Supplemental Material: Instantiation of the Proposed Templates in the Multiple Case Study

May, 2021

1 SoftEng–Case, Mind family, Mind#2 Replication

Replication	**Mind#2** internal replication based on Mind#1 original experiment
Original Study	**Goal**: 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 at the ETSII of the US in first half of the 2013-2014 academic year. This replication was carried out at the same site in first half of the 2014-2015 academic year.
Purposes	• Confirm results • Overcome some limitations of the baseline experiment

Change #1	Increased treatment duration (Mind#2)
Description	Originally, for 4 weeks Mindfulness was practiced 4 days a week in 10-minute sessions. In this replication, the sessions were 12 minutes long and for 6 weeks. With the purpose of making the benefits of Mindfulness more evident.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because increasing the duration of treatment better reflects the effect of Mindfulness practice. • This change increases internal validity because increasing the duration of treatment strengthens its effect over that of other possible factors.

Change #2	Random assignment of subjects to groups (Mind#2)
Description	Originally, the assignment of subjects to treatment was not randomized. In this replication, it becomes random. With the purpose of remedying the threats to the internal validity of the quasi-experiments.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases conclusion validity because it improves the power of the applicable statistical tests. • This change increases internal validity because it eliminates assignment bias.

Change #3	Null treatment of the control group (Mind#2)
Description	Originally, the control group received a placebo treatment consisting of a public speaking workshop. In this replication, the control group received no treatment, as the public speaking workshop was postponed until after the second task. With the purpose of mitigating the potential distorting factor of placebo on experimental results.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases internal validity because it eliminates the possibility that the placebo could have an effect on the results.
Comments	The new public speaking workshop is online mainly

2 SoftEng–Case, Mind family, Mind#3 Replication

Replication	**Mind#3** internal replication based on Mind#2 original experiment
Original Study	**Goal**: 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 at the ETSII of the US in first half of the 2014-2015 academic year. This replication was carried out at the same site in first half of the 2015-2016 academic year.
Purposes	• Confirm results

Change #1	Reordering of conceptual modelling exercises (Mind#3)
Description	Originally, the Eramus problem was carried out and after the treatment, the EoDProjects problem was carried out. In this replication, the order of the exercises was swapped. With the purpose of studying whether it influences the results.
Dimension	Protoco, specifically, the guides.
Effects on validity	• This change increases internal validity because it allows to study the effect of the difference between tasks on the experimental results. • This change increases conclusion validity because it allows to analyse and adjust the effect of the difference between tasks on the experimental results.

3 SoftEng–Case, Req family, Q–2009 Replication

Replication	**Q–2009** internal replication based on Q–2007 original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2007 . This replication was carried out at the same site in 2009.
Purposes	• Confirm results

Change #1	Analysis of effectiveness (Q–2009)
Description	Originally, analysts' effectiveness in interview sessions is analyzed. In this replication, effectiveness is not analyzed. With the purpose of avoiding the high cost of transcribing and analyzing all the interviews.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change decreases construct validity because the dependent variable effectiveness is not considered.

Change #2	Retention capacity analysis (Q–2009)
Description	Originally, the retention capacity is analyzed. In this replication, retention capacity is not analyzed. With the purpose of reducing the cost of transcribing and analyzing all the interviews.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change decreases construct validity because the retention capacity is not analyzed.

Change #3	Development experience (Q–2009)
Description	Originally, no account is taken of development experience. In this replication, experience in development is considered to calculate the independent variable experience. With the purpose of -.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because there is one more variable in the construct.

Change #4	Language of the interviews (Q–2009)
Description	Originally, interviews are conducted in Spanish. In this replication, interviews are conducted in English. With the purpose of conducting the interviews in English as it is a requirement of the master's degree to which the students belonged.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases internal validity because the results are independent of the language in which the interview is conducted.

Change #5	Unavailability for interviews (Q–2009)
Description	Originally, a person responds in interviews. In this replication, these person is changed. With the purpose of making this feasible.
Dimension	Experimenter, specifically, the monitor.
Effects on validity	• This change increases internal validity because it eliminates experimenter bias.

4 SoftEng–Case, Req family, Q–2011 Replication

Replication	**Q–2011** internal replication based on Q–2009 original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2009. This replication was carried out at the same site in 2011.
Purposes	• Generalize results

Change #1	Group interviews (Q–2011)
Description	Originally, interviews between subjects (analysts) and experimenter are individual. In this replication, interviews are in groups. With the purpose of reducing the cost and effort involved in conducting individual interviews and experimenter fatigue.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because all subjects receive the same information.

