Threats to Validity and their Control Actions – Results of a Systematic Literature Review

Amadeu Anderlin Neto and Tayana Conte
{neto.amadeu, tayana}@icomp.ufam.edu.br

USES – Grupo de Usabilidade e Engenharia de Software
PPGI – Programa de Pós-Graduação em Informática
Instituto de Computação (IComp), Universidade Federal do Amazonas (UFAM) – Manaus, AM – Brazil

Technical Report USES
TR-USES-2014-0002
March 2014

Programa de Pós-Graduação em Informática
Universidade Federal do Amazonas
Manaus, Amazonas 69077-000
URL: http://www.ufam.edu.br

Threats to Validity and their Control Actions: Results of a Systematic Literature Review

Amadeu Anderlin Neto and Tayana Conte
{neto.amadeu, tayana}@icomp.ufam.edu.br

USES – Grupo de Usabilidade e Engenharia de Software
PPGI – Programa de Pós-Graduação em Informática
Instituto de Computação, UFAM, Manaus, AM – Brazil

Universidade Federal do Amazonas, Programa de Pós-Graduação em Informática, RT-USES-2014-0002
March 2014

Abstract

This technical report presents the results of a Systematic Literature Review (SLR). The SLR aims at identifying threats to validity and control actions which occur in controlled experiments. We describe the threats to validity and possible ways to control them, based on evidence from literature. In addition, we present relationships that were identified during the analysis of the data extracted. When we select an action to control a threat to validity, this action may cause another threat.

Keywords: empirical software engineering, threats to validity, control actions.

1. Introduction

This technical report presents the results of a SLR which was conducted to identify threats to validity and actions to address them. In addition, we found relationships between the threats to validity and control actions. These relationships were classified into *"controlled by"* and "*may cause*". The first one means that a control action can mitigate a threat to validity. On the other hand, the second one means that a control action may cause a threat to validity.

According to Conte (2009), SLRs are based on a well-defined research strategy, which aims to detect the maximum possible relevant bibliographic material. We define a protocol revision that specifies the research question and methods that will be used to perform the review. A SLR proposes a fair assessment of the research topic. We use a rigorous review methodology, reliable, and capable of auditing (Kitchenham and Charters, 2007).

2. Research Strategy

Initially, we used the SLR performed by Sjoberg et al. (2005). This SLR aimed at investigating the quality of experiments published in Software Engineering. As far as we know, this SLR covered the majority of venues in Software Engineering. However, the research questions (RQs) explored by Sjoberg et al. (2005) are different from ours. We are interested in threats to validity described in controlled experiments, control actions, and relationships between them (see our RQs in Table 1).

Table 1. Research Questions of this SLR.

Research Questions	Goal
RQ1 - How many and which threats to validity and actions to address them are described in papers that report controlled experiments?	RQ1 aims at finding out which threats to validity and control actions are reported in controlled experiments. An expected result concerning this RQ is to build a list that contains the different threats to validity and their respective control actions.
RQ2 - Which relationships are there between threats to validity and actions to address them?	RQ2 aims at identifying relationships that occur between threats to validity and control actions. These relationships provide information about consequences of choosing a control action for a threat to validity. An expected result concerning this RQ is to build a conceptual model to assist researchers in the identification and mitigation of threats to validity.

To attend our specific goal, a paper from Sjoberg's SLR was excluded if: (a) it was not available for download; or (b) it did not present threats to validity. Following this exclusion criteria, out of a total of 103 papers, we extracted data from 47 (see Table 2). Extracted data comprised the threats to validity and control actions reported in the 47 controlled experiments. Thereafter, we produced a list of threats to validity and control actions, and identified relationships between them. The whole process was reviewed by an independent researcher. In case of any disagreement, the researchers had a discussion in order to address them.

Table 2. Results of Sjoberg's SLR: Jan. 1993 – Dez. 2002.

Journal/Conference	Initial	Final	Percentage
Empirical Software Engineering (EMSE)	22	16	34.04%
IEEE Transactions on Software Engineering (TSE)	17	8	17.02%
Information and Software Technology (IST)	8	6	12.77%
Journal of Systems and Software (JSS)	24	5	10.64%
International Conference on Software Engineering (ICSE)	12	5	10.64%
IEEE International Symposium on Software Metrics (METRICS)	10	5	10.64%
IEEE International Symposium on Empirical Software Engineering (ISESE)	3	2	4.25%
IEEE Software (IEEE SW)	4	0	0.00%
Software Maintenance and Evolution (SME)	2	0	0.00%
ACM Transactions on Software Engineering Methodology (TOSEM)	1	0	0.00%
Software: Practice and Experience (SP&E)	0	0	0.00%
IEEE Computer (IEEE Comp)	0	0	0.00%
Total:	103	47	100%

However, we did not identify enough data to answer our RQs. We believe it happened because Sjoberg et al. (2005) selected papers published between 1993 and 2002. According to Carver et al. (2004), since then a growing number of researchers were including experimental results in their research work. Thus, we identified the need to conduct a new SLR.

The goals of our SLR was: (a) to increase the evidences of threats to validity, control actions, and relationships between them identified in the preliminary literature review; and (b) to identify new threats to validity, control actions, and relationships which were not found in the preliminary literature review.

