Supplemental Material: Instantiation of the Proposed Templates in the Multiple Case Study Using CÆSAR LaTeX Template

Margarita Cruz, Beatriz Bernárdez, Amador Durán, Cathy Guevara-Vega, Antonio Ruiz-Cortés

May, 2022

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

Replication	**Mind#2** (`https://doi.org/10.1016/j.jss.2016.06.104`) 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 University of Seville in first semester of the 2013-2014 academic year. This replication was carried out at the same site in first semester of the 2014-2015 academic year.
Purposes	• Confirm results • Overcome some limitations of the baseline experiment
Comments	First replication after original experiment

Change #1	Random assignment of subjects to groups (Mind#2)
Description	Originally, group assignment was based on subjects' preferences.. In this replication, assignment to groups was randomized. With the purpose of mitigating selection and allocation bias and avoiding limitations of statistical analysis.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change substantially increases (+3) conclusion validity because it improves the power of the applicable statistical tests. • This change substantially increases (+3) internal validity because it eliminates assignment bias.

Change #2	Increased duration of the treatment (Mind#2)
Description	Originally, the treatment was applied for 4 weeks, 4 days a week, in 10-minute sessions. In this replication, the treatment was applied for 6 weeks, 4 days a week, in 12-minute sessions. With the purpose of increasing the effect of the treatment.
Dimension	Operationalization, specifically, the treatment.
Effects on validity	• This change moderately increases (+2) construct validity because it reflects better the effect of mindfulness practice. • This change moderately increases (+2) internal validity because it reinforces the effect of the treatment over other uncontrolled factors.
Comments	Available time during class break was 20 minutes

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 treatment.
Effects on validity	• This change moderately increases (+2) internal validity because it eliminates the possibility that the placebo could have any effect on the experimental outcomes.
Comments	The new public speaking workshop was online mainly

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

Replication	**Mind#3** (https://doi.org/10.1109/TSE.2020.2991699) 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 design.
Effects on validity	• This change slightly increases (+1) internal validity because it allows to study the effect of the difference between tasks on the experimental results . • This change substantially increases (+3) conclusion validity because it allows to analyze and adjust the effect of the difference between tasks on the experimental results.

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

Replication	**Q–2009** (https://doi.org/10.20868/UPM.thesis.40566) internal replication based on Q–2007 original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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	Effectiveness analysis (Q–2009)
Description	Originally, analysts' effectiveness in interview sessions was analyzed. In this replication, effectiveness was 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 substantially decreases (-3) construct validity because the dependent variable is not analysed.

Change #2	Retention capacity analysis (Q–2009)
Description	Originally, the retention capacity was analyzed. In this replication, retention capacity was 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 substantially decreases (-3) construct validity because the retention capacity is not analysed.

Change #3	Analysts' experience in development (Q–2009)
Description	Originally, experience in development was considered to calculate the independent variable experience. In this replication, experience in development was considered to calculate the independent variable experience. With the purpose of analyse its effect on the results.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) construct validity because the possible influence of one more variable is analysed.

Change #4	Interview language (Q–2009)
Description	Originally, interviews were conducted in Spanish. In this replication, interviews were conducted in English. With the purpose of use the language of the students' master's degree.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change moderately decreases (-2) internal validity because there may be comprehension problems that affect the results.

Change #5	Change of experimenter (Q–2009)
Description	Originally, an experimenter played the role of client in the interviews by answering questions from the experimental subjects (analysts). In this replication, the experimenter (client) was replaced. With the purpose of replacing the experimenter who was not available.
Dimension	Experimenter, specifically, the monitor.
Effects on validity	• This change moderately increases (+2) internal validity because experimenter bias is eliminated.