Change #2	Experience determination (Q–2011)
Description	Originally, experience in requirements analysis is considered. In this replication, experience is determined by years of experience and the skill the subject claims to have . With the purpose of -.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because more variables are taken into account. • This change decreases conclusion validity because as the procedure becomes tedious.

Change #3	Duration of interviews (Q–2011)
Description	Originally, the duration of the interviews is 30 min. In this replication, the duration of the interviews is 60 min. With the purpose of group interviews.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because it increases the duration of the interviews in order to better understand the requirements.

Change #4	Time elapsed before submission of information (Q–2011)
Description	Originally, The subject (analyst) has 7 days to present in writing the information gathered in the interview. In this replication, the written presentation is immediately after the interview. With the purpose of avoiding the loss of information.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because the information is collected after the interview so that it is not forgotten.

Change #5	Submission time (Q–2011)
Description	Originally, the time elapsed between the interview and the written presentation of the information collected is not measured. In this replication, the time elapsed between the interview and the written presentation of the information is set at 120 min. With the purpose of the written presentation to be made immediately after the interview.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because it is analyzed whether time influences the information collected.

Change #6	Unavailability for interviews (Q–2011)
Description	Originally, a person responds in interviews. In this replication, these person is changed. With the purpose of making this feasible.
Dimension	Experimenter, specifically, the monitor.
Effects on validity	• This change increases internal validity because it eliminates experimenter bias.

5 SoftEng–Case, Req family, Q–2012 Replication

Replication	**Q–2012** external replication based on Q–2011 original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2011 . This replication was carried out at the same site in 2012.
Purposes	• Confirm results

Change #1	Professional subjects (Q–2012)
Description	Originally, the subjects are Master's students. In this replication, the subjects are professionals. With the purpose of replication at the International Working Conference on Requirements Engineering.
Dimension	Population, specifically, the experience.
Effects on validity	• This change increases external validity because the effect on professionals is analyzed.

Change #2	Development skill (Q–2012)
Description	Originally, subjects have little or no development experience. In this replication, the subjects are professionals with experience in development. With the purpose of replication at the International Working Conference on Requirements Engineering.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases external validity because the population is extended.

Change #3	Reduction of consolidation time (Q–2012)
Description	Originally, the duration of the interviews is 120 min. In this replication, the duration of the interviews is 30 min. With the purpose of complying with the time.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because the influence of time is analyzed.

Change #4	Elimination of the training period (Q–2012)
Description	Originally, the experiment is carried out at the end of the course, i.e. after the training period. In this replication, no training period. With the purpose of replication at the International Working Conference on Requirements Engineering.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because the influence of training and experience is analyzed.

6 SoftEng–Case, Req family, E–2012A Replication

Replication	**E–2012A** internal replication based on Q–2012 original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2012 . This replication was carried out at the same site in 2012.
Purposes	• Confirm results

Change #1	Problem domain (E–2012A)
Description	Originally, knowledge is defined as familiarity through subjective assessment. In this replication, knowledge is defined as an independent variable with two levels: known and unknown problem. With the purpose of finding out in the experimental population (graduate students) whether or not they know a certain problem domain.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because a variable is added which improves the construct.

Change #2	Repeated measures design (E–2012A)
Description	Originally, the interviews to know the requirements are carried out on two different days, to avoid fatigue in the experimenter. In this replication, the design is changed to a design of repeated measurements (within-subjects). With the purpose of having a design that requires fewer subjects.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases conclusion validity because the applicable tests.

Change #3	Interview type (E–2012A)
Description	Originally, interviews between subjects (analysts) and experimenters are in groups. In this replication, interviews are individual. With the purpose of having two experimenters (responders) with two languages.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because it allows to analyse the differences between the two monitors regarding the information provided. • This change decreases 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 (E–2012A)
Description	Originally, there are no blocking variables. In this replication, there is a blocking variable per language. With the purpose of analyzing the influence of using the mother tongue in relation to a second language.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases internal validity because the language is blocked from influencing the results.

Change #5	Lock by monitor (E–2012A)
Description	Originally, there are no blocking variables. In this replication, there is one blocking variable per experimenter (respondent). With the purpose of conducting the interview in the subject's language.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases internal validity because the monitor is blocked from influencing the results.