To conduct our SLR, we used the same procedure described by Sjoberg et al. (2005). Therefore, we scanned nine journals and four conferences proceedings from 2003 and 2011. The journals were *ACM Transactions on Software Engineering and Methodology* (TOSEM), *Empirical Software Engineering* (EMSE), *IEEE Computer* (IEEE Comp),

IEEE Software (IEEE SW), *IEEE Transactions on Software Engineering* (TSE), *Information and Software Technology* (IST), *Journal of Systems and Software* (JSS), *Software Maintenance and Evolution* (SME), and *Software: Practice and Experience* (SP&E). The conferences were the *ACM / IEEE Empirical Software Engineering and Metrics* (ESEM), the *IEEE International Symposium on Empirical Software Engineering* (ISESE), the *IEEE International Symposium on Software Metrics* (METRICS), and the *International Conference on Software Engineering* (ICSE). It is important to mention that in 2007, ISESE and METRICS merged into ESEM.

We added a new exclusion criteria to attend our research questions: the paper does not present threats to validity nor actions to address them. Furthermore, we used search engines for the following venues: ESEM, ICSE, IEEE Comp, IEEE SW, ISESE, IST, JSS, METRICS, and TSE. We used search engines in these journals and conferences because the number of published papers was too high for a manual review. We created the following search string:

- controlled experiment OR randomized experiment OR randomised experiment OR quasi-experiment OR quasi experiment OR formal experiment OR empirical study OR empirical evaluation OR empirical analysis.

This search string was based on terms identified in papers from the SLR of Sjoberg et al. (2005). As result, we identified 159 papers published in 9 journals and 4 conferences from 2003 and 2011 (see Table 3). By adding the total number of papers identified in our SLR (159) to the number of papers identified in Sjoberg's SLR (47), we obtained a total of 206 papers reporting on controlled experiments and threats to validity.

Table 3. Results of SLR: Jan. 2003 – Dez. 2011.

Journal/Conference	Filter 1	Filter 2	Final	Percentage
EMSE	225	109	42	26.42%
IST	157	115	26	16.35%
ESEM	117	77	21	13.21%
TSE	92	80	20	12.58%
ISESE	92	67	17	10.69%
JSS	159	126	15	9.43%
METRICS	23	17	7	4.40%
TOSEM	139	35	4	2.52%
ICSE	48	46	4	2.52%
SME	195	96	2	1.26%
SP&E	578	211	1	0.63%
IEEE Comp	8	6	0	0.00%
IEEE SW	15	5	0	0.00%
Total:	1848	990	159	100%

In the first filter, we read the title and abstract of the papers. If a paper satisfied at least one inclusion criterion (see Table 4), the paper was selected. Otherwise, the paper was discarded. In the second filter, we read sections of the paper that reported the experiment process. In this stage, if a paper satisfied at least one exclusion criterion, the paper was discarded. Otherwise, the paper was selected to data extraction.

We extracted the following data: (a) threats to validity; and (b) actions to address the identified threats. As a result, we built a list of threats to validity and control actions. Finally, we identified relationships between threats and control actions. All steps were reviewed by an independent researcher. We conducted meetings in order to address any disagreements.

Table 4. Inclusion and Exclusion Criteria

Inclusion criteria (a paper must…)	Exclusion criteria (a paper should not be…)
- Use of at least two treatment conditions. - Have subjects or teams as experimental units. - Measure performance of a Software Engineering task. - Be related to a cause-effect investigation, i.e., the use of a treatment had to precede the measure of an outcome.	- Editorials, prefaces, article summaries, interviews, news, reviews, correspondence, discussions, comments, reader's letters, and summaries of tutorials, workshops, panels, and poster sessions. - Correlation studies - studies that are based solely on calculations using existing data (e.g., from data mining), and evaluations of simulated teams based on data for individuals were excluded. - Multiple case studies - studies that use projects or companies as treatment groups, in which data is collected at several levels (treatment defined, but no experimental unit defined). - Papers that, at the outset, would not provide sufficient data for our analyses (e.g., summaries of research programs). - Usability experiments, because those are part of another discipline (human computer interaction). - Papers that do not describe threats to validity.

3. Results

In this section, we present the SLR results. First, we present an overview on the data obtained from the SLR. Second, we describe threats to validity and actions to address them (RQ1).

Controlled experiments, as defined in Section 1, are reported in 206 (10.56%) of the 1951 papers scanned for this SLR, see Table 5. EMSE, ESEM, ISESE and METRICS, which focus specifically on empirical software engineering, reported more than half of selected papers, 53.40%. In 1993, we did not identify papers which satisfied our inclusion criteria. We can observe that in 2010 there was a decrease of publications which address our RQs. One potential reason was that several controlled experiments published in IST did not report threats to validity (7 papers) or did not use people or teams as experimental units (4 papers), see exclusion criteria in Table 4.

Regarding the validity types described in the controlled experiments (see Table 6), it is possible to observe that most papers concerned internal validity. According to Wohlin et al. (2012), internal validity has the highest priority. Therefore, researcher should prioritize the identification and control of the largest number of threats to internal validity possible.

It is worth mentioning that the classification presented in Table 6 was reported by the authors of the respective papers. In 29 papers (row Unknown in Table 6), the authors did not classify the threats in any of the four validity types.

In summary, the SLR identified a total of:

- 36 threats to internal validity and 86 control actions;
- 9 threats to external validity and 23 control actions;
- 10 threats to construct validity and 21 control actions; and
- 8 threats to validity and 19 control actions.

The references with their respective instances of threats to validity and control actions are presented in Appendix A. In the next subsections we present details of the identified threats to validity and their respective control actions, organized by the validity types.