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

Replication	**Q–2011** (`https://doi.org/10.20868/UPM.thesis.40566`) internal replication based on Q–2009 original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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	• Confirm results • Generalize results

Change #1	Group interviews (Q–2011)
Description	Originally, interviews between subjects (analysts) and experimenter were individual. In this replication, interviews were conducted in groups. With the purpose of reduce the cost and effort of conducting individual interviews.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately decreases (-2) internal validity because there may be comprehension problems that affect the results.

Change #2	Experience determination (Q–2011)
Description	Originally, experience in requirements analysis was determined by their years of experience. In this replication, experience was further determined by the subject's abilities. With the purpose of consider the subject's abilities.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) construct validity because there is a new variable to better capture the construct. • This change moderately decreases (-2) conclusion validity because the procedure becomes tedious.

Change #3	Duration of interviews (Q–2011)
Description	Originally, the duration of the interviews was 30 min. In this replication, the duration of the interviews was 60 min. With the purpose of the interviews are group interviews.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately increases (+2) internal validity because increasing the duration of interviews allows for a better understanding of the requirements.

Change #4	Requirements report (Q–2011)
Description	Originally, after the interview, the subjects (analysts) had 7 days to report the requirements. In this replication, subjects (analysts) reported the requirements immediately after the interview. With the purpose of avoid loss of information.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately increases (+2) 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 required by the analyst to report the requirements was not measured. In this replication, the consolidation time was 120 minutes. With the purpose of submit the requirements report immediately after the interview.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately increases (+2) internal validity because it is analysed whether time influences the information collected.

Change #6	Changing experimenter (Q–2011)
Description	Originally, an experimenter plays the role of client in the interviews by answering questions from the experimental subjects (analysts). In this replication, the experimenter (client) was replaced. With the purpose of replacing the experimenter who was not available.
Dimension	Experimenter, specifically, the monitor.
Effects on validity	• This change moderately increases (+2) internal validity because experimenter bias is eliminated.

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

Replication	**Q–2012** (https://doi.org/10.20868/UPM.thesis.40566) external replication based on Q–2011 original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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 were Master's students. In this replication, the subjects were professionals. With the purpose of use participants in the International Working Conference on Requirements Engineering as subjects.
Dimension	Population, specifically, the experience.
Effects on validity	• This change substantially increases (+3) external validity because the results extend to subjects who are professionals.

Change #2	Development experience (Q–2012)
Description	Originally, subjects had little or no development experience. In this replication, the subjects were professionals with experience in development. With the purpose of analyse the influence of experience on development.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change substantially increases (+3) construct validity because the new independent variable 'experience on development' is defined.

Change #3	Consolidation time (Q–2012)
Description	Originally, the subjects (analysts) had 120 minutes to report the requirements (consolidation time). In this replication, the consolidation time was 30 minutes. With the purpose of adapt to the time available.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately decreases (-2) internal validity because the time available may affect the results.

Change #4	Elimination of the training period (Q–2012)
Description	Originally, the experiment was carried out at the end of the course, i.e. after the training period. In this replication, there was no training period. With the purpose of adapting to the International Working Conference on Requirements Engineering.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change does not affect (0) internal validity because the subjects are experienced professionals and do not need a period of training.

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

Replication	**E–2012A** (https://doi.org/10.20868/UPM.thesis.40566) internal replication based on Q–2012 original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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 of the problem domain was defined, by subjective assessment, as 'familiarity'. In this replication, knowledge was defined as an independent variable with two levels: known and unknown problem domain. With the purpose of analyze whether knowledge of the problem domain affected the results.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) construct validity because the new independent variable 'problem domain' is defined.

Change #2	Repeated measures design (E–2012A)
Description	Originally, the interviews were conducted on two different days. In this replication, the design was changed to a design of repeated measurements (within-subjects). With the purpose of adapt the experiment as there were fewer experimental subjects.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change moderately increases (+2) conclusion validity because improves the power of the applicable statistical tests. • This change substantially increases (+3) internal validity because the change of experimental design better reflects the construct.

