A model–based approach for specifying changes in replicationsof controlled experiments using templates and patterns

Supplemental Material: Instantiation of the Proposed Template in the multiple-case studies and References of the Related Work

Abstract—This supplemental material contains the detailed instantiation of the proposal template in the multiple—case studies whose details have not **b**een previously pu**b**lished elsewhere, and the 25 references of the related work section not included in the main article.

1 Proposal template in the multiple-case studies

This section contains the instantiation of the proposal template in the multiple-case studies.

Table 1: Template instantiation in SoftEng-Case, Mind family, Mind#2 Replication

Replication	***Mind#2*** Internal replication based on ***Mind#1*** original experiment
Goal of experiment	To study whether mindfulness practice (cause) improves productivity in conceptual modelling (effect) in software engineering students (population).
Description	A group of students from the Software Engineering Degree at the University of Seville (sample) attended 10-minute mindfulness sessions for 4 weeks, 4 days per week (experimental group treatment), while a second group of students attended a public speaking workshop as a placebo (control group treatment). The performance of both groups was compared in terms of quality (similarity to the reference solution) and productivity (similarity in percentage per unit time) (metrics).
Site and Date	The base experiment was carried out in *E.T.S. Ingeniería Informática, University of Seville* in the first half of the 2013-2014 academic year and this replication, in *E.T.S. Ingeniería Informática, University of Seville* in in the first half of the 2014-2015 academic year.
Purpose	Confirm results Overcome some limitations of the baseline experiment

Table 2: Template instantiation in SoftEng-Case, Mind family, Mind#2 Replication

Change *1*	**Increased treatment duration** **Originally**, for 4 weeks Mindfulness was practiced 4 days a week in 10-minute sessions **In replication** In replication the sessions were 12 minutes long and for 6 weeks **in order to** make more evident the benefits of Mindfulness
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases construct validity since increasing the duration of treatment better reflects the effect of Mindfulness practice. The change increases internal validity since increasing the duration of treatment strengthens its effect over that of other possible factors.
Change *#2*	**Random assignment of subjects to groups** **Originally**, the assignment of subjects to treatment was not randomized **In replication** it becomes random **in order to** remedy threats to the internal validity of quasi-experiments
Modified Dimension	**Protocol**, specifically experimental design
Threat to validity	The change increases the validity of the conclusion, since it improves the power of the applicable statistical tests. The change increases internal validity, since it eliminates assignment bias.
Change *#3*	**Null treatment of the control group** **Originally**, an public speaking workshop was given to the control group as a placebo **In replication** the oratory workshop took place after the experiment **in order to** avoid a possible effect of such a workshop on the measurements of dependent variables
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases internal validity because it eliminates the possibility that the placebo could have an effect on the results.

Table 4: Template instantiation in SoftEng-Case, Req family, Q-2009 Replication

Replication	***Q-2009*** Internal replication based on ***Q-2007*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Description	The effectiveness of the analysts is analysed, using the interview as a requirements analysis technique. In the interviews, the experimenter acted as a client answering the questions of the experimental subjects (analysts) about two possible problems,one in the known domain and one in the unknown domain. To measure the effectiveness of consolidation, after some time, the analyst presents in writing what he/she remembers from the interview and the number of problem elements mentioned by the analyst is counted.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2007 and this replication, in *Polytechnic University of Madrid* in 2009
Purpose	Confirm results
Change *1*	**Analysis of effectiveness** **Originally**, analysts' effectiveness in interview sessions is analysed **In replication** effectiveness is not analysed **because of** the high cost of transcribing and analyzing all interviews
Modified Dimension	**Operationalization**, specifically the effect
Threat to validity	The change decreases the construct validity because the dependent variable effectiveness is not considered.
Change *2*	Retention capacity analysis **Originally**, the retention capacity is analyzed **In replication** retention capacity is not analysed **because of** the high cost of transcribing and analyzing all interviews
Modified Dimension	**Operationalization**, specifically the effect
Threat to validity	The change decreases the construct validity since the retention capacity is not analysed.
Change *3*	**Development experience** **Originally**, no account is taken of development experience **In replication** experience in development is considered to calculate the independent variable experience
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases the construct validity since there is one more variable in the construct.
Change *4*	**Language of the interviews** **Originally**, interviews are conducted in Spanish **In replication** interviews are conducted in English **because of** English was a requirement of the master to which the students belonged
Modified Dimension	**Protocol**, specifically experimental material
Threat to validity	The change increases the internal validity since the results are independent of the language in which the interview is conducted.
Change *5*	**Unavailability for interviews** **Originally**, a person responds in interviews **In replication** these person is changed **because** is not available.
Modified Dimension	**Stakeholder**, specifically the *monitor*
Threat to validity	The change increases internal validity because it eliminates experimenter bias.

Table 5: Template instantiation in SoftEng-Case, Req family, Q-2011 Replication

Replication	***Q-2011*** Internal replication based on ***Q-2009*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2009 and this replication, in *Polytechnic University of Madrid* in 2011
Purpose	Generalise results
Change *1*	**Group interviews** **Originally**, interviews between subjects (analysts) and experimenter are individual **In replication** interviews are in groups **because of** the cost and effort involved in conducting individual interviews and the experimenter's fatigue
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases the internal validity since all subjects receive the same information.
Change *2*	**Experience determination** **Originally**, experience in requirements analysis is considered **In replication** experience is determined by years of experience and the skill the subject claims to have
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases construct validity because more variables are taken into account. The change decreases the conclusion validity as the procedure becomes tedious.
Change *3*	Duration of interviews **Originally**, the duration of the interviews is 30 min. **In replication** the duration of the interviews is 60 min . **because of** the interview is in group
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity as it increases the duration of the interviews in order to better understand the requirements.
Change *4*	**Time elapsed before submission of information Originally**, The subject (analyst) has 7 days to present in writing the information gathered in the interview. **In replication** the written presentation is immediately after the interview. **in order to** avoid loss of information
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity as the information is collected after the interview so that it is not forgotten.