Change #6	Number of monitors (E–2012A)
Description	Originally, there is a experimenter (respondent). In this replication, there are two experimenters (respondents). With the purpose of alleviating the effects of experimenter fatigue and learning (respondents) .
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because avoiding monitor fatigue and learning.

Change #7	Number of problems (E–2012A)
Description	Originally, all subjects perform the same problem (experimental object). In this replication, there are two problems. With the purpose of having groups of subjects due to the blocking variables.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases internal validity because differences can be analyzed.

Change #8	Duration of interviews (E–2012A)
Description	Originally, the duration of the interviews is 60 min. In this replication, the duration of the interviews is 30. With the purpose of conducting individual interviews.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because monitor fatigue is reduced.

Change #9	Consolidation time (E–2012A)
Description	Originally, the time elapsed between the interview and the written presentation is 30 min. In this replication, the time elapsed between the interview and the written presentation is 90 min. With the purpose of following the recommendations.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases construct validity because more time is available and the information collected can be more accurate.

Change #10	Inter-subject design (E–2012A)
Description	Originally, the difficulty of the problem is not measured. In this replication, the difficulty variable indicates the difficulty of the problem. With the purpose of analyzing the difficulty of the problems .
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because there is a new variable to better capture the construct.

7 SoftEng–Case, Req family, E–2012B Replication

Replication	**E–2012B** internal replication based on E–2012A original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2012 . This replication was carried out at the same site in 2012.
Purposes	• Confirm results

Change #1	New problems (E–2012B)
Description	Originally, two problem domains are used in the experiment, one known domain (DC) and the other unknown domain (DD). In this 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). With the purpose of -.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change does not affect validity because difficulty is similar.

Change #2	Order of problems (E–2012B)
Description	Originally, first the known domain problem is performed and then the unknown domain problem. In this replication, the order of the problems is swapped. With the purpose of -.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because it allows to analyze whether the order affects the results. • This change increases conclusion validity because it allows to analyze and adjust the effect of the difference between tasks on the experimental results.

Change #3	Timing of the experiment (E–2012B)
Description	Originally, the experiment was carried out at the beginning of the course. In this replication, the experiment is carried out after the subjects have received training in Requirements Engineering. With the purpose of -.
Dimension	Context, specifically, the Moment of Realisation.
Effects on validity	• This change dicreases internal validity because replication takes place at the end of the course and may influence the results.

8 SoftEng–Case, Req family, E–2013 Replication

Replication	**E–2013** internal replication based on E–2012B original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2012 . This replication was carried out at the same site in 2013.
Purposes	• Generalize results

Change #1	Inter-subject design (E–2013)
Description	Originally, the design is of repeated measurements. In this replication, the design is between-subjects. With the purpose of avoiding the learning effect.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases internal validity because it prevents bias due to the order in which treatments are administered.
Comments	In the between-subjects design, each subject undergoes only one treatment to avoid the learning effect

Change #2	Previous training (E–2013)
Description	Originally, no short training (warming up) before the course. In this replication, the brief training (warming up) is 1 week. With the purpose of analyzing the effect of training.
Dimension	Operationalization, specifically, the treatments .
Effects on validity	• This change increases construct validity because the effect of training is better reflected. • This change increases internal validity because the effect of training is analyzed against other factors.

9 SoftEng–Case, Req family, E–2014 Replication

Replication	**E–2014** internal replication based on E–2013 original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2013 . This replication was carried out at the same site in 2014.
Purposes	• Generalize results

Change #1	Previous training (E–2014)
Description	Originally, there are two people answering the interviews. In this replication, only one person responds. With the purpose of resolving the unavailability of one of the persons.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change increases internal validity because with only one monitor, there is no difference in the information received by the subjects. • This change decreases conclusion validity because the process becomes tedious for a single monitor.

Change #2	Previous training (E–2014)
Description	Originally, the brief training (warming up) is 1 week. In this replication, the brief training (warming up) is 6 week. With the purpose of analyzing the effect of training.
Dimension	Operationalization, specifically, the treatments .
Effects on validity	• This change increases construct validity because the effect of training is better reflected. • This change increases internal validity because the effect of training is analyzed against other factors.