Change #3	Design of repeated measures (E–2012A)
Description	Originally, interviews between subjects (analysts) and experimenter (clients) were in group. In this replication, interviews were individual. With the purpose of conduct interviews with 2 experimenters (clients) and in two languages.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately increases (+2) internal validity because experimenter bias is eliminated. • This change moderately decreases (-2) conclusion validity because although there were two responding monitors, the process was tedious for these monitors due to the larger number of individual interviews.

Change #4	Blocking by language (E–2012A)
Description	Originally, there were no blocking variables. In this replication, there was a blocking variable per language. With the purpose of prevent language from influencing results.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change substantially increases (+3) internal validity because language is prevented from influencing the results.

Change #5	Blocking by experimenter (E–2012A)
Description	Originally, there were no blocking variables. In this replication, there was a blocking variable per experimenter (respondent). With the purpose of prevent experimenter from influencing results.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change substantially increases (+3) internal validity because experimenter is prevented from influencing the results.

Change #6	Number of experimenter (E–2012A)
Description	Originally, there was an experimenter (clients). In this replication, there were two experimenters (clients). With the purpose of alleviate the effects of fatigue and learning of the experimenter.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately increases (+2) internal validity because experimenter fatigue and learning is prevented from influencing the results.

Change #7	Number of problems (E–2012A)
Description	Originally, there was only one problem. In this replication, there were two problems. With the purpose of use two problems due to the new design and the blocking variables.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change moderately increases (+2) internal validity because in the new design 2 blocking variables have been defined.

Change #8	Interview duration (E–2012A)
Description	Originally, the duration of the interviews was 60 min. In this replication, the duration of the interviews was 30. With the purpose of the interviews were individual.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately increases (+2) internal validity because the interviews were individual and this could affect results.

Change #9	Time of consolidation (E–2012A)
Description	Originally, the subjects (analysts) had 30 minutes to report the requirements (consolidation time). In this replication, the consolidation time was 90 minutes. With the purpose of more time available for reporting requirements.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change slightly increases (+1) interna validity because more time is available and the information collected can be more accurate.

Change #10	Problem difficulty (E–2012A)
Description	Originally, the difficulty of the problem was not measured. In this replication, the new variable 'problem difficulty' is defined. With the purpose of analyse the influence of the difficulty of the problem on the results.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) construct validity because there is a new variable to better capture the construct.

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

Replication	**E–2012B** (https://doi.org/10.20868/UPM.thesis.40566) internal replication based on E–2012A original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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 were used, one known domain and one unknown domain. In this replication, the problems have been changed although they are still known domain and unknown domain. With the purpose of using different problems for requirements analysis.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change does not affect (0) internal validity because the difficulty of the problems is similar to the previous ones.

Change #2	Order of problems (E–2012B)
Description	Originally, first the known domain problem was performed and then the unknown domain problem. In this replication, the order of the problems was swapped. With the purpose of analyse whether it affects the results.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change slightly increases (+1) internal validity because allows to analyse whether the order affects the results. • This change slightly increases (+1) conclusion validity because it allows to analyse 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 was carried out after the subjects have received training in Requirements Engineering. With the purpose of analyse whether it affects the results.
Dimension	Context, specifically, the the time of performance.
Effects on validity	• This change moderately decreases (-2) internal validity because the replication was carried out at the end of the course and may influence the results.

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

Replication	**E–2013** (https://doi.org/10.20868/UPM.thesis.40566) internal replication based on E–2012B original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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	• Confirm results

Change #1	Inter-subject design (E–2013)
Description	Originally, the design was of repeated measurements. In this replication, the design was between-subjects. With the purpose of avoiding the learning effect.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change substantially increases (+3) internal validity because in the between-subjects design, each subject receives a single treatment to avoid the learning effect.
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, there was no short training (warm-up) before the course. In this replication, short training (warm-up) was 1 week. With the purpose of to analyse the effect of training.
Dimension	Operationalization, specifically, the treatments .
Effects on validity	• This change moderately increases (+2) internal validity because the effect of training is analysed against other factors. • This change substantially increases (+3) construct validity because the effect of training is better reflected.