Change *5*	Submission time **Originally**, the time elapsed between the interview and the written presentation of the information collected is not measured **In replication** the time elapsed between the interview and the written presentation of the information is set at 120 min. **because of** the written presentation is immediately after the interview
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity since it is analysed whether time influences the information collected.
Change *6*	**Unavailability for interviews** **Originally**, a person responds in interviews **In replication** these person is changed **because** is not available.
Modified Dimension	**Stakeholder**, specifically the *monitor*
Threat to validity	The change increases internal validity because it eliminates experimenter bias.

Table 6: Template instantiation in SoftEng-Case, Req family, Q-2012 Replication

Replication	***Q-2012*** External replication based on ***Q-2011*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2011 and this replication, in *Polytechnic University of Madrid* in 2012
Purpose	Confirm results
Change *1*	**Professional subjects** **Originally**, the subjects are Master's students **In replication** the subjects are professionals **because of** replication is performed at the International Working Conference on Requirements Engineering
Modified Dimension	**Population**, specifically the experience
Threat to validity	The change increases the external validity since the effect on professionals is analysed.
Change *2*	**development skill** **Originally**, subjects have little or no development experience **In replication** the subjects are professionals with experience in development **because of** replication is performed at the International Working Conference on Requirements Engineering
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases external validity since the population is extended.
Change *3*	Reduction of consolidation time **Originally**, the duration of the interviews is 120 min. **In replication** the duration of the interviews is 30 min . **because of** time constraints
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity because the influence of time is analysed.
Change *4*	**Elimination of the training period** **Originally**, the experiment is carried out at the end of the course, i.e. after the training period **In replication** no training period **because of** replication is performed at the International Working Conference on Requirements Engineering
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity because the influence of training and experience is analysed.

Table 7: Template instantiation in SoftEng-Case, Req family, E-2012A Replication

Replication	***E-2012A*** Internal replication based on ***Q-2012*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2012 and this replication, in *Polytechnic University of Madrid* in 2012
Purpose	Generalise results
Change *1*	**Problem domain** **Originally**, knowledge is defined as familiarity through subjective assessment **In replication** knowledge is defined as an independent variable with two levels: known and unknown problem **because of** in the experimental population (post-graduate students) it is possible to know whether or not they know a certain domain of the problem
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases the construct validity because a variable is added which improves the construct.
Change *2*	*Repeated measures design* **Originally**, the interviews to know the requirements are carried out on two different days, to avoid fatigue in the experimenter **In replication** the design is changed to a design of repeated measurements (within-subjects) **because of** this design does not require a large number of subjects
Modified Dimension	**Protocol**, specifically the experimental design
Threat to validity	The change increases the conclusion validity because of the applicable tests.
Change *3*	**Interview type** **Originally**, interviews between subjects (analysts) and experimenters are in groups **In replication** interviews are individual . **because of** there are two experimenters (responders) with two languages
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases the internal validity since it allows to analyse the differences between the two monitors regarding the information provided. The change decreases the conclusion validity because, although there are two monitors responding, the process can become tedious for these monitors due to the increased number of individual interviews.
Change *4*	**Blocking by language** **Originally**, there are no blocking variables **In replication** there is a blocking variable per language **because of** subjects who use their mother tongue will be more effective than subjects who use a second language
Modified Dimension	**Protocol**, specifically the experimental design
Threat to validity	The change increases internal validity since the language is blocked from influencing the results.

Change *5*	**Lock by monitor** **Originally**, there are no blocking variables **In replication** there is one blocking variable per experimenter (respondent) **because** experimental subjects conduct the interview in their own language.
Modified Dimension	**Protocol**, specifically the experimental design
Threat to validity	The change increases internal validity since the monitor is blocked from influencing the results.
Change *6*	**Number of monitors** **Originally**, there is a experimenter (respondent) **In replication** there are two experimenters (respondents) **In order to** alleviate the effects of fatigue and learning of the experimenter (respondents)
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity by avoiding monitor fatigue and learning.
Change *7*	**Number of problems** **Originally**, there is the same problem (experimental object) for all subjects **In replication** there are two problems **because of** groups are made due to blocking variables
Modified Dimension	**Protocol**, specifically the experimental design
Threat to validity	The change increases internal validity and differences can be analysed.
Change *8*	**Duration of interviews** **Originally**, the duration of the interviews is 60 min. **In replication** the duration of the interviews is 30 min. **because** the interview is individual
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity because it reduces monitor fatigue.
Change *9*	**Consolidation time** **Originally**, the time elapsed between the interview and the written presentation is 30 min. **In replication** the time elapsed between the interview and the written presentation is 90 min. **because** the recommended duration of 90 minutes
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity because more time is available and the information collected can be more accurate.
Change *10*	**Originally**, the difficulty of the problem is not measured **In replication** the difficulty variable indicates the difficulty of the problem **because** there are two problems
Modified Dimension	**Problem difficulty** **Operationalization**, specifically the cause
Threat to validity	The change increases the construct validity as there is a new variable to better capture the construct.