10 SoftEng–Case, Req family, E–2015 Replication

Replication	**E–2015** internal replication based on E–2014 original experiment
Original Study	**Goal**: Study the influence of the analyst's experience and domain knowledge on the effectiveness of requirements analysis **Description**: The effectiveness of the analysts is analyzed, 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 at Polytechnic University of Madrid in 2014 . This replication was carried out at the same site in 2015.
Purposes	• Generalize results

Change #1	Previous training (E–2015)
Description	Originally, the brief training (warming up) is 1 week. In this replication, the brief training (warming up) is 2 week. With the purpose of analyzing the effect of training.
Dimension	Operationalization, specifically, the treatments .
Effects on validity	• This change increases construct validity because the effect of training is better reflected. • This change increases internal validity because the effect of training is analyzed against other factors.

11 SoftEng–Case, Code family, VV–UPM1 Replication

Replication	**VV–UPM1** internal replication based on VV–UPM original experiment
Original Study	**Goal**: 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 at Polytechnic University of Madrid in -. This replication was carried out at the same site in -.
Purposes	• Generalize results

Change #1	Visibility (VV–UPM1)
Description	Originally, the visibility of the fault is not analyzed. In this replication, the influence of the visibility of the fault is analyzed. With the purpose of drawing new conclusions.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change increases construct validity because the effect for the evaluation of each technique is better reflected.

Change #2	Two versions of each programme (VV–UPM1)
Description	Originally, the influence of the program is not analyzed. In this replication, two versions of each program are implemented and is a new factor. With the purpose of preventing errors from masking each other as the programs are not very long.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because the influence of the program version is analyzed.

Change #3	Increase in the number of failures (VV–UPM1)
Description	Originally, three of the fault types appear only once while the other three types appear twice. In this replication, all types of faults are duplicated. With the purpose of increasing the number of failures.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases internal validity because by increasing the number of faults to be detected, the comparison of techniques is facilitated.

Change #4	Test cases (VV–UPM1)
Description	Originally, subjects generate their test cases to detect code failures. In this 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. With the purpose of testing whether the visibility of faults influences their detection.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases internal validity because generating and executing test cases reinforces the effect of the technique.

Change #5	Discarding a programme (VV–UPM)
Description	Originally, four programs are used. In this replication, three programs are used, one is discarded . With the purpose of balancing the design.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change decreases internal validity because one of the programmes on which failures were detected is removed.

Change #6	Techniques applied by each subject (VV–UPM1)
Description	Originally, each subject applies a technique. In this replication, each subject applies the three techniques. With the purpose of balancing the design.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases internal validity because the applications of each technique are increased and comparability is facilitated.

12 SoftEng–Case, Code family, VV–UPV Replication

Replication	**VV-UPV** internal replication based on VV–UPM original experiment
Original Study	**Goal**: 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 at Polytechnic University of Madrid in -. This replication was carried out at Polytechnic University of Valencia in -.
Purposes	• Generalize results

Change #1	Discarding a technique (VV–UPV)
Description	Originally, the three verification and validation techniques are used: code reading, equivalence partitioning and branch testing. In this replication, the code reading technique is omitted. With the purpose of spending less time.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change decreases construct validity because one of the levels of the techniques factor is removed.
Comments	There were time constraints

Change #2	Duration of sessions (VV–UPV)
Description	Originally, the duration of the 3 sessions is 4h. each, i.e. the time is unlimited. In this replication, the duration of each of the 3 sessions is 2h.. With the purpose of spending less time.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change decreases internal validity because the time constraint may influence the results.
Comments	There were time constraints

Change #3	Training period (VV–UPV)
Description	Originally, subjects receive three four-hour training sessions to learn how to apply the techniques. In this replication, the training consists of two two-hour tutorials. With the purpose of spending less time as the subjects were already familiar with the technique.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect validity because the subjects are already familiar with the techniques and the training is not operationalised.

Change #4	Training in each technique (VV–UPV)
Description	Originally, the training in the use of the techniques is before the experiment is executed. In this 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. With the purpose of spending less time as the subjects were already familiar with the technique.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect validity because subjects were already familiar with the techniques.

Change #5	Application of techniques (VV–UPV)
Description	Originally, subjects apply a technique to a program in each session. In this replication, subjects apply the same technique to different programs in each session. With the purpose of spending less time.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change decreases internal validity because the three techniques are not analyzed and compared.