Table 5. Number of papers that report threats to validity in controlled experiments.

	Year																		
Journal/Conference	**94**	**95**	**96**	**97**	**98**	**99**	**00**	**01**	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**	**11**	**Total**
EMSE	0	0	1	4	5	1	2	2	1	2	3	8	7	4	4	7	2	5	58
IST	0	0	0	2	1	0	0	3	0	2	6	2	0	1	1	10	1	3	32
TSE	0	1	0	1	1	0	1	3	1	4	1	3	2	1	4	0	0	5	28
ESEM	-	-	-	-	-	-	-	-	-	-	-	-	-	6	2	2	4	7	21
JSS	0	0	0	0	0	0	4	1	0	4	1	2	2	1	3	2	0	0	20
ISESE	-	-	-	-	-	-	-	-	2	2	2	3	10	-	-	-	-	-	19
METRICS	-	0	1	0	0	0	0	3	1	2	5	0	-	-	-	-	-	-	12
ICSE	1	0	0	0	1	0	1	1	1	0	0	1	0	0	0	1	1	1	9
TOSEM	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	4
SME	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	2
SP&E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
IEEE Comp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IEEE SW	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	1	1	2	7	8	1	8	13	6	16	18	20	22	15	14	23	9	22	206

Table 6. Validity Types per Paper

	Year																		
Validity Type	**94**	**95**	**96**	**97**	**98**	**99**	**00**	**01**	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**	**11**	**Total**
Internal	1	1	2	6	8	1	8	11	5	13	15	16	16	14	11	20	7	17	172
External	1	1	2	6	8	1	8	11	5	13	15	15	16	14	10	20	6	18	170
Construct	0	0	0	1	0	0	1	5	2	8	9	6	9	10	7	19	6	11	94
Conclusion	0	0	0	0	0	0	1	2	2	6	5	3	5	6	4	14	5	6	59
Unknown	0	0	0	1	0	0	0	2	1	2	2	4	5	1	3	3	1	4	29

3.1. Internal Validity

In this section, we present threats to internal validity extracted from papers scanned for this SLR. In total, 36 threats to internal validity were identified. Furthermore, we extracted control actions for each threat. We identified a total of 86 control actions.

To facilitate the identification of threats to validity and control actions during the reading, we defined an identifier for each one. For instance, INT-T03, where:

- INT is the validity type, in this case internal. Other possibilities for this prefix are: EXT (external), COT (construct), and COS (conclusion);
- T informs that the identifier is related to a threat to validity. In case of a control action this letter would be C;
- 03 is a sequential number, in this case informing that the identifier is related to the third identified threat to internal validity.

The first threat to internal validity is INT-T01. This threat occurs during subject selection. Generally, subjects have different experience levels. This fact can cause a risk for an experiment. Perepletchikov and Ryan (2011) presented concerns in selecting subjects with comparable experience. We identified five actions to control this threat (see Table 7).

Table 7. Threat INT-T01 and its control actions.

Threat description	Control action description
- INT-T01 – Differences among subjects related to experience.	- INT-C01 – To characterize subjects' experience through questionnaire. - INT-C03 – To assign subjects randomly to groups. - INT-C04 – To characterize subjects' experience through pretest. - INT-C16 – To assign treatments randomly to subjects. - INT-C25 – To group subject according to their experience level.

The control actions presented in Table 7 may be applied together or individually. Actions INT-C01 and INT-C04 aim at getting subject information. These actions were applied by Perepletchikov and Ryan (2011) through a user profile questionnaire and a pretest task. Thus, researchers can form balanced groups (applying the balancing design principle). On the other hand, actions INT-C03 and INT-C16 address the randomization design principle. Mouchawrab et al. (2011) state that the subjects were randomly selected from different blocks to form groups. Biffl and Halling (2003) also randomly assigned subjects to a defect detection approach. Finally, action INT-C25 aims at grouping subjects according to experience level (applying the blocking design principle). Biffl and Halling (2003) also grouped subjects to one of three student populations: students with excellent skills, students with medium skills, and students with little skills.

The second identified threat is INT-T02. The experiment results may be influenced by the level of familiarity with the experimental material. Tonella and Ceccato (2005), for instance, mention the level of familiarity with Java and JDK. Five control actions were identified for this threat (Table 8).

Actions INT-C39 and INT-C60 aim at establishing a familiarity level among subjects with the experimental material. Vokac et al. (2004), for instance, applied a familiarization task to all subjects. Knodel et al. (2008) explained all the windows and views needed for the experiment. Actions INT-C17 and INT-C69, on the other hand, are applied in order to provide familiarity with the experimental material. Lucia et al. (2009) organized training before the experiment aiming to give the students acceptable

familiarity. In the experiment presented by Babar et al. (2007), subjects used the material for more than six weeks in order to get familiarity. Finally, action INT-C30 suggests to use experimental material familiar to all subjects (Berling and Thelin, 2004).

Table 8. Threat INT-T02 and its control actions.

Threat description	Control action description
- INT-T02 – Familiarity with experimental material.	- INT-C17 – To provide training to improve familiarity with experimental material. - INT-C30 – To use familiar experimental material to all subjects. - INT-C39 – To apply pretest in order to level the subjects' familiarity with experimental material. - INT-C60 – To explain the experimental material. - INT-C69 – To establish long time of use experimental material.

The third identified threat, INT-T03, concerns the influence of learning effects. Subjects can acquire knowledge by using the experimental material and applying treatments. Briand et al. (2005) state that subjects would remember characteristics of treatments from previous experiment runs. Thirteen control actions were identified for this threat, see Table 9.