Table 8: Template instantiation in SoftEng-Case, Req family, E-2012B Replication

Replication	***E-2012B*** Internal replication based on ***E-2012A*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2012 and this replication, in *Polytechnic University of Madrid* in 2012
Purpose	Confirm results
Change *1*	**New problems** **Originally**, two problem domains are used in the experiment, one known domain (DC) and the other unknown domain (DD) **In replication** the problem domains used in the experiment have been modified, but one is still a known domain (DC) and the other is an unknown domain (DD)
Modified Dimension	**Protocol**, specifically the experimental material
Threat to validity	The change does not affect the validity
Change *2*	**Order of problems** **Originally**, first the known domain problem is performed and then the unknown domain problem. **In replication** the order of the problems is swapped
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases the internal validity since it allows to analyse whether the order affects the results. The change increases the validity of the conclusion as it allows to analyse and adjust the effect of the difference between tasks on the experimental results.
Change *3*	**Timing of the experiment** **Originally**, the experiment was carried out at the beginning of the course; **In replication** the experiment is carried out after the subjects have received training in Requirements Engineering
Modified Dimension	Context, specifically the Moment of Realisation
Threat to validity	The change decreases internal validity since replication takes place at the end of the course and may influence the results.

Table 9: Template instantiation in SoftEng-Case, Req family, E-2013 Replication

Replication	***E-2013*** Internal replication based on ***E-2012B*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2012 and this replication, in *Polytechnic University of Madrid* in 2013
Purpose	Generalise results
Change *1*	**Inter-subject design** **Originally**, the design is of repeated measurements **In replication** the design is between-subjects **In order to** avoid the learning effect
Modified Dimension	**Protocol**, specifically experimental design
Threat to validity	The change increases internal validity because it prevents bias due to the order in which treatments are administered.
Commentary	In the between-subjects design, each subject undergoes only one treatment to avoid the learning effect.
Change *2*	**Previous training** **Originally**, no short training (warming up) before the course **In replication** the brief training (warming up) is 1 week **Because** we want to study the effect of training.
Modified Dimension	**Operationalization**, specifically teh cause
Threat to validity	The change increases construct validity because the effect of training is better reflected. The change increases the internal validity because the effect of training is analysed against other factors.

Table 10: Template instantiation in SoftEng-Case, Req family, E-2014 Replication

Replication	***E-2014*** Internal replication based on ***E-2013*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2013 and this replication, in *Polytechnic University of Madrid* in 2014
Purpose	Generalise results
Change *1*	**Only one monitor** **Originally**, in the interviews, there are two respondents **In replication** there is only one responder **because of** the unavailability of one of the respondents
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases internal validity since, with only one monitor, there is no difference in the information received by the subjects. The change decreases the conclusion validity since, by having only one monitor, the process becomes tedious for the monitor.
Change *2*	**Previous training** **Originally**, the brief training (warming up) is 1 week **In replication** the brief training (warming up) is 6 week **In order to** explore the warming up effect
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases construct validity since the effect of training is better reflected. The change increases the internal validity since the effect of training is analysed against other factors.

Table 11: Template instantiation in SoftEng-Case, Req family, E-2015 Replication

Replication	***E-2015*** Internal replication based on ***E-2014*** original experiment
Goal of experiment	Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* in 2014 and this replication, in *Polytechnic University of Madrid* in 2015
Purpose	Generalise results
Change *1*	**Previous training** **Originally**, the brief training (warming up) is 1 week **In replication** the brief training (warming up) is 2 week **In order to** explore the warming up effect
Modified Dimension	**Operationalization**,specifically the cause
Threat to validity	The change increases construct validity since the effect of training is better reflected. The change increases the internal validity since the effect of training is analysed against other factors.

Table 12: Template instantiation in SoftEng-Case, Code family, VV-UPM1 Replication

Replication	***VV-UPM1*** Internal replication based on ***VV-UPM*** original experiment
Goal of experiment	To evaluate the effectiveness of three code verification and validation techniques
Description	Subjects evaluate the three techniques by applying each technique to each of the C programs containing the errors to be detected. Previously, the subjects receive training in the use of each of the error detection techniques.
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* and this replication, in *Polytechnic University of Madrid*
Purpose	Generalise results
Change *1*	**Visibility** **Originally**, the visibility of the fault is not analysed **In replication** the influence of the visibility of the fault is analysed **in order to** draw new conclusions
Modified Dimension	**Operationalization**, specifically the effect
Threat to validity	The change increases construct validity since the effect for the evaluation of each technique is better reflected.
Commentary	Laboratory package developed by Kamsties and Lott is used
Change *2*	**Two versions of each programme** **Originally**, the influence of the programme is not analysed **In replication** two versions of each program are implemented and is a new factor **because** the programs are not very long and therefore the errors are masked from each other
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases the construct validity since the influence of the programme version is analysed.
Change *3*	**Increase in the number of failures** **Originally**, three of the fault types appear only once while the other three types appear twice **In replication** all types of faults are duplicated **because** there are two versions of each program
Modified Dimension	**Protocol**, specifically the effect
Threat to validity	The change increases the internal validity, since by increasing the number of faults to be detected, the comparison of techniques is facilitated.
Change *4*	**Test cases** **Originally**, subjects generate their test cases to detect code failures **In replication** first, the subjects apply the technique to generate the test cases and then execute the test cases provided to them to detect program failures **in order to** check whether the visibility of faults influences their detection
Modified Dimension	**Protocol**, specifically the experimental material
Threat to validity	The change increases internal validity because generating and executing test cases reinforces the effect of the technique.

Change *5*	**Discarding a programme** **Originally**, four programs are used **In replication** three programs are used, one is discarded **in order to** balance the design
Modified Dimension	**Protocol**, specifically the experimental design
Threat to validity	The change decreases internal validity because one of the programmes on which failures were detected is removed.
Change *6*	**Techniques applied by each subject** **Originally**, each subject applies a technique **In replication** each subject applies the three techniques **because** the design is changed
Modified Dimension	**Protocol**, specifically the experimental design
Threat to validity	The change increases internal validity since the applications of each technique are increased and comparability is facilitated.

