixing faults in c and java source code: Abbreviated vs. full-word identifier names. *ACM Transactions on Software Engineering and Methodology*, 26(2):6:1–6:43, 2017. (pages 51, 52).
- [210] G. Scanniello, M. Staron, H. Burden, and R. Heldal. On the effect of using sysml requirement diagrams to comprehend requirements: Results from two controlled experiments. In *Evaluation and Assessment in Software Engineering (EASE)*, page 49, 2014. (pages 51, 52).
- [211] C. Seaman. Qualitative methods in empirical studies of software engineering. *IEEE Transactions on software engineering*, 25(4):557–572, 1999. (page 63).
- [212] M. Shahin, P. Liang, and Z. Li. Do architectural design decisions improve the understanding of software architecture? two controlled experiments. In *IEEE International Conference on Program Comprehension (ICPC)*, pages 3–13, 2014. (pages 51, 52).
- [213] S. Shamshiri, J. M. Rojas, J. P. Galeotti, N. Walkinshaw, and G. Fraser. How do automatically generated unit tests influence software maintenance? In *International Conference on Software Testing, Verification and Validation (ICST)*, pages 250–261, 2018. (pages 51, 52).
- [214] M. Shepperd. Replication studies considered harmful. In *2018 IEEE/ACM 40th International Conference on Software Engineering: New Ideas and Emerging Technologies Results (ICSE-NIER)*, pages 73–76. IEEE, 2018. (page 63).
- [215] M. Shepperd, N. Ajenka, and S. Counsell. The role and value of replication in empirical software engineering results. *Information and Software Technology*, 99:120–132, 2018. (page 70).

- [216] M. Shepperd, C. Mair, and M. Jørgensen. An experimental evaluation of a de-biasing intervention for professional software developers. In *ACM Symposium on Applied Computing (SAC)*, pages 1510–1517, 2018. (pages 51, 52).
- [217] F. Shull, V. Basili, J. C. Carver, J. C. Maldonado, G. H. Travassos, M. Mendonça, and S. Fabbri. Replicating software engineering experiments: addressing the tacit knowledge problem. In *Proceedings of EMSE'2002*, pages 7–16. IEEE, 2002. (pages 59, 62, 70).
- [218] F. Shull, J. C. Carver, S. Vegas, and N. Juristo. The role of replications in empirical software engineering. *Empirical Software Engineering*, 13(2):211–218, 2008. (pages 28, 30, 33, 34, 60).
- [219] C. A. Siebra and M. A. B. Mello. The importance of replications in software engineering - a case study in defect prediction. In *ACM Research in Adaptive and Convergent Systems (RACS)*, pages 376–381, 2015. (pages 51, 52).
- [220] J. Siegmund, N. Siegmund, and S. Apel. Views on internal and external validity in empirical software engineering. In *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on*, volume 1, pages 9–19. IEEE, 2015. (pages 29, 30, 70).
- [221] N. Siegmund, S. Sobernig, and S. Apel. Attributed variability models: Outside the comfort zone. In *ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*, pages 268–278, 2017. (pages 51, 52).
- [222] P. Singh and B. Suri. Quality metrics for conceptual model of data warehouse. In *CSI International Conference on Software Engineering (CONSEG 2013)*, pages 98–103, 2013. (pages 51, 52).
- [223] F. L. Siqueira. Comparing the comprehensibility of requirements models: An experiment replication. *Information and Software Technology*, 96:1–13, 2018. (pages 51, 52).
- [224] D. I. K. Sjøberg, T. Dyba, and M. Jorgensen. The future of empirical methods in software engineering research. In *2007 Future of Software Engineering*, pages 358–378. IEEE Computer Society, 2007. (page 63).
- [225] D. I. K. Sjøberg, J. E. Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N. K. Liborg, and A. C. Rekdal. A survey of controlled experiments in softwa-