Table 9. Threat INT-T03 and its control actions.

Threat description	Control action description
- INT-T03 – Subject can acquire knowledge during the experiment.	- INT-C09 – To use different material in different runs. - INT-C10 – To establish design that balances learning. - INT-C13 – To establish short execution time. - INT-C19 – To establish order of material use. - INT-C21 – To define criteria for interaction between researcher and subjects. - INT-C23 – To allow material access only during the experiment. - INT-C26 – To conduct only one run with one treatment. - INT-C35 – To provide training to mitigate learning. - INT-C41 – To establish task order. - INT-C49 – To apply the treatment right after the training. - INT-C64 – To formulate tasks with short execution times. - INT-C81 – To formulate independent tasks. - INT-C83 – To conduct pilot study to verify learning provided by experimental material.

Actions INT-C41, INT-C64 and INT-C81 aim at mitigating learning provided by the designated tasks. If there is a relationship between tasks, for instance, this fact may result in the undesired subject learning effect (Canfora et al., 2006). On the other hand, actions INT-C09, INT-C19, INT-C23 and INT-C83 aim at minimizing learning provided by the experimental material. When subjects use the same experimental material in more than one run, their performance may also be affected. For example, Mouchawrab et al. (2011) used different experimental material in each runs conducted by each group. Similarly, if subjects access experimental material during intervals (Mouchawrab et al., 2011), they may improve their knowledge about the experimental material content. Actions INT-C26 and INT-C49 aim at controlling learning provided by treatments. As well as presented previously, subjects can study the treatment during experiment intervals. This fact may bias the experiment.

Another way to control threat INT-T03 is to reduce the execution time (INT-C13). In this case, subjects have little time to obtain knowledge that may influence their performance. Sabaliauskaite et al. (2003) stated that in their experiment threat INT-T03 did not occur because the experiment lasted no longer than 2.5 hours. Action INT-C10 aims at avoiding learning provided by the experimental design. Wohlin et al. (2012)

present a list of designs that can be used. The researcher is responsible to choose an appropriate design. Action INT-C21 aims at controlling learning provided by the researcher. This may occur through interaction between researcher and subjects. Subjects can get information which improves their performance. Heijstek et al. (2011) interacted with subjects by means of a strict interaction protocol to enable reacting to questions or remarks in a uniform way. Finally, action INT-C35 aims at reducing learning effects, balancing it with training (Lucia et al., 2009).

Table 10 presents the threat INT-T04 and its control actions. If the experimental material is different for subject, results may be biased. Ras and Rech (2009) mentioned differences in experimental material as a relevant risk in controlled experiments. We extracted six control actions for this threat.

Table 10. Threat INT-T04 and its control actions.

Threat description	Control action description
- INT-T04 – Differences among the experimental material.	- INT-C02 – To use the same experimental material for all subjects. - INT-C11 – To formulate experimental material as equivalent as possible. - INT-C15 – To review experimental material (different people, if possible). - INT-C24 – To conduct pilot study in order to evaluate experimental material. - INT-C65 – To select experts to formulate experimental material. - INT-C71 – To select experimental material used in previous studies.

Action INT-C02 suggests that all subjects use same experimental material. Therefore, any bias provided by experimental material would be the same for all subjects. In the experiment presented by Lucia et al. (2011), subjects used the same experimental material. If it is necessary formulate different experimental material, action INT-C11 proposes to minimize the differences. In this context, Bernardez et al. (2004) prepared the experimental material thoroughly with a single difference. Furthermore, experimental material must be reviewed (INT-C15) and formulated by experts (INT-C65). Deligiannis et al. (2004) had the experimental material reviewed by two independent professionals. It is desirable also to conduct a pilot study to evaluate experimental material (INT-C24), as done by Land et al. (2005). Finally, action INT-C71 suggests using experimental material from previous studies. This suggestion considers that previous studies had already evaluated and validated experimental material. Petersen et al. (2008) state that experimental material used in several former experiments is thoroughly tested.

Trouble around precision of data collection is the concern of threat INT-T05. Data may not be precise if subjects are responsible to collect them. An example is the experiment described by Biffl and Halling (2003) where subjects take all their notes on paper. We identified five control actions for this threat (see Table 11).

Table 11. Threat INT-T05 and its control actions.

Threat description	Control action description
- INT-T05 – Precision of collected data.	- INT-C38 – To use a tool for data collection. - INT-C44 – To reward subject's performance. - INT-C50 – To confront data collected manually and automatically, discarding data in case of inconsistencies. - INT-C70 – To ask subjects to be precise in data collection. - INT-C75 – To confront data collected by subjects and observers, discarding data in case of inconsistencies.

Action INT-C38 suggests the use of a tool to support data collection, as used by Abrahão et al. (2003). By using tool support, the precision on data collection may be improved. Researcher can also offer a reward for subject's performance (INT-C44). Bandi et al. (2003), for instance, rewarded subjects on the correctness of their answers. Being rewarded subjects may conduct data collection more carefully. Actions INT-C50 and INT-C75 aim at analyzing consistency of collected data based on two different perspectives. If there are inconsistencies, data should be discarded. In the experiment conducted by Biffl and Halling (2003), supervisors checked data collected from subjects and from a tool. Finally, action INT-C70 suggests that researchers explicitly ask subjects to be as precise as possible on the data of interest (Conte et al., 2007).