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

Replication	**E–2014** (https://doi.org/10.20868/UPM.thesis.40566) internal replication based on E–2013 original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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	• Confirm results

Change #1	Only one experimenter (E–2014)
Description	Originally, two experimenters played the role of clients in the interviews by answering questions from the experimental subjects (analysts). In this replication, there was only one experimenter. With the purpose of having only one experimenter as the other was not available.
Dimension	Protocol, specifically, the guides.
Effects on validity	• This change moderately increases (+2) internal validity because all subjects (analysts) receive the same information as there is only one experimenter (client). • This change slightly decreases (-1) conclusion validity because by having only one experimenter (client), the process can become tedious for the experimenter.

Change #2	Training increased (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 analyse the effect of increasing training.
Dimension	Operationalization, specifically, the treatments .
Effects on validity	• This change moderately increases (+2) construct validity because the effect of training is better reflected. • This change moderately increases (+2) internal validity because the effect of training is analysed against other factors.

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

Replication	**E–2015** (`https://doi.org/10.20868/UPM.thesis.40566`) internal replication based on E–2014 original experiment
Original Study	**Goal**: To study the influence of analysts' experience and domain knowledge on the effectiveness of requirements elicitation **Description**: Analysts' effectiveness was analysed using interviews as the requirements elicitation technique. In the interviews, the experimenter played the role of the customer by answering questions from the experimental subjects (analysts) about two possible problems, one from a known domain and another one from an unknown domain. To measure the effectiveness of consolidation, after some time, the analyst wrote down what he/she remembered from the interview and the number of problem items mentioned by the analyst was measured
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	• Confirm results

Change #1	Increasing the duration of training (E–2015)
Description	Originally, short training (warm-up) was 1 week. In this replication, short training (warm-up) was 2 week. With the purpose of analyse the effect of increasing training.
Dimension	Operationalization, specifically, the treatments .
Effects on validity	• This change moderately increases (+2) internal validity because the effect of training is analysed against other factors. • This change moderately increases (+2) construct validity because the effect of training is better reflected.

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

Replication	**VV–UPM1** (https://doi.org/10.1109/ESEM.2009.5314236) 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 2001. This replication was carried out at the same site in 2002.
Purposes	• Generalize results

Change #1	Visibility of fault (VV–UPM1)
Description	Originally, the visibility of fault was not analysed. In this replication, the visibility of fault or the number of people who detected the fault was analysed. With the purpose of to draw new conclusions.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change moderately increases (+2) 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 programme is a factor (independent variable) although its influence was not analysed. In this replication, two versions of each program are implemented and is a new factor. With the purpose of to draw new conclusions.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) construct validity because the possible influence of programme version is analysed.

Change #3	Increase the number of faults (VV–UPM1)
Description	Originally, there are three types of faults that appeared once. The other three types of faults appeared twice. In this replication, all types of faults doubled. With the purpose of increase the number of faults as there were two versions of each programme.
Dimension	Operationalization, specifically, the metrics.
Effects on validity	• This change moderately increases (+2) 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, the subjects generated their test cases to detect faults in the code. In this replication, first, the subjects applied the technique to generate the test cases and then executed the test cases provided to them in order to detect program failures. With the purpose of check whether the visibility of faults influences their detection.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change slightly increases (+1) internal validity because the generation and execution of test cases reinforces the effect of the technique.

Change #5	Discarding a programme (VV–UPM)
Description	Originally, four programs were used. In this replication, three programs were used, one was discarded. With the purpose of balance the design.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change moderately decreases (-2) 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 applied a technique. In this replication, each subject applied the three techniques. With the purpose of to facilitate the comparison of techniques.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change moderately increases (+2) internal validity because the applications of each technique are increased and comparability is facilitated.