Table 13: Template instantiation in SoftEng-Case, Code family, VV-UPV Replication

Replication	***VV-UPV*** External replication based on ***VV-UPM*** original experiment
Goal of experiment	To evaluate the effectiveness of three code verification and validation techniques
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* and this replication, in *Polytechnic University of Valencia*
Purpose	Generalise results
Change *1*	**Discarding one techniques** **Originally**, the three verification and validation techniques are used: code reading, equivalence partitioning and branch testing **In replication** the code reading technique is omitted **because of** time constraints
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change decreases the construct validity since one of the levels of the techniques factor is removed.
Comment	The baseline experiment are UPM replications treated as one
Change *2*	**Duration of sessions** **Originally**, the duration of the 3 sessions is 4h. each, i.e. the time is unlimited **In replication** the duration of each of the 3 sessions is 2h. **because of** time constraints
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change decreases the internal validity since the time constraint may influence the results.
Change *3*	**Training period** **Originally**, subjects receive three four-hour training sessions to learn how to apply the techniques **In replication** the training consists of two two-hour tutorials **because** he subjects are already familiar with the techniques
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change does not affect validity since the subjects are already familiar with the techniques and the training is not operationalised.
Change *4*	**Training in each technique** **Originally**,the training in the use of the techniques is before the experiment is executed **In replication** Each tutorial is carried out before the application of the technique, in the first 2 sessions; i. e., the training is interspersed with the operation of the experiment **because** he subjects are already familiar with the techniques
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change does not affect validity since subjects were already familiar with the techniques.

Change *5*	**Application of the techniques** **Originally**, subjects apply a technique to a program in each session **In replication** subjects apply the same technique to different programs in each session **because of** time constraints
Modified Dimension	**Protocol**, specifically the experimental design
Threat to validity	The change decreases the internal validity since the three techniques are not analysed and compared.
Change *6*	**Test cases in separate session** **Originally**, The subjects execute test cases with the application of the technique; that is to say in each session **In replication** Subjects run test cases for one of the programs they have tested in a separate session, i.e. in session 3 **because of** time constraints
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases the internal validity since the effect of conducting the test cases in a separate session is analysed.

Table 14: Template instantiation in SoftEng-Case, Code family, VV-Uds Replication

Replication	***VV-Uds*** External replication based on ***VV-UPM*** original experiment
Goal of experiment	To evaluate the effectiveness of three code verification and validation techniques
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* and this replication, in *Polytechnic University of Sevilla*
Purpose	Generalise results
Change *1*	**Duration of sessions** **Originally**, the duration of the 3 sessions is 4h. each, i.e. the time is unlimited **In replication** the duration of each of the 3 sessions is 2h. **because of** time constraints
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change decreases the internal validity since the time restriction may affect the results.
Change *2*	**Test cases in separate session** **Originally**, the subjects execute test cases with the application of the technique; i. e. in each session **In replication** the subjects execute test cases for one of the programs they have tested in a later session, i.e. in session 4 **because of** time constraints
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change decreases the internal validity since the time restriction may affect the results.
Change *3*	**Subjects work in pairs** **Originally**, subjects work individually **In replication** subjects work in pairs **because** there are not enough computers
Modified Dimension	**Context**, specifically, form of work
Threat to validity	The change increases internal validity because the effect of working together can be analysed.
Change *4*	**Reduction of training period** **Originally**, subjects receive three four-hour training sessions to learn how to apply the techniques **In replication** the training consists of two two-hour tutorials **because** he subjects are already familiar with the techniques
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change does not affect validity since the subjects are already familiar with the techniques and the training is not operationalised.
Change *5*	**Training in each technique** **Originally**,the training in the use of the techniques is before the experiment is executed **In replication** each tutorial is conducted before the application of the technique in each of the three sessions in which each technique is examined; i.e., the training is interspersed with the operation of the experiment **because** he subjects are already familiar with the techniques
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change does not affect validity since subjects were already familiar with the techniques.

Table 15: Template instantiation in SoftEng-Case, Code family, VV-ORT Replication

Replication	***VV-ORT*** External replication based on ***VV-UPM*** original experiment
Goal of experiment	To evaluate the effectiveness of three code verification and validation techniques
Site and Date	The base experiment was carried out in *Polytechnic University of Madrid* and this replication, in *University ORT Uruguay*
Purpose	Generalise results
Change *1*	**Discarding one techniques** **Originally**, the three techniques of verification and validation are used: code reading, equivalence partitioning and branch testing **In replication** the code reading technique is omitted **because of** time constraints
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change decreases the construct validity since one of the levels of the techniques factor is removed.
Change *2*	**Discard a programme** **Originally**, three program codes are used **In replication** one of the programs is discarded **because of** time constraints
Modified Dimension	**Protocol**, specifically experimental design
Threat to validity	The change decreases internal validity because one of the programmes on which failures were detected is removed.
Change *3*	**Duration of sessions** **Originally**, the experiment is carried out in three sessions each of four hours **In replication** the experiment is executed in a single session **because of** time constraints
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change decreases the internal validity since the time restriction may affect the results.
Change *4*	**Implementation of techniques by programmes** **Originally**, subjects apply a different technique to evaluate a program in each of the three sessions **In replication** the subjects apply the two techniques to the two programs in a single session **because of** time constraints
Modified Dimension	**Protocol**, specifically experimental design
Threat to validity	The change decreases the internal validity since one of the techniques and one of the programmes is not used.