Change #6	Test cases in separate session (VV–UPV)
Description	Originally, subjects execute test cases with the application of the technique; that is to say in each session. In this replication, subjects execute test cases for one of the programs they have tested in a separate session, i.e. in session 3. With the purpose of spending less time.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change increases internal validity because the effect of conducting the test cases in a separate session is analyzed.

13 SoftEng–Case, Code family, VV–US Replication

Replication	**VV-US** internal replication based on VV–UPM original experiment
Original Study	**Goal**: 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 at Polytechnic University of Madrid in -. This replication was carried out at University of Seville in -.
Purposes	• Generalize results

Change #1	Duration of sessions (VV–US)
Description	Originally, the duration of the 3 sessions is 4h. each, i.e. the time is unlimited. In this replication, the duration of each of the 3 sessions is 2h.. With the purpose of spending less time.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change decreases internal validity because the time constraint may influence the results.
Comments	There were time constraints

Change #2	Test cases in separate session (VV–US)
Description	Originally, subjects execute test cases with the application of the technique; that is to say in each session. In this replication, subjects execute test cases for one of the programs they have tested in a later session, i.e. in session 4. With the purpose of spending less time.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change decreases internal validity because time restriction may affect the results.

Change #3	Subjects work in pairs (VV–US)
Description	Originally, subjects work individually. In this replication, subjects work in pairs. With the purpose of using computers.
Dimension	Context, specifically, the working methods.
Effects on validity	• This change increases internal validity because the effect of joint work can be analyzed.

Change #4	Training period (VV–US)
Description	Originally, subjects receive three four-hour training sessions to learn how to apply the techniques. In this replication, the training consists of two two-hour tutorials. With the purpose of spending less time as the subjects were already familiar with the technique.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect validity because the subjects are already familiar with the techniques and the training is not operationalised.

Change #5	Training in each technique (VV–US)
Description	Originally, the training in the use of the techniques is before the experiment is executed. In this 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. With the purpose of spending less time as the subjects were already familiar with the technique.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect validity because subjects were already familiar with the techniques.

14 SoftEng–Case, Code family, VV–ORT Replication

Replication	**VV–ORT** internal replication based on VV–UPM original experiment
Original Study	**Goal**: 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 at Polytechnic University of Madrid in -. This replication was carried out at University ORT Uruguay in -.
Purposes	• Generalize results

Change #1	Discarding a technique (VV–ORT)
Description	Originally, the three verification and validation techniques are used: code reading, equivalence partitioning and branch testing. In this replication, the code reading technique is omitted. With the purpose of spending less time.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change decreases construct validity because one of the levels of the techniques factor is removed.
Comments	There were time constraints

Change #2	Discarding a programme (VV–ORT)
Description	Originally, three program codes are used. In this replication, one of the programs is discarded. With the purpose of spending less time.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change decreases internal validity because one of the programmes on which failures were detected is removed.

Change #3	Duration of sessions (VV–ORT)
Description	Originally, the duration of the 3 sessions is 4h. each, i.e. the time is unlimited. In this replication, the experiment is executed in a single session. With the purpose of spending less time.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change decreases internal validity because the time constraint may influence the results.
Comments	There were time constraints

Change #4	Implementation of techniques by programmes (VV–ORT)
Description	Originally, subjects apply a different technique to evaluate a program in each of the three sessions. In this replication, subjects apply the two techniques to the two programs in a single session. With the purpose of spending less time.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change decreases internal validity because one of the techniques and one of the programmes is not used.

15 Agrobio–Case, Soil family, Soil–2018 Replication

Replication	**Soil–2018** internal replication based on Soil–2016 original experiment
Original Study	**Goal**: To evaluate the effect of a bio-surfactant on the assisted phytoremediation of contaminated soil **Description**: It is analyzed whether the addition of the biosurfactant JBR-425 to two types of copper (Cu)-contaminated soils (cause) reduces the Cu concentration in these soils 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 at IRNAS–CSIC in October 2016. This replication was carried out at the ETSIA of the US in March 2018.
Purposes	• Generalize results • Overcome some limitations of the baseline experiment

Change #1	Growing medium (Soil–2018)
Description	Originally, the experiment was carried out in a cultivation chamber. In this replication, was carried out in a greenhouse. With the purpose of simulate natural conditions.
Dimension	Context, specifically, the growing medium.
Effects on validity	• This change increases external validity because it allows generalization of results through replication under conditions closer to natural conditions.