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

Replication	**VV-UPV** (https://doi.org/10.1109/ICST.2012.113) 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 2001. This replication was carried out at Polytechnic University of Valencia in around 2005.
Purposes	• Generalize results

Change #1	Discarding technique (VV–UPV)
Description	Originally, the three verification and validation techniques were used: code reading, equivalence partitioning and branch testing. In this replication, the code reading technique was omitted. With the purpose of adapt to the time available.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately decreases (-2) construct validity because one of the levels of the techniques factor is removed.
Comments	

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

Change #3	Training period (VV–UPV)
Description	Originally, subjects received three four-hour training sessions to learn how to apply the techniques. In this replication, the training consisted of two two-hour tutorials. With the purpose of take advantage of the fact that the subjects are already familiar with the techniques.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect (0) 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, training in the use of the techniques was prior to the execution of the experiment. In this replication, at the beginning of each session, a tutorial on the use of the technique was conducted. With the purpose of take advantage of the fact that the subjects are already familiar with the techniques.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect (0) affect validity because the subjects are already familiar with the techniques.

Change #5	Application of techniques (VV–UPV)
Description	Originally, subjects applied a technique to a program in each session. In this replication, subjects applied the same technique to different programs in each session. With the purpose of adapt to the time available.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change substantially decreases (-3) internal validity because the three techniques are not analysed and compared.

Change #6	Test cases in separate session (VV–UPV)
Description	Originally, subjects performed the test cases with the application of the technique, i.e. in each session. In this replication, subjects performed the test cases of one of the programmes they tested in a separate session. With the purpose of adapt to the time available..
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change moderately decreases (-2) internal validity because the effect of conducting the test cases in a separate session is analysed.

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

Replication	**VV-US** (https://doi.org/10.1016/j.infsof.2012.07.016) 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 2001. This replication was carried out at University of Seville in around 2005.
Purposes	• Generalize results

Change #1	Session duraction (VV–US)
Description	Originally, the duration of the 3 sessions was 4h. each, i.e. the time was unlimited. In this replication, the duration of each of the 3 sessions was 2h. With the purpose of adapt to the time available.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change moderately decreases (-2) internal validity because time constraints may influence results.
Comments	

Change #2	Test cases in separate session (VV–US)
Description	Originally, subjects performed the test cases with the application of the technique, i.e. in each session. In this replication, subjects performed the test cases of one of the programmes they tested in a later session. With the purpose of adapt to the time available.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change slightly decreases (-1) internal validity because time constraints may influence results.

Change #3	Subjects work in pairs (VV–US)
Description	Originally, subjects worked individually. In this replication, subjects worked in pairs. With the purpose of use available computers.
Dimension	Context, specifically, the working methods.
Effects on validity	• This change slightly decreases (-1) internal validity because it could affect the results.

Change #4	Reduction of training period (VV–US)
Description	Originally, subjects received three four-hour training sessions to learn how to apply the techniques. In this replication, the training consisted of two two-hour tutorials. With the purpose of take advantage of the fact that the subjects are already familiar with the techniques.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect (0) affect validity because tthe subjects are already familiar with the techniques.

Change #5	Training of each technique (VV–US)
Description	Originally, training in the use of the techniques was prior to the execution of the experiment. In this replication, each tutorial was conducted in each of the three sessions in which each technique was examined, i.e. the training was interspersed with the performance of the experiment.. With the purpose of take advantage of the fact that the subjects are already familiar with the techniques.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change does not affect (0) affect validity because the subjects are already familiar with the techniques.