Table 16: Template instantiation in Agrobiology-Case, Soil family, Soil-2018 Replication

Replication	***Soil-2018*** Internal replication based on ***E-2016*** original experiment
Goal of experiment	To evaluate the effect of a bio-surfactant on the assisted phytoremediation of contaminated soil
Description	Whether the addition of the biosurfactant JBR-425 to two types of copper (Cu) contaminated soils (cause) reduces the Cu concentration in these soils is analysed by determining the Cu extracted (effect) by barley (*Hordeum vulgare*) and mustard (*Brassica juncea*) plants germinated in these soils.
Site and Date	The base experiment was carried out in *Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS–CSIC)* in *October 2016* and this replication, in *Escuela Técnica Superior de Ingeniería Agronómica (ETSIA)-University of Seville* in *March 2018*
Purpose	Generalise results Overcome some limitations of the baseline experiment
Change *1*	**Growing medium** **Originally**, the experiment was carried out in a cultivation chamber. **In replication**, was carried out in a greenhouse **In order to** simulate natural conditions
Modified Dimension	**Context**, specifically, the Growing medium of experimental unit
Threat to validity	The change increases the external validity since it allows to generalise the results carrying out the replication in conditions closer to natural ones
Change *2*	**Plant types** **Originally**, two plants were used: *Hordeum vulgare* and *Brassica juncea* **In replication**, only *Brassica juncea* was used **Because** in the original experiment it was demonstrated that only *Brassica juncea* was a metal accumulator plant
Modified Dimension	**Protocol**, specifically measuring instruments
Threat to validity	The change increases construct validity due the effect can be measured in the mustard (*Brassica juncea*), i.e. the extracted Cu
Change *3*	**Soil types** **Originally**, there were two types of soil: Coria (pH=7.8) and Constantina (pH=5.5) **In replication**, only Constantina soil was used **Because** it was demonstrated that in the soil of Coria the metal was strongly adsorbed and the phytoextraction did not affect the biomass production
Modified Dimension	**Population**, specifically the type of soil
Threat to validity	The change increases the construct validity since it ensures that the metal can be extracted from the soil. The extracted Cu is the effect to be measured.
Change *4*	**Copper dose reduction** **Originally**, Copper (Cu) doses were 0, 500 and 1000 mg kg^{-1} **In replication**, Cu doses were adjusted to 0, 125, 250 and 500 mg kg^{-1} **Because of** Cu doses of 1000 mg kg^{-1} was toxic to the plant
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases construct validity because the Cu dose is adjusted to non-toxic levels for the plant The change increases internal validity because the effect is analysed at valid Cu levels

Change *5*	**Form of application of copper** **Originally**, Cu was applied as Copper Nitrate **In replication**, Cu was applied as Copper Sulfate **Because of** is more accessible and the concentrations applied do not affect the plant
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change does not affect validity since the reactive is equivalent
Commentary	The change affects how Cu is applied (treatment)
Change *6*	**Soil aging time** textbfOriginally, the soil aging time (from the time Cu is applied until the plant is sown) is 45 days **In replication**, soil aging time is 15 days **Because of** time constraints and so that Cu is not so much retained
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases construct validity because the plant extracts the metal more easily and the effect is better measured. The change decreases the validity of the external because it reduces the generalizability of results to soils contaminated for a longer period of time.
Change *7*	**Increased biomass obtained** **Originally**, there were 6 treatments corresponding to the 3 levels of Cu and with/without *surfactant* (to facilitate Cu extraction). There were 2 soils and 2 types of plants. This represents 24 experimental units (3x2x2x2). For each experimental unit, 3 pots were prepared. In total there are 72 pots (3x2x2x2x3) **In replication**, there were 8 treatments corresponding to 4 level of Cu and with/without *surfactant*. There were 1 soil and 1 type of plant. This represents 8 experimental units. For each experimental unit, 4 pots were prepared and placed on a tray. In total there were 32 pots (4x2x4) distributed in 8 trays with 4 pots each. The trays are distributed completely randomly. This is repeated 3 times. The experimental unit was the tray **Because** by cultivating 4 pots in each tray sufficient biomass can be obtained
Modified Dimension	**Protocol**, specifically experimental design
Threat to validity	The change increases the validity of the conclusion because increasing the number of subjects (n) improves the power of the statistical tests, reducing the probability of obtaining a false negative (type II error).
Change *8*	**To obtain biomass at fruiting stage** **Originally**, the biomass is collected when the plants have between 2 and 3 true leaves **In replication**, the plants are rinsed when they have between 2 and 3 real leaves and 4 plants are left by pot. The biomass is collected when the plants reach the fructification stage **In order to** avoid competition between plants, let the plants complete their vegetative cycle and thus obtain more biomass
Modified Dimension	**Operationalization**, specifically the measurement procedure
Threat to validity	The change increases the construct validity since more biomass is obtained for further analysis.
Change *9*	**Increase in soil volume** **Originally**, the pots are 300 ml tube type **In replication**, the pots are 500 ml bucket type. **Because** a greater volume of soil allows for greater root development and greater biomass production
Modified Dimension	**Protocol**, specifically experimental material
Threat to validity	The change increases the construct validity since more biomass is obtained for further analysis.