- re engineering. *Software Engineering, IEEE Transactions on*, 31(9):733–753, 2005. (pages 51, 59, 62, 70).
- [226] M. Solari and S. Matalonga. A controlled experiment to explore potentially undetectable defects for testing techniques. In *International Conference on Software Engineering and Knowledge Engineering (SEKE)*, pages 106–109, 2014. (pages 51, 52).
- [227] M. Solari, S. Vegas, and N. Juristo. Content and structure of laboratory packages for software engineering experiments. *Information and Software Technology*, 2017. In Press. (pages 58, 59, 60, 61).
- [228] E. Souza, A. Moreira, J. Araújo, S. Abrahão, E. Insfran, and D.S.D. Silveira. Comparing business value modeling methods: A family of experiments. *Information and Software Technology*, 104:179–193, 2018. (pages 51, 52).
- [229] C. Sun, Y. Zai, and H. Liu. Evaluating and comparing fault-based testing strategies for general boolean specifications: A series of experiments. *Computer Journal*, 58(5):1199–1213, 2013. (pages 51, 52).
- [230] P. Sun and K.T. Stolee. Exploring crowd consistency in a mechanical turk survey. In *International Workshop on CrowdSourcing in Software Engineering (CSI-SE)*, pages 8–14, 2016. (pages 51, 52).
- [231] W. Sun, M. I. Aguirre-Urreta, and G. M. Marakas. Effectiveness of pair and solo programming methods: A survey and an analytical approach. In *Americas Conference on Information Systems (AMCIS)*, 2015. (pages 51, 52).
- [232] D. Taibi, A. Janes, and V. Lenarduzzi. How developers perceive smells in source code: A replicated study. *Information and Software Technology*, 92:223–235, 2017. (pages 51, 52).
- [233] A. Teran-Somohano, O. Dayibas, L. Yilmaz, and A. Smith. Toward a model-driven engineering framework for reproducible simulation experiment lifecycle management. In *Winter Simulation Conference*, volume 2015-January, pages 2726–2737, 2015. (pages 51, 52).
- [234] M. Torchiano, G. Scanniello, F. Ricca, G. Reggio, and M. Leotta. Do uml object diagrams affect design comprehensibility? results from a family of four controlled experiments. *Journal of Visual Languages and Computing*, 41:10–21, 2017. (pages 51, 52).

- [235] A. Durán Toro and B. 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. (page 87).
- [236] P. Tramontana, M. Risi, and G. Scanniello. Studying abbreviated vs. full-word identifier names when dealing with faults: An external replication. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2014. (pages 51, 52).
- [237] F. M. Uyaguari. *Impacto de la instrumentación en los experimentos de Ingeniería del Software: Aplicación a experimentos de Test-Driven Development*. PhD thesis, Universidad Politécnica de Madrid, 2017. (page 21).
- [238] N. J. van Eck and L. Waltman. VOSviewer version 1.6.11. Centre for Science and Technology Studies (CWTS) of Leiden University. <https://www.vosviewer.com/>, 2019. (page 47).
- [239] S. Vegas, O. Dieste, and N. Juristo. Difficulties in running experiments in the software industry: Experiences from the trenches. In *International Workshop on Conducting Empirical Studies in Industry (CESI)*, pages 3–9, 2015. (pages 51, 52).
- [240] S. Vegas, N. Juristo, A. Moreno, Ma. Solari, and P. Letelier. Analysis of the influence of communication between researchers on experiment replication. In *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering*, pages 28–37. ACM, 2006. (pages 9, 60).
- [241] A. Vetrò, W. Böhm, and M. Torchiano. On the benefits and barriers when adopting software modelling and model driven techniques - an external, differentiated replication. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, volume 2015-November, pages 168–171, 2015. (pages 51, 52).
- [242] L. Vidacs and M. Pinzger. Co-evolution analysis of production and test code by learning association rules of changes. In *IEEE International Workshop on Machine Learning Techniques for Software Quality Evaluation*, pages 31–36, 2018. (pages 51, 52).
- [243] R. H. Von Alan, S. T March, J. Park, and S. Ram. Design science in information systems research. *MIS quarterly*, 28(1):75–105, 2004. (page 11).

- [244] C. Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Evaluation and Assessment in Software Engineering (EASE)*, pages 38:1–38:10, 2014. (page 47).
- [245] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Publishing Company, Incorporated, 2012. (pages 4, 23, 24, 25, 27, 28, 60, 71, 72, 74, 75, 78, 79, 89, 91, 117, 118, 135, 147).
- [246] F. Wu. Empirical tests of scale-free characteristic in open source software: A replicated case study. *Advanced Materials Research*, 622:1933–1936, 2013. (pages 41, 51, 52).
- [247] H. Yang. Improved software cost estimation method based on cocomo model and linear regression. *Advanced Materials Research*, 989-994:1497–1500, 2014. (pages 41, 51, 52).
- [248] S. Young, T. Abdou, and A. Bener. A replication study: Just-in-time defect prediction with ensemble learning. In *International Conference on Software Engineering (ICSE)*, pages 42–47, 2018. (pages 51, 52).
- [249] N. Yusop, M. Kamalrudin, M. M. Yusof, and S. Sidek. Eliciting security requirements for mobile apps: A replication study. *Journal of Theoretical and Applied Information Technology*, 95(15):3613–3622, 2017. (pages 51, 52).