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

Replication	**VV–ORT** (https://doi.org/10.1016/j.infsof.2012.07.016) 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 2001. This replication was carried out at University ORT Uruguay in around 2005.
Purposes	• Generalize results

Change #1	Discarding a technique (VV–ORT)
Description	Originally, the three verification and validation techniques were used: code reading, equivalence partitioning and branch testing. In this replication, the code reading technique was omitted. With the purpose of adapt to the time available.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately decreases (-2) construct validity because one of the levels of the techniques factor is removed.
Comments	

Change #2	Discard a programme (VV–ORT)
Description	Originally, four programs were used. In this replication, three programs were used, one was discarded. With the purpose of balance the design and adapt to the time available.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change moderately decreases (-2) 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 was 4h. each, i.e. the time was unlimited. In this replication, the experiment was carried out in a single session. With the purpose of adapt to the time available.
Dimension	Protocol, specifically, the guides .
Effects on validity	• This change moderately decreases (-2) internal validity because time constraints may influence influence results.
Comments	

Change #4	Technique per programme (VV–ORT)
Description	Originally, subjects applied a different technique to evaluate a program in each of the three sessions. In this replication, the subjects applied two techniques to two programmes in a single session. With the purpose of adapt to the time available.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change moderately decreases (-2) 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** (`https://hdl.handle.net/11441/132481`) internal replication based on Soil–2016 original experiment
Original Study	**Goal**: To study whether the addition of rhamnolipids (cause) improves Cu phytoextraction (effect) in Cu-contaminated soils (population) **Description**: Samples are taken from two different soils (Coria and Constantina), which are then artificially contaminated with Cu, distributed in pots, and left to age for 45 days. Then, barley and mustard seeds are sown, fertiliser is added, and JBR-425 rhamnolipid is added after 15 days. After 30 days, the Cu content of plant biomass and the availability of Cu in soil are analyzed by two different methods: extraction with CaCl2 and EDTA. Cu was applied in three concentrations: 0, 500 and 1000 mg/kg, so 6 treatment groups were obtained combining the 3 concentrations of Cu with the addition or not of JBR-425. The control group was taken as the one that was neither contaminated with copper nor added with JBR-425
Site and Date	The base experiment was carried out at ETSIA of the University of Seville in October 2015. This replication was carried out at same site in March 2018.
Purposes	• Confirm results • Generalize results

Change #1	Greenhouse cultivation (Soil–2018)
Description	Originally, the experiment was carried out in a culture chamber. In this replication, the experiment was carried out in a greenhouse. With the purpose of simulate natural conditions.
Dimension	Context, specifically, the the medium.
Effects on validity	• This change substantially increases (+3) external validity because generalises the results for conditions closer to natural conditions.

Change #2	Only mustard is grown (Soil-2018)
Description	Originally, two plants were used: barley and mustard. In this replication, only the mustard plant was used. With the purpose of use the plant with the highest Cu phytoextraction capacity.
Dimension	Protocol, specifically, the measuring instruments.
Effects on validity	• This change substantially increases (+3) construct validity because using mustard the effect is best measured by phytoextracted Cu.

Change #3	Only Constantina soil is used (Soil-2018)
Description	Originally, two soil types were sampled: Coria (pH=7.8) and Constantina (pH=5.5). In this replication, soil samples were taken only from Constantina. With the purpose of use a soil from which the metal can be extracted. In the soil of Coria Cu is strongly absorbed and cannot be phytoextracted.
Dimension	Population, specifically, the type of soil.
Effects on validity	• This change substantially increases (+3) construct validity because it ensures that the metal can be phytoextracted, which is the effect to be measured.

Change #4	Cu dose reduction (Soil-2018)
Description	Originally, the doses of Cu added to the soils were 0, 500 and 1000 mg/kg. In this replication, the doses of Cu added to the soils were 0, 125, 250 and 500 mg/kg. With the purpose of avoiding toxic levels of Cu for the plants.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) construct validity because at high Cu doses, the plants die and the effect cannot be measured.

Change #5	Cu applied as Cu sulphate r (Soil-2018)
Description	Originally, cu was applied as Cu nitrate. In this replication, cu was applied as Cu Sulphate. With the purpose of use the most accessible reagent.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change does not affect (0) affect validity because it is only a change of one reagent for an equivalent one.