Table 17: Template instantiation in Agrobiology-Case, Soil family, Soil-2019 Replication

Replication	***Soil-2019*** Internal replication based on ***Soil-2016*** original experiment
Description	To evaluate the effect of a bio-surfactant on the assisted phytoremediation of contaminated soil
Site and Date	The base experiment was carried out in *Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS–CSIC)* in *October 2016* and this replication, in *ETSIA-University of Seville* in *March 2019*
Purpose	Extend results
Change *1*	**Growing medium** **Originally**, the experiment was carried out in a cultivation chamber **In replication**, was carried out in a greenhouse **In order to** simulate natural conditions
Modified Dimension	**Context**, specifically the environment
Threat to validity	The change increases the external validity since it allows to generalise the results performing the replication in conditions closer to the natural ones
Change *2*	**Types of plants** **Originally**, two types of plants were used: *Hordeum vulgare* L. and *Brassica juncea* L. **In replication**, only *Brassica juncea* L. was used **Because** in the original experiment it was demonstrated that only *Brassica juncea* L. was a metal accumulator plant
Modified Dimension	**Protocol**, specifically measuring instruments
Threat to validity	The change increases the construct validity because in *Brassica juncea* it is possible to measure the effect, i.e. the extracted Cu
Commentary	By using only one type of plant, it does not affect the results. It is not operationalised.
Change *3*	**Types of naturally contaminated soils** **Originally**, there were two types of soil: Coria (pH=7.8) and Constantina (pH=5.5) **In replication**, there were three types of soil: Miraflores-1 (pH=x, Pb=158 y Zn=125, Cu=36) and Miraflores-2 (pH=y, Pb=375 Zn=192 Cu=206) and Lebrija (not contaminated by metals) **In order to** experiment with naturally contaminated soils. Miraflores soils are urban gardens with natural contamination and Lebrija soil was used as control
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases internal validity because soils of different characteristics are compared. The change increases conclusion validity because the construct is better reflected.
Commentary	The soils of Miraflores are urban gardens with natural contamination and the soil of Lebrija was used as a control.
Change *4*	**Artificial Cu dose is cancelled** **Originally**, Copper (Cu) doses were 0, 500 and 1000 mg kg^{-1} applied as Copper Nitrate **In replication**, the soils are not artificially contaminated with Cu **Because of** these soils are urban orchard-gardens with natural pollution (Cu levels 36 and 206 mg kg^{-1})
Modified Dimension	**Operationalization**, specifically the cause
Threat to validity	The change increases the internal validity since Cu levels are those present in the soil and are not toxic to the plant.

Change *5*	**Increase in soil types** **Originally**, there were 6 treatments corresponding to the 3 levels of Cu and with/without *surfactant* (to facilitate Cu extraction). There were 2 soils and 2 types of plants. This represents 24 experimental units (3x2x2x2). For each experimental unit, 3 pots were prepared. In total there are 72 pots (3x2x2x2x3) **In replication**, there were 2 treatments corresponding to with/without *surfactant*. There were 3 soils and 1 type of plant. This represents 6 experimental units. Each experimental unit was repeated 4 times (2x3x4 pots) forming a block. This is repeated 3 times to have 3 blocks. In total there are 72 pots (2x3x4x3). Within each block, pots are randomly distributed. The experimental unit is the pot. **Because** plant types and soils have been adjusted
Modified Dimension	**Protocol**, specifically experimental design
Threat to validity	The change increases internal validity because it allows the effect on different soil types to be compared.
Change *6*	**Biomass collection in fruiting** **Originally**, the biomass is collected when the plants have between 2 and 3 true leaves **In replication**, the plants are rinsed when they have between 2 and 3 real leaves and only 1 plant is left by pot. The biomass is collected when the plants reach the fructification stage. **In order to** avoid competition between plants, let the plants complete their vegetative cycle and thus obtain more biomass
Modified Dimension	**Operationalization**, specifically the measurement procedure
Threat to validity	The change increases construct validity due to more biomass being obtained for further analysis
Change *7*	**Increasing the volume of soil** **Originally**, the pots are 300 ml tube type **In replication**, the pots are 500 ml bucket type. **Because** a greater volume of soil allows for greater root development and greater biomass production
Modified Dimension	**Protocol**, specifically experimental material
Threat to validity	The change increases construct validity due to more biomass being obtained for further analysis

Table 18: Template instantiation in Agrobiology-Case, Harvest family, Harvesting System-2017 Replication

Replication	***Harvesting System-2017*** Internal replication based on ***Harvesting System-2016*** original experiment
Goal of experiment	Evaluate the effects of two harvesting methods and two conservation methods, designed for small producers, on the quality of the stored olive fruit and the quality of the extracted oil.
Description	Four experimental factors were studied: Variety, recollection, conservation, and storage time. Three varieties were studied: Arbequina, Picual and Verdial. Two recollection methods were compared: (1) a prototype of a manual inverted umbrella and (2) traditional harvest with nets. Two conservation methods were compared for each type of recollection: (a) cold storage, 5 grades and (b) ambient temperature. The fruit were stored up to 14 days, while at day 0, 4, 8, 14 fruit was inspected and oil extracted for phyisco-chemical analysis. The harvesting took place in an olive groove in Bollullos par del Condado (Huelva, Spain). The fruit and oil extraction and evaluation was carried out in Instituto de la Grasa (CSIC) Sevilla.
Site and Date	The base experiment was carried out in *Instituto de la Grasa (CSIC) Sevilla* in *October 2016* and this replication, in *Instituto de la Grasa (CSIC) Sevilla* in *2017*
Purpose	Confirm results
Change *1*	**Different climatic conditions** **Originally**, the weather conditions are those of 2016 **In replication**, the climatic conditions are different as they correspond to 2017 **In order to** analyze data corresponding to different campaigns

Table 19: Template instantiation in Agrobiology-Case, Olive family, Olive-Des Replication