Threat INT-T06, as INT-T01, regards subject selection. The level of subject's ability has variation. Besides worrying about experience, Perepletchikov and Ryan (2011) selected subjects with comparable ability. We identified seven actions to control this threat (see Table 12).

Table 12. Threat INT-T06 and its control actions.

Threat description	Control action description
- INT-T06 – Differences among subjects related to ability.	- INT-C03 – To assign subjects to groups randomly. - INT-C07 – To characterize subjects' ability through questionnaire. - INT-C08 – To group subjects according to their ability level. - INT-C16 – To assign subjects to treatments randomly. - INT-C27 – To select subjects based on historical data. - INT-C33 – To characterize subjects' ability through pretest. - INT-C37 – To provide training to level subjects' ability.

Actions INT-C07, INT-C27 and INT-C33 aim at obtaining subject information to form balanced groups (balancing design principle). Actions INT-C07 and INT-C33 were applied by Perepletchikov and Ryan (2011) through a User Profile questionnaire and a pretest task. Actions INT-C03 and INT-C16 are related to applying the randomization principle. Mouchawrab et al. (2011) state that subjects were randomly selected from different blocks to form groups. Biffl and Halling (2003) randomly assigned subjects to a defect detection approach. Action INT-C08 aims at grouping subjects according to ability level (blocking design principle). Lucia et al. (2011) classified the subjects as having low and high ability. Finally, action INT-C37 aims at establishing a balance between subjects through training, as done by Briand et al. (2011).

Threat INT-T07 and its actions are presented in Table 13. This threat occurs if subjects drop out of the experiment (El-Attar and Miller, 2009). Four control actions were identified.

Table 13. Threat INT-T07 and its control actions.

Threat description	Control action description
- INT-T07 – Subjects may drop out of the experiment.	- INT-C05 – To assign subjects to groups at the day of the experiment. - INT-C13 – To establish short execution time. - INT-C14 – To discard incomplete or incorrect data. - INT-C45 – To select volunteer subjects. - INT-C58 – To reward subjects' participation.

Action INT-C05 aims at forming groups on the day of the experiment. In this way, if any subject drops out of the experiment, groups will not be harmed. Pfahl et al. (2003), for instance, assigned groups only on the second day of the experiment, i.e., directly before the treatment. Action INT-C45 suggests the selection of volunteer subjects. El-Attar and Miller (2009) selected volunteer subjects in order to avoid mortality. The authors consider that volunteer subjects hardly abandon experiments. Action INT-C13

aims at establishing a short execution time. Thus, the possibility of subjects' dropping out is reduced. Thelin et al. (2001) handled this threat applying the experiment during 3h. It is also possible to discard incomplete or incorrect data (INT-C14), as done by Andersson et al. (2003). Finally, researcher can reward subjects' participation in the experiment (INT-C58). Thelin et al. (2003) gave grade in a course to reward subject participation.

Table 14 presents threat INT-T08, related to subject communication during the experiment. The exchange of information during the experiment may affect subjects' performance, as mentioned by Lucia et al. (2011). We identified five control actions for this threat.

Table 14. Threat INT-T08 and its control actions.

Threat description	Control action description
- INT-T08 – Subjects communication during the experiment.	- INT-C44 – To reward subjects' performance. - INT-C54 – To distribute subjects with enough space between them. - INT-C56 – To introduce an observer in the experimental environment. - INT-C66 – To ask subjects to not communicate during the experiment. - INT-C67 – Not allowing sharing experimental material during the experiment. - INT-C85 – To discard data of subjects who communicated.

Huang and Holcombe (2009) introduced a competition scheme in the experiment to avoid sharing. They informed that the top three teams would receive prizes and that the best team would get the opportunity of joining a contest at IBM. These actions can be seen as rewarding the subjects' performance (INT-C44). By distributing subjects with enough space between then (INT-C54), researchers minimize chances of information sharing. Land et al. (2005) used a large room with sufficient physical space between subjects. Another action to control threat INT-T08 is introducing an observer in the experimental environment (INT-C56), as done by Lucia et al. (2011). Gupta and Jalote (2007) asked subjects to not communicate during the experiment (INT-C66) and did not allow sharing experimental material (INT-C67). Finally, action INT-C85 suggests discarding data if subjects communicate potentially affecting their performance, as done by Briand et al. (2005).

Threat INT-T09 regards lack of motivation (Thelin et al., 2003). For this threat, we identified seven control actions (see Table 15).

Table 15. Threat INT-T09 and its control actions.

Threat description	Control action description
- INT-T09 – Lack of motivation.	- INT-C06 – To participate is a mandatory part of a course. - INT-C32 – To select benefic activities. - INT-C44 – To reward subjects' performance. - INT-C45 – To select volunteer subjects. - INT-C46 – To select subjects from a real context. - INT-C55 – To inform subjects that future activity may be similar. - INT-C79 – To inform subjects that treatment may be benefic.

Action INT-C06 regards subjects' mandatory participation in the experiment (Thelin et al., 2004b) (e.g., so that subjects will able to finish a course). Another control action for this threat is rewarding subjects' performance (INT-C44) Bandi et al. (2003). Researchers can also select volunteer subjects (INT-C45). Cornelissen et al. (2011) assumed that volunteer subjects are properly motivated, since agreeing to participate usually means that they are motivated. Furthermore, another control action is selecting subjects from a real context (INT-C46). In this case, researcher can inform subjects that

the technology under evaluation may support their activities (Carver et al., 2008). Finally, actions INT-C32, INT-C55 and INT-C79 are related to procedures that may be useful to subjects in the future. Abrahão et al. (2003) motivated subjects by explaining to them that similar approaches could be useful for their projects.