Change #2	Plant types (Soil-2018)
Description	Originally, two plants were used: *Hordeum vulgare* and *Brassica juncea*. In this replication, only *Brassica juncea* was used . With the purpose of to use only *Brassica juncea*, as it is the only plant that accumulates metals.
Dimension	Protocol, specifically, the measuring instruments.
Effects on validity	• This change increases construct validity because (*Brassica juncea*) is the only plant in which the effect can be measured, i.e. the extracted Cu.
Comments	By using only one type of plant, it does not affect the results. It is not operationalised

Change #3	Soil types (Soil-2018)
Description	Originally, there were two types of soil: Coria (pH=7.8) and Constantina (pH=5.5). In this replication, only Constantina soil was used. With the purpose of of using only the soil of Constantina, as the metal is strongly adsorbed on the soil of Coria and the phytoextraction does not affect the biomass production.
Dimension	Population, specifically, the type of soil.
Effects on validity	• This change increases construct validity because the metal can be extracted from the soil. The extracted Cu is the effect to be measured.

Change #4	Copper dose reduction (Soil-2018)
Description	Originally, Copper (Cu) doses were 0, 500 and 1000 mg kg^{-1} . In this replication, Cu doses were adjusted to 0, 125, 250 and 500 mg kg^{-1} . With the purpose of of not using the Cu dose of 1000 mg kg^{-1} as it is toxic to the plant.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because the Cu dose is adjusted to non-toxic levels for the plant. • This change increases internal validity because the effect is analyzed at valid Cu levels.

Change #5	Form of application of copper (Soil-2018)
Description	Originally, Cu was applied as Copper Nitrate . In this replication, Cu was applied as Copper Sulfate . With the purpose of of using the most accessible reactive. The concentrations applied do not affect the plant.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change does not affect validity because reactives are equivalent.

Change #6	Soil aging time (Soil-2018)
Description	Originally, the soil aging time (from the time Cu is applied until the plant is sown) is 45 days . In this replication, soil aging time is 15 days . With the purpose of to reduce the time and reduce Cu retention.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases construct validity because the plant extracts the metal more easily and the effect is better measured. • This change decreases external validity because it reduces the generalizability of results to soils contaminated for a longer period of time.

Change #7	Increased biomass obtained (Soil-2018)
Description	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 this 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. With the purpose of of growing 4 pots in each tray to obtain sufficient biomass.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases conclusion validity 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	Biomass obtained in the fruiting phase (Soil-2018)
Description	Originally, the biomass is collected when the plants have between 2 and 3 true leaves. In this 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. With the purpose of avoid competition between plants, let the plants complete their vegetative cycle and thus obtain more biomass.
Dimension	Operationalization, specifically, the measurement procedure.
Effects on validity	• This change increases construct validity because more biomass is obtained for further analysis.

Change #9	Biomass obtained in the fruiting phase (Soil-2018)
Description	Originally, the pots in which the plants are sown are 300 ml tubes. In this replication, the pots are 500 ml bucket type. With the purpose of of having a larger soil volume to allow for more root development and higher biomass production.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because more biomass is obtained for further analysis.

16 Agrobio–Case, Soil family, Soil–2019 Replication

Replication	**Soil–2019** internal replication based on Soil–2016 original experiment
Original Study	**Goal**: To evaluate the effect of a bio-surfactant on the assisted phytoremediation of contaminated soil **Description**: It is analyzed whether the addition of the biosurfactant JBR-425 to two types of copper (Cu)-contaminated soils (cause) reduces the Cu concentration in these soils 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 at IRNAS–CSIC in October 2016. This replication was carried out at the ETSIA of the US in March 2019.
Purposes	• Generalize results

Change #1	Growing medium (Soil–2019)
Description	Originally, the experiment was carried out in a cultivation chamber. In this replication, was carried out in a greenhouse. With the purpose of simulate natural conditions.
Dimension	Context, specifically, the growing medium.
Effects on validity	• This change increases external validity because it allows generalization of results through replication under conditions closer to natural conditions.