Replication	***Olive-Des*** Internal replication based on ***Olive-2015*** original experiment
Goal of experiment	Extraction phenolic compounds (EPC) from virgin olive oil with green solvents (Deep Eutectic Solvents DES)
Description	The aim is to test different non-toxic solvents to extract phenolic compounds from virgin olive oil.
Site and Date	The base experiment was carried out in *Instituto de la Grasa (CSIC) Sevilla* in *2015* and this replication, in *Instituto de la Grasa (CSIC) Sevilla* in *2015*
Purpose	Extend results
Change *1*	**Solvent extraction 1** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and Glycerol (1:2)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *2*	**Solvent extraction 2** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and Lactic acid (1:2)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *3*	**Solvent extraction 3** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and Urea (1:2)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *4*	**Solvent extraction 4** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and Sucrose (1:1)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *5*	**Solvent extraction 5** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and Sucrose (4:1)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *6*	**Solvent extraction 6** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and 1,4-Butanediol (1:5)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*

Change *7*	**Solvent extraction 7** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and Xylitol (2:1)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *8*	**Solvent extraction 8** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and 1,2-Propanediol (1:1)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *9*	**Solvent extraction 9** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride and Malonic acid (1:1)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *10*	**Solvent extraction 10** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *Choline chloride, Urea and Glycerol (1:1:1)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*
Change *11*	**Solvent extraction 11** **Originally**, the extraction is done with a solution 80% (v/v) *methanol and water* **In replication**, the extraction is done with a solution *D-(−)-Fructose D-(+)-Glucose and Sucrose (1:1:1)* **In order to** analyze non-toxic alternatives for the extraction of *phenolic compounds*

Table 20: Template instantiation in Agrobiology-Case, Diet family, Diet-Hiper Replication

Replication	***Diet-Hiper*** Internal replication based on ***Diet-Normo*** original experiment
Goal of experiment	To analyse the effect of a diet rich in oleic acid in *hypertriglyceridemic* subjects.
Description	Subjects are fed a diet rich in oleic acid and are regularly tested for, among other things, cholesterol levels.
Site and Date	The base experiment was carried out in *Instituto de la Grasa (CSIC) Sevilla* in *2006* and this replication, in *Instituto de la Grasa (CSIC) Sevilla* in *2006*
Purpose	Extend results
Change *1*	**Hypertensive subjects** **Originally**, Subjects have blood pressure levels within the limits considered normal **In replication**, Subjects are *hypertensive* **In order to** study the effect of the diet rich in oleic acid on *hypertriglyceridemic* subjects who are also *hypertensive*

Table 21: Template instantiation in Comp-Case, Testing family, Test-NF Replication

Replication	***Test-NF*** Internal replication based on ***Test-F*** original experiment
Description	Prioritization of test case execution to accelerate the detection of faults in highly configurable systems
Site and Date	The base experiment was carried out in *ETSII-University of Seville* in *2015* and this replication, in *ETSII-University of Seville* in *2015*
Purpose	Extend results
Change *1*	**Non-functional objectives** **Originally**, *objective functions* are *functional* **In replication**, *objective functions* are *non–functional* **In order to** compare differences in favour of multi–objective prioritization over mono–objective prioritization using non–functional objectives
Modified Dimension	**Operationalization**, specifically the effect
Threat to validity	The change does not affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected

Table 22: Template instantiation in Comp-Case, Testing family, Test–F&NF Replication

Replication	***Test–F&NF*** Internal replication based on ***Test–F*** original experiment
Description of experiment	Prioritization of test case execution to accelerate the detection of faults in highly configurable systems
Site and Date	The base experiment was carried out in *ETSII-University of Seville* in *2015* and this replication, in *ETSII-University of Seville* in *2015*
Purpose	Extend results
Change *1*	**Functional and non-functional objectives** **Originally**, *objective functions* are *functional* **In replication**, *objective functions* combine *functional* and *non–functional* **In order to** analyse whether *multi–objective* prioritization using *functional* and *non–functional* objectives outperform prioritization driven by a single objective, either functional or non–functional
Modified Dimension	**Operationalization**, specifically the effect
Threat to validity	The change does not affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected

Table 23: Test–FvsNF replication specification using the template

Replication	***Test–FvsNF*** Internal replication based on ***Test–F*** original experiment
Description of experiment	Prioritization of test case execution to accelerate the detection of faults in highly configurable systems
Site and Date	The base experiment was carried out in *ETSII-University of Seville* in *2015* and this replication, in *ETSII-University of Seville* in *2015*
Purpose	Extend results
Change *1*	**Comparison of objectives** **Originally**, *objective functions* are *functional* **In replication**, *objective functions* combine *functional* and *non–functional* **In order to** analyze the domain of *non-functional objectives* over *functional objectives*, especially when these are combined in a *multi-objective* perspective
Modified Dimension	**Operationalization**, specifically the effect
Threat to validity	The change does not affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected

Table 24: Template instantiation in Comp-Case, SPL family, SPL–Pr&Com Replication

Replication	***SPL–Pr&Com*** Internal replication based on ***SPL–Pr*** original experiment
Goal of experiment	Comparison of test case prioritization criteria for Software Product Lines (SPL)
Site and Date	The base experiment was carried out in *ETSII-University of Seville* in *2014* and this replication, in *ETSII-University of Seville* in *2014*
Purpose	Generalise results
Change *1*	**Set of tests** **Originally**, only a test suite was generated **In replication**, for each model, 2-wise test suite was generated **In order to** obtain a list of products covering all the possible pairs of features on each model
Modified Dimension	**Protocol**, specifically the guides
Threat to validity	The change increases construct validity by increasing the number of tests.
Change *2*	**Generation of the test suite** **Originally**, a test suite was randomly generated using SPLAR tool **In replication**, test suite was randomly generated using SPLCAT tool **Because** SPLCAT increase the fault detection rate and thus it is considered as an extra prioritization approach in our comparison
Modified Dimension	**Protocol**, specifically the experimental material
Threat to validity	the change does not affect validity because because only the tool used changes

2 References of the related work

This section contains the 25 references of the related work section not included in the main article. [1], [2], [3], [4], [5] [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25].