Table 16 presents threat INT-T10. This threat arises when experimental material is inconsistent, ambiguous, incorrect, or incomplete. Arisholm et al. (2007), for instance, mentioned model misspecification as a threat. Five control actions were identified for this threat.

Table 16. Threat INT-T10 and its control actions.

Threat description	Control action description
- INT-T10 – Specification problems of experimental material.	- INT-C15 – To review experimental material (different people, if possible). - INT-C24 – To conduct a pilot study to evaluate experimental material. - INT-C43 – To search for experimental material in literature. - INT-C65 – To select experts to formulate experimental material. - INT-C71 – To select experimental material used in previous studies.

Actions INT-C43 and INT-C71 suggest that experimental material may be retrieved from literature or be the same used in previous studies. Petersen et al. (2008) used already evaluated and validated experimental materials from a former experiment. Experimental material should also be reviewed by different people (INT-C15), as done by Ferrari and Madhavji (2008). Furthermore, it is recommended that experimental material should be formulated by experts (INT-C65). The experimental material used by El-Attar and Miller (2009), for instance, was formulated by different professionals. Finally, action INT-C24 suggests that experimental material should be evaluated through a pilot study (Land et al., 2005).

We did not identify control actions for threat *"Different experimental environment"* (INT-T11) (Knodel et al., 2008). This threat occurs when subjects are assigned to different experimental environments, where facts that happen in the environment may affect the results.

Threat INT-T12 is presented in Table 17. If the experiment is too long, fatigue effects can affect subjects. This threat is mentioned by Mouchawrab et al. (2011). Nine control actions were identified for this threat.

Table 17. Threat INT-T12 and its control actions.

Threat description	Control action description
- INT-T12 – Fatigue effects can influence subjects' performance.	- INT-C13 – To establish a short execution time. - INT-C22 – To use different experimental material in more than one run. - INT-C28 – To allow breaks between runs. - INT-C31 – To conduct runs in consecutive days. - INT-C45 – To select volunteer subjects. - INT-C52 – To establish interaction with subjects. - INT-C64 – To formulate short tasks. - INT-C78 – To allow subjects to choose the experiment execution day. - INT-C82 – To conduct a pilot study in order to verify whether the execution time causes fatigue.

Action INT-C45 suggests that subjects selected should be volunteers. These subjects will be motivated to participate in the experiment (El-Attar and Miller, 2009). Actions INT-C13 and INT-C64 aim at establishing short execution times for the experiment

runs. Canfora et al. (2007) conducted a short execution time experiment to avoid subjects' fatigue. Researchers can also interact with subjects (INT-C52) (Hayes and Offutt, 2006). As well as lectures, since without interaction experiments can become boring. If the experiment has more than one run, it is suggested that the experimental material should be different (INT-C22), as done by Mouchawrab et al. (2011). The researcher can also allow breaks during the experiment (INT-C28) or conduct runs in consecutive days (INT-C31). Briand et al. (2005) allowed breaks between runs. In addition, researchers can allow subjects to choose the experiment execution date (INT-C78) (Berling and Thelin, 2004) providing convenience. Finally, action INT-C82 suggests conducting a pilot study to check whether execution time causes fatigue, as done by Mendonça and Oliveira (2006).

Table 18 presents threat INT-T13, which may occurs if an experiment has breaks. During those breaks, everyday facts may happen and affect subjects' performance (Berling and Thelin, 2004). We identified two control actions for this threat.

Table 18. Threat INT-T13 and its control actions.

Threat description	Control action description
- INT-T13 – Historical facts may change subjects' performance during experiments with long breaks.	- INT-C13 – To establish a short execution time. - INT-C26 – To conduct only one run with one treatment.

Action INT-C13 suggests that the researcher should establish a short execution time, as done by Thelin et al. (2001), so that the subjects would not need any breaks. Thelin et al. (2001) also conducted only one run (INT-C26), avoiding undesired historical facts to occur.

Table 19 presents threat INT-T14, which occurs when subjects exchange information during breaks (not during experiment execution, as in threat INT-T08). Mouchawrab et al. (2011) mentioned diffusion or imitation of treatments as threats to internal validity. For this threat we identified two control actions.

Table 19. Threat INT-T14 and its control actions.

Threat description	Control action description
- INT-T14 – Communication between subjects during breaks.	- INT-C23 – To allow access to experimental material only during experiment. - INT-C29 – To ask subjects to not communicate during breaks.

Action INT-C23 suggests that experimental material should be available to subjects only during the experiment (Mouchawrab et al., 2011), avoiding information exchange. Finally, action INT-C29 suggests researcher to ask subjects to not communicate during breaks (Robillard and Murphy, 2007), so that information about treatments and the use of experimental material will not be shared.

Threat INT-T15 is presented in Table 20. Trainings may be unbalanced: differences among training material, duration, and explanations can affect subjects' understanding (Hochstein et al., 2008). We identified only one control action for this threat.

Table 20. Threat INT-T15 and its control action.

Threat description	Control action description
- INT-T15 – Unbalanced training.	- INT-C57 – To establish procedures for consistent training.

By establishing training procedures (INT-C57), information provided to subjects is more likely to be consistent and balanced. Conte et al. (2007), for instance, prepared equivalent training courses.

Subjects' language being different from the language of the experimental material (INT-T16) also represents a potential threat (Genero et al., 2007). For this threat, we identified three control actions (see Table 21).

Table 21. Threat INT-T16 and its control actions.

Threat description	Control action description
- INT-T16 – Subjects' language different from language of experimental material.	- INT-C36 – To use a translation support. - INT-C51 – To apply pretest in order to verify subjects' knowledge in the language of the experimental material. - INT-C74 – To select subjects familiar with the experimental material language.

Action INT-C36 suggests the researcher to allow the use of some support. Sandahl et al. (1998), for instance, allowed subjects to use a dictionary. Action INT-C51 aims at obtaining information about subject's knowledge, as done by Dag et al. (2006), providing useful information to decide whether subjects are able participate in the experiment. Finally, action INT-C74 suggests that researcher select subjects familiar with the experimental material language (Qattous et al., 2010).

We did not identify any control actions for threat *"Subjects selection based on historical data"* (INT-T17) and *"Pretest application can change subject's performance"* (INT-T18). The first one occurs when a researcher select subjects based on historical data (Mouchawrab et al., 2011). In this case, researcher may cause bias if select people with better results. The second one occurs when a pretest activity influences experiment execution (Jansen et al., 2009), with subjects getting information that may affect their performance.

Table 22 presents threat INT-T19, which may occur when researcher conduct an experiment replication. If the original experimental package is changed, those changes may influence the treatment results. Maldonado et al. (2006), for instance, made some changes in the experimental procedures during their replication. We identified only one control action for this threat.

Table 22. Threat INT-T19 and its control action.

Threat description	Control action description
- INT-T19 – Changing original experimental package, in case of replication.	- INT-C47 – To conduct a pilot study to understand experimental package.

Action INT-C47 suggests researchers to conduct a pilot study to understand the experimental package (Maldonado et al., 2006), getting a clearer picture of potential impacts caused by changes. Using pilot studies, researchers can check which changes can safely performed without major implications.

Table 23 presents threat INT-T20, which may happen when an experiment has limited time to accomplish tasks (Mouchawrab et al., 2011). In this situation, subject may perform tasks fast in order to attend the fixed established execution time. Therefore, tasks may become error prone. We identified two control actions for threat INT-T20.

Table 23. Threat INT-T20 and its control actions.

Threat description	Control action description
- INT-T20 – Establish time to experiment execution.	- INT-C61 – To conduct a pilot study to check if experiment has enough execution time. - INT-C63 – To establish the same execution time to all subjects.

Action INT-C61 aims at conducting a pilot study in order to assess adequacy of the established time (Cornelissen et al., 2011), so that the researcher can evaluate if he should adjust the execution time or even the provided tasks. Action INT-C63 suggests that all subjects should have the same experiment execution time. Therefore, the time effect on the results would be equivalent to each subject (Knodel et al., 2008).

Threat INT-T21 occurs when experimental tasks, which subjects have to perform, are different (Deligiannis et al., 2004). In this scenario, treatment results may be affected by differences among tasks. For this threat, we identified four control actions (see Table 24).

Table 24. Threat INT-T21 and its control actions.

Threat description	Control action description
- INT-T21 – Differences among experimental tasks.	- INT-C12 – To formulate tasks as equivalent as possible. - INT-C53 – To use the same tasks for all subjects. - INT-C62 – To conduct a pilot study to assess tasks adequacy. - INT-C77 – To formulate tasks based on literature.

Action INT-C12 suggests that tasks should be as equivalent as possible. Bunse (2006), for instance, ensured that the tasks are highly similar. It is also possible to use the same experimental tasks for all subjects (INT-C53) (Lucia et al., 2011), when appropriate. In addition, action INT-C62 aims at conducting a pilot study in order to assess the adequacy of different tasks, as done by Perepletchikov and Ryan (2011). Finally, Wettel et al. (2011) formulated tasks based on literature (INT-C77), improving task reliability.

Threats INT-T22, INT-T23, INT-T24 and INT-T26 have only one control action each (see Table 25). The first one occurs when a group gives its best to show that a certain treatment is competitive (Thelin et al., 2003). Thus, results may present bias, because subjects did not behave naturally. We identified action INT-C34 to address threat INT-T22. This action suggests that researcher informs that all subjects will use the same treatments, as mentioned by Thelin et al. (2003).

Table 25. Threats INT-T22, INT-T23, INT-T24, INT-T26 and their control actions.

Threat description	Control action description
- INT-T22 – Competition between groups.	- INT-C34 – To inform that all subjects will use the same treatments.
- INT-T23 – Subjects may have performed similar studies previously.	- INT-C18 – To select subjects without previous experience.
- INT-T24 – Presentation order of experimental material to subjects.	- INT-C20 – To establish presentation order of experimental material.
- INT-T26 – Training may be insufficient.	- INT-C48 – To allow subjects ask questions.

If subjects had performed similar studies previously (INT-T23), there is a threat for the experiment (Canfora et al., 2007), since experience might improve subjects' performance. We identified action INT-C18 to control threat INT-T23. This action suggests that researcher get information about subjects' experience in experiments (Abrahão et al., 2003).

Threat INT-T24 arises when presentation order of experimental material may affect subjects' performance, as mentioned by Ras and Rech (2009). We identified action INT-C20 to control this threat, by establishing the presentation order of the experimental material (Ras and Rech, 2009).

Table 25 also presented threat INT-T26. This threat occurs when training does not provide all necessary information about a treatment (Budgen et al., 2011). Then, experiment execution may be affected. For this threat, we identified only one control

action (INT-C48). This control action allows subjects to ask questions (Staron et al., 2006). Thus, subjects can clarify any doubts about the experiment execution.

Table 26 presents threat INT-T25, which happens if subjects do not follow the process correctly (as described in the experiment planning), potentially harming experiment results. Biffl and Halling (2003) identified this threat in their experiment. We extracted four control actions for this threat.

Table 26. Threat INT-T25 and its control actions.

Threat description	Control action description
- INT-T25 – Lack of conformity with process to be followed.	- INT-C56 – To introduce observers in the experimental environment. - INT-C59 – To establish guidelines to support subjects. - INT-C68 – To provide training on experiment execution conformity. - INT-C86 – To discard data from subjects who did not follow the process.

Action INT-C56 suggests introducing researchers in the experimental environment. Researchers should interfere if they observe any shortcomings. Biffl and Halling (2003) applied this action to ensure process adherence. It is also possible to establish guidelines (INT-C59) in order to avoid subjects from deviating from the process. Muller (2004) enforced the process rigorously and left the subjects no possibility for variation. In addition, subjects can receive training in order to provide conformity (INT-C68) (Itkonen et al., 2007). Finally, action INT-C86 aims at discarding data from subjects who presented unconformity (Arisholm et al., 2006).

We did not identify any control actions for threat *"Reward for participation, not for performance"* (INT-T27) (Thelin et al., 2004b). By establishing rewards not related to performance, subjects may not act in serious and correct ways. Consequently, results may not represent the real subjects' performance.

Table 27 presents threat INT-T28, related to subjects that do not have enough technical skills to perform the experiment. This was stated as a major threat in the experiment conducted by Ahmed et al. (2005). We identified three control actions for this threat.

Table 27. Threat INT-T28 and its control actions.

Threat description	Control action description
- INT-T28 – Technical factors may cause difficulties for subjects.	- INT-C40 – To discard data of subjects with technical problems. - INT-C42 – To select experienced subjects. - INT-C48 – To allow subjects ask questions.

Action INT-C40 suggests that data from subjects who faced technical problems should be discarded. Erdogmus et al. (2005) excluded data from subjects who did not provide reasonable answers. To avoid this threat, researcher can also select experienced subjects (INT-C42), as done by Svahnberg and Wohlin (2005). Furthermore, Cornelissen et al. (2011) allowed subjects to ask questions to address any doubts (INT-C48).

Threats INT-T29, INT-T33, and INT-T36 are presented in Table 28. The first one occurs when subjects are trained by different people. Thus, the training may provide different information to each group (Muller, 2004). For this threat, we identified only one control action, establishing procedures to conduct training (INT-C57) to assure the information transmitted to subjects is consistent. Conte et al. (2007) prepared equivalent training courses.

Threat INT-T33 may happen if subjects perform offline experiments (e.g., at home), allowing them to use others information sources to perform experimental tasks

(Chatterji et al., 2011). The identified control action was INT-C76, avoiding that subjects consult the internet, as done by Chatterji et al. (2011).

Table 28 also presents threat INT-T36, which arises when tasks are not adequate to be executed in the experiment (Knodel et al., 2008). We also identified only one action for this threat. Action INT-C62 aims at conducting a pilot study in order to assess the adequacy of the tasks and their influence on the experiment results (Knodel et al., 2008).

Table 28. Threats INT-T29, INT-T33, INT-T36 and their control actions.

Threat description	Control action description
- INT-T29 – Different people training subjects.	- INT-C57 – To establish procedures for consistent training.
- INT-T33 – Subjects can look for answer elsewhere.	- INT-C76 – Not allowing internet access.
- INT-T36 – Tasks adequacy to be performed in the experiment.	- INT-C62 – To conduct a pilot study to assess tasks adequacy.

We did not identify any control action for threat *"Differences among subjects related to gender"* (INT-T30) (Salleh et al., 2010). This threat occurs when subject gender may cause bias to the experiment.

Threat INT-T31 is presented in Table 29. This threat occurs when researchers unconsciously (or consciously) influence training. Thus, training to a group can be better than to another (Muller, 2005). We identified three control actions for this threat.

Table 29. Threat INT-T31 and its control actions.

Threat description	Control action description
- INT-T31 – Training conducted by only one person.	- INT-C57 – To establish procedures for consistent training. - INT-C84 – To provide training with different instructors.

Action INT-C57 can be applied to threat INT-T31 with the same purpose as to threat INT-T15. Furthermore, action INT-C84 suggests training to be provided by more than one instructor (Muller, 2005). Thus, the influence of the researcher on different groups is minimized.

Threat INT-T32 is presented in Table 30. This threat occurs when subjects do not finish the experimental tasks (Cornelissen et al., 2011) and when procedures for experiment execution are not followed. For this threat, we identified two control actions.

Table 30. Threat INT-T32 and its control actions.

Threat description	Control action description
- INT-T32 – Incomplete or incorrect data collection.	- INT-C14 – To discard incomplete or incorrect data. - INT-C80 – To provide examples on how to fill out the data collection form.

Action INT-C14 suggests that incomplete or incorrect data should be discarded (Andersson et al., 2003). Thus, researcher avoid that this data influences the outcome. Action INT-C80 aims at explaining how to fill out the data collection forms (Canfora et al., 2006). Thus, subjects can clarify any doubts about using the