Change #6	Reduction of soil ageing time (Soil-2018)
Description	Originally, soil ageing time (from Cu application to sowing) was 45 days. In this replication, the ageing time was 15 days. With the purpose of avoid excessive fixation of Cu to the soil which hinders its phytoextraction.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) construct validity because the plant extracts the metal more easily as it is not so retained and reflects the construction better. • This change moderately decreases (-2) external validity because it reduces the generalisation of results to contaminated soils that have been contaminated for a longer period of time.

Change #7	Increase in the number of pots per treatment type (Soil-2018)
Description	Originally, there were (3x2x2x2) 24 experimental units (3 Cu levels, 2 JBR, 2 plant types and 2 soil types). For each experimental unit a pot is prepared and repeated 3 times. In this replication, there are (4x2) 8 experimental units (4 Cu and 2 JBR levels). For each experimental unit 4 pots are prepared and placed in a tray. This is repeated 3 times. With the purpose of increase the number of pots to obtain sufficient biomass.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change substantially increases (+3) conclusion validity because increasing the number of subjects (n) improves the power of statistical tests, reducing the probability of obtaining a false negative (type II error).

Change #8	Biomass obtained at fruiting stage (Soil-2018)
Description	Originally, biomass was collected when plants had 2 to 3 true leaves. In this replication, biomass was collected when plants reached the fruiting stage. With the purpose of obtain more biomass as plants complete their vegetative cycle.
Dimension	Operationalization, specifically, the measurement procedure.
Effects on validity	• This change moderately increases (+2) construct validity because further biomass has been obtained for further analysis.
Comments	Modify the measurement procedure by collecting the leaves later

Change #9	Increasing the volume of the pots (Soil-2018)
Description	Originally, the pots were of the 300 ml tube type. In this replication, the pots were of the 500 ml bucket type. With the purpose of achieve greater root development and produce more biomass.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change slightly increases (+1) construct validity because further biomass has been obtained for further analysis.

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

Replication	**Soil–2019** (`https://hdl.handle.net/11441/132478`) internal replication based on Soil–2016 original experiment
Original Study	**Goal**: To study whether the addition of rhamnolipids (cause) improves Cu phytoextraction (effect) in Cu-contaminated soils (population) **Description**: Samples are taken from two different soils (Coria and Constantina), which are then artificially contaminated with Cu, distributed in pots, and left to age for 45 days. Then, barley and mustard seeds are sown, fertiliser is added, and JBR-425 rhamnolipid is added after 15 days. After 30 days, the Cu content of plant biomass and the availability of Cu in soil are analyzed by two different methods: extraction with CaCl2 and EDTA. Cu was applied in three concentrations: 0, 500 and 1000 mg/kg, so 6 treatment groups were obtained combining the 3 concentrations of Cu with the addition or not of JBR-425. The control group was taken as the one that was neither contaminated with copper nor added with JBR-425
Site and Date	The base experiment was carried out at ETSIA of the University of Seville in October 2016. This replication was carried out at the same site in March 2019.
Purposes	• Confirm results • Generalize results

Change #1	Growing in greenhouses (Soil–2019)
Description	Originally, the experiment was carried out in a culture chamber. In this replication, the experiment was carried out in a greenhouse. With the purpose of simulate natural conditions.
Dimension	Context, specifically, the growing medium.
Effects on validity	• This change substantially increases (+3) external validity because generalises the results for conditions closer to natural conditions.

Change #2	Mustard is only grown (Soil-2019)
Description	Originally, two plants were used: barley and mustard. In this replication, only the mustard plant was used. With the purpose of use the plant with the highest Cu phytoextraction capacity.
Dimension	Protocol, specifically, the measuring instruments.
Effects on validity	• This change substantially increases (+3) construct validity because using mustard the effect is best measured by phytoextracted Cu.