References

[1] Albayrak, O., and Carver, J. C. Investigation of individual factors impacting the effectiveness of requirements inspections: A replicated experiment. *Empirical Software Engineering 19*, 1 (2014), 241–266.

[2] Apa, C., Dieste, O., et al. Effectiveness for detecting faults within and outside the scope of testing techniques: an independent replication. *Empirical Software Engineering 19*, 2 (2014), 378–417.

[3] Assar, S., Borg, M., and Pfahl, D. Using text clustering to predict defect resolution time: a conceptual replication and an evaluation of prediction accuracy. *Empirical Software Engineering 21*, 4 (2016), 1437–1475.

[4] Cavalcanti, G., Accioly, P., and Borba, P. Assessing semistructured merge in version control systems: A replicated experiment. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (2015), pp. 267–276.

[5] Fernández, D. M., Wagner, S., Kalinowski, M., Felderer, M., Mafra, P., Vetrò, A., Conte, T., Christiansson, M. T., Greer, D., Lassenius, C., Männistö, T., Nayabi, M., Oivo, M., Penzenstadler, B., Pfahl, D., Prikladnicki, R., Ruhe, G., Schekelmann, A., Sen, S., Spínola, R. O., Tuzcu, A., de la Vara, J. L., and Wieringa, R. Naming the pain in requirements engineering: Contemporary problems, causes, and effects in practice. *Empirical Software Engineering 22*, 5 (2017), 2298–2338.

[6] Fernández-Sáez, A. M., Genero, M., Caivano, D., and Chaudron, M. R. V. Does the level of detail of uml diagrams affect the maintainability of source code?: a family of experiments. *Empirical Software Engineering 21*, 1 (2016), 212–259.

[7] Fucci, D., Scanniello, G., Romano, S., Shepperd, M., Sigweni, B., Uyaguari, F., Turhan, B., Juristo, N., and Oivo, M. An external replication on the effects of test-driven development using a multi-site blind analysis approach. In *Proceedings of EMSE'16* (2016), ACM.

[8] Fucci, D., and Turhan, B. A replicated experiment on the effectiveness of test-first development. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (2013), pp. 103–112.

[9] Fucci, D., and Turhan, B. On the role of tests in test-driven development: A differentiated and partial replication. *Empirical Software Engineering 19*, 2 (2014), 277–302.

[10] Guerrouj, L., Di Penta, M., Guéhéneuc, Y.-G., and Antoniol, G. An experimental investigation on the effects of context on source code identifiers splitting and expansion. *Empirical Software Engineering 19*, 6 (2014), 1706–1753.

[11] Gómez, M. N., and Acuña, S. A replicated quasi-experimental study on the influence of personality and team climate in software development. *Empirical Software Engineering 19*, 2 (2014), 343–377.

[12] Herbold, S., Trautsch, A., and Grabowski, J. Global vs. local models for cross-project defect prediction: A replication study. *Empirical Software Engineering 22*, 4 (2017), 1866–1902.

[13] Itkonen, J., and Mäntylä, M. V. Are test cases needed? replicated comparison between exploratory and test-case-based software testing. *Empirical Software Engineering 19*, 2 (2014), 303–342.

[14] Jung, J., Hoefig, K., Domis, D., Jedlitschka, A., and Hiller, M. Experimental comparison of two safety analysis methods and its replication. In *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)* (2013), pp. 223–232.

[15] Kosar, T., Gaberc, S., Carver, J. C., and Mernik, M. Program comprehension of domain-specific and general-purpose languages: replication of a family of experiments using integrated development environments. *Empirical Software Engineering 23*, 5 (2018), 2734–2763.

[16] Mello, R. M. D., Oliveira, R. F., and Garcia, A. F. 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)* (2017), pp. 68–77.

[17] Mondal, M., Rahman, M. S., Roy, C. K., and Schneider, K. A. Is cloned code really stable? *Empirical Software Engineering 23*, 2 (2018), 693–770.

[18] Nielebock, S., Krolikowski, D., Krüger, J., Leich, T., and Ortmeier, F. Commenting source code: is it worth it for small programming tasks? *Empirical Software Engineering* (2018).

[19] Reimanis, D., Izurieta, C., Luhr, R., Xiao, L., Cai, Y., and Rudy, G. A replication case study to measure the architectural quality of a commercial system. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (2014), ACM, p. 31.

[20] Riaz, M., King, J., Slankas, J., Williams, L., Massacci, F., Quesada-López, C., and Jenkins, M. Identifying the implied: Findings from three differentiated replications on the use of security requirements templates. *Empirical Software Engineering 22*, 4 (2017), 2127–2178.

[21] Ricca, F., Torchiano, M., Leotta, M., Tiso, A., Guerrini, G., and Reggio, G. On the impact of state-based model-driven development on maintainability: a family of experiments using unimod. *Empirical Software Engineering 23*, 3 (2018), 1743–1790.

[22] Rodrigues Santos, A., do Carmo Machado, I., Santana de Almeida, E., Siegmund, J., and Apel, S. Comparing the influence of using feature-oriented programming and conditional compilation on comprehending feature-oriented software. *Empirical Software Engineering* (2018).

[23] Runeson, P., Stefik, A., and Andrews, A. 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 (2014), 1781–1808.

[24] Salleh, N., Mendes, E., and Grundy, J. Investigating the effects of personality traits on pair programming in a higher education setting through a family of experiments. *Empirical Software Engineering 19*, 3 (2014), 714–752.

[25] Scanniello, G., Marcus, A., and Pascale, D. Link analysis algorithms for static concept location: an empirical assessment. *Empirical Software Engineering 20*, 6 (2015), 1666–1720.