Change #2	Plant types (Soil-2019)
Description	Originally, two plants were used: *Hordeum vulgare* and *Brassica juncea*. In this replication, only *Brassica juncea* was used . With the purpose of to use only *Brassica juncea*, as it is the only plant that accumulates metals.
Dimension	Protocol, specifically, the measuring instruments.
Effects on validity	• This change increases construct validity because (*Brassica juncea*) is the only plant in which the effect can be measured, i.e. the extracted Cu.
Comments	By using only one type of plant, it does not affect the results. It is not operationalised

Change #3	Soil types (Soil-2019)
Description	Originally, there were two types of soil: Coria (pH=7.8) and Constantina (pH=5.5). In this 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). With the purpose of of experimenting with naturally contaminated soils. The Miraflores soils are urban gardens with natural contamination and the Lebrija soil was used as a control.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases internal validity because soils of different characteristics are compared. • This change decreases conclusion validity because the construct is better reflected.
Comments	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 (Soil-2019)
Description	Originally, Copper (Cu) doses were 0, 500 and 1000 mg kg^{-1} . In this replication, the soils are not artificially contaminated with Cu. With the purpose of of using soils from urban gardens with natural contamination (Cu levels of 36 and 206 mg kg^{-1}).
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change increases internal validity because Cu levels are those present in the soil and are not toxic to the plant.

Change #5	Design change (Soil-2019)
Description	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 this 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. With the purpose of of adjusting the types of plants and soils .
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change increases internal validity because it allows the effect on different soil types to be compared.

Change #6	Biomass obtained in the fruiting phase (Soil-2019)
Description	Originally, the biomass is collected when the plants have between 2 and 3 true leaves. In this 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. With the purpose of avoid competition between plants, let the plants complete their vegetative cycle and thus obtain more biomass.
Dimension	Operationalization, specifically, the measurement procedure.
Effects on validity	• This change increases construct validity because more biomass is obtained for further analysis.

Change #7	Biomass obtained in the fruiting phase (Soil-2019)
Description	Originally, the pots in which the plants are sown are 300 ml tubes. In this replication, the pots are 500 ml bucket type. With the purpose of of having a larger soil volume to allow for more root development and higher biomass production.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because more biomass is obtained for further analysis.

17 Agrobio–Case, Harvest family, Harvesting System–2017 Replication

Replication	**Harvesting System–2017** internal replication based on Harvesting System–2016 original experiment
Original Study	**Goal**: 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
Site and Date	The base experiment was carried out at IG–CSIC in 2016. This replication was carried out at the same site in 2017.
Purposes	• Confirm results

Change #1	Different climatic conditions (Harvesting System–2017)
Description	Originally, the weather conditions are those of 2016. In this replication, climatic conditions are different as they correspond to 2017. With the purpose of of analyzing data from different years.
Dimension	Context, specifically, the environment.
Effects on validity	• This change increases external validity because it allows generalization of results to different seasons.

18 Agrobio–Case, Olive family, Olive–Des Replication

Replication	**Olive–Des** internal replication based on Olive–2015 original experiment
Original Study	**Goal**: To test different non-toxic solvents to extract phenolic compounds from virgin olive oil **Description**: Different green solvents (Deep Eutectic Solvents DES) for the extraction of phenolic compounds (EPC) from virgin olive oil are analyzed
Site and Date	The base experiment was carried out at IG–CSIC in 2015. This replication was carried out at the same site in 2015.
Purposes	• Generalize results

Change #1	Solvent extraction 1 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and Glycerol (1:2)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #2	Solvent extraction 2 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and Lactic acid (1:2)* . With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #3	Solvent extraction 3 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and Urea (1:2)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #4	Solvent extraction 4 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and Sucrose (1:1)* . With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #5	Solvent extraction 5 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and Sucrose (4:1)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #6	Solvent extraction 6 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and 1,4-Butanediol (1:5)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #7	Solvent extraction 7 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and Xylitol (2:1)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #8	Solvent extraction 8 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and 1,2-Propanediol (1:1)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #9	Solvent extraction 9 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride and Malonic acid (1:1)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #10	Solvent extraction 10 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *Choline chloride, Urea and Glycerol (1:1:1)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

Change #11	Solvent extraction 11 (Olive–Des)
Description	Originally, the extraction is done with a solution 80% (v/v) *methanol and water* . In this replication, the extraction is done with a solution *D-(–)-Fructose D-(+)-Glucose and Sucrose (1:1:1)*. With the purpose of of analysing non-toxic alternatives for extraction.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because non-toxic alternatives for the extraction of *phenolic compounds* are analyzed.