Change #3	Naturally contaminated soils (Soil-2019)
Description	Originally, two soil types were sampled: Coria (pH=7.8) and Constantina (pH=5.5). In this replication, three soil types were sampled: Miraflores-1 and Miraflores-2 (with Pb, Zn and Cu) and Lebrija (not contaminated by metals). With the purpose of use naturally contaminated soils. The Miraflores soils are urban gardens with natural contamination and the Lebrija soil is used as a control.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change substantially increases (+3) conclusion validity because the construct is better reflected. • This change substantially increases (+3) internal validity because soils of different characteristics are compared.

Change #4	Cu null dose (Soil-2019)
Description	Originally, the doses of Cu added to the soils were 0, 500 and 1000 mg/kg. In this replication, soils are not artificially contaminated with Cu. With the purpose of experiment with the Cu already in the soil.
Dimension	Operationalization, specifically, the treatments.
Effects on validity	• This change moderately increases (+2) internal validity because cu levels are those present in the soil, are non-toxic to the plant and better reflects the construct.
Comments	These soils are urban gardens with natural contamination (Cu levels 36 and 206 mg/kg)

Change #5	Number of pots per treatment type (Soil-2019)
Description	Originally, there were (3x2x2x2) 24 experimental units (3 Cu levels, 2 JBR, 2 plant types and 2 soil types). For each experimental unit a pot is prepared and repeated 3 times. In this replication, there are (3x2) 6 experimental units (3 soils and 2 JBR levels). For each experimental unit 4 pots are prepared forming a block. This is repeated 3 times. With the purpose of analyse the influence of two naturally contaminated soils and a third uncontaminated one.
Dimension	Protocol, specifically, the experimental design.
Effects on validity	• This change substantially increases (+3) conclusion validity because increasing the number of subjects (n) improves the power of statistical tests, reducing the probability of obtaining a false negative (type II error).

Change #6	biomass at fruiting phase (Soil-2019)
Description	Originally, biomass was collected when plants had 2 to 3 true leaves. In this replication, biomass was collected when plants reached the fruiting stage. With the purpose of obtain more biomass as plants complete their vegetative cycle.
Dimension	Operationalization, specifically, the measurement procedure.
Effects on validity	• This change moderately increases (+2) construct validity because further biomass has been obtained for further analysis..

Change #7	Increase the volume of pots (Soil-2019)
Description	Originally, the pots were of the 300 ml tube type. In this replication, the pots were of the 500 ml bucket type. With the purpose of achieve greater root development and produce more biomass.
Dimension	Protocol, specifically, the experimental material.
Effects on validity	• This change slightly increases (+1) construct validity because further biomass has been obtained for further analysis.

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

Replication	**Harvesting System–2017** (`https://hdl.handle.net/11441/133365`) 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 substantially increases (+3) external validity because it allows generalization of results to different seasons.

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

Replication	**Olive–Des** (https://doi.org/10.1016/j.foodchem.2015.10.131) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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 moderately increases (+2) 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** (https://doi.org/10.1016/j.jnutbio.2007.03.002) 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 substantially increases (+3) external validity because allows to analyse whether the results are generalisable to *hypertensive* subjects.

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

Replication	**Test–NF** (`https://doi.org/10.1016/j.jss.2016.09.045`) 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 (0) affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected.

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

Replication	**Test–F&NF** (https://doi.org/10.1016/j.jss.2016.09.045) 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 (0) affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected.

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

Replication	**Test–FvsNF** (https://doi.org/10.1016/j.jss.2016.09.045) 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 (0) affect validity because dependent variables of the original are replaced by other dependent variables, however the validity is not affected.

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

Replication	**SPL–Pr&Com** (`https://doi.org/10.1109/ICST.2014.15`) 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 substantially increases (+3) 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 substantially increases (+3) construct validity because it increases the number of tests.