19 Agrobio–Case, Diet family, Diet–Hiper Replication

Replication	**Diet-Hiper** internal replication based on Diet-Normo original experiment
Original Study	**Goal**: To analyze 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 at IG–CSIC in 2006. This replication was carried out at the same site in 2006.
Purposes	• Generalize results

Change #1	Hypertensive subjects (Diet-Hiper)
Description	Originally, subjects have blood pressure levels within the limits considered normal. In this replication, subjects are *hypertensive* . With the purpose of of studying the effect of a diet rich in oleic acid in *hypertriglyceridemic* subjects who are also *hypertensive*.
Dimension	Population, specifically, the blood pressure level of the subjects.
Effects on validity	• This change increases external validity because the effect on *hypertensive* subjects is analyzed.

20 Comp–Case, Testing family, Test–NF Replication

Replication	**Test–NF** internal replication based on Test–F original experiment
Original Study	**Goal**: Prioritization of test case execution to accelerate the detection of faults in highly configurable systems **Description**: The order of test cases is determined by prioritization objectives. In this study, 63 combinations of up to three prioritization targets were studied to accelerate fault detection in the Drupal framework.
Site and Date	The base experiment was carried out at the ETSII of the US in 2015. This replication was carried out at the same site in 2015.
Purposes	• Generalize results

Change #1	Non-functional objectives (Test-NF)
Description	Originally, *objective functions* are *functional*. In this replication, the *objective functions* are *non–functional* . With the purpose of comparing the differences in favour of multi-objective prioritization over single-objective prioritization using non-functional objectives.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change does not affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected.

21 Comp–Case, Testing family, Test—F&NF Replication

Replication	**Test–F&NF** internal replication based on Test–F original experiment
Original Study	**Goal**: Prioritization of test case execution to accelerate the detection of faults in highly configurable systems **Description**: The order of test cases is determined by prioritization objectives. In this study, 63 combinations of up to three prioritization targets were studied to accelerate fault detection in the Drupal framework.
Site and Date	The base experiment was carried out at the ETSII of the US in 2015. This replication was carried out at the same site in 2015.
Purposes	• Generalize results

Change #1	Functional and non-functional objectives (Test–F&NF)
Description	Originally, *objective functions* are *functional*. In this replication, *objective functions* combine *functional* and *non–functional*. With the purpose of analysing whether *multi–objective* prioritization using *functional* and *non–functional* objectives outperform prioritization driven by a single objective, either functional or non–functional.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change does not affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected.

22 Comp–Case, Testing family, Test–FvsNF Replication

Replication	**Test–FvsNF** internal replication based on Test–F original experiment
Original Study	**Goal**: Prioritization of test case execution to accelerate the detection of faults in highly configurable systems **Description**: The order of test cases is determined by prioritization objectives. In this study, 63 combinations of up to three prioritization targets were studied to accelerate fault detection in the Drupal framework.
Site and Date	The base experiment was carried out at the ETSII of the US in 2015. This replication was carried out at the same site in 2015.
Purposes	• Generalize results

Change #1	Comparison of objectives (Test–FvsNF)
Description	Originally, *objective functions* are *functional*. In this replication, *objective functions* combine *functional* and *non–functional* . With the purpose of analyzing the domain of *non-functional objectives* over *functional objectives*, especially when these are combined in a *multi-objective* perspective.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change does not affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected.

23 Comp–Case, SPL family, SPL–Pr&Com Replication

Replication	**SPL–Pr&Com** internal replication based on SPL–Pr original experiment
Original Study	**Goal**: Comparison of test case prioritization criteria for Software Product Lines (SPL) **Description**: The applicability of test case prioritisation techniques to SPL testing is analysed. Five different prioritisation criteria based on common feature model metrics are proposed and their effectiveness in increasing the early failure detection rate, i.e. a measure of how quickly failures are detected, is compared.
Site and Date	The base experiment was carried out at the ETSII of the US in 2014. This replication was carried out at the same site in 2014.
Purposes	• Generalize results

Change #1	Set of tests (SPL–Pr&Com)
Description	Originally, only a test suite was generated. In this replication, for each model, 2-wise test suite was generated. With the purpose of obtaining a list of products that covers all possible pairs of characteristics in each model.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because it increases the number of tests.

Change #2	Generation of the test suite (SPL–Pr&Com)
Description	Originally, a test suite was randomly generated using SPLAR tool. In this replication, test suite was randomly generated using SPLCAT tool. With the purpose of increasing the failure detection rate by SPLCAT. It is therefore considered as an additional prioritisation approach in our comparison.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change increases construct validity because it increases the number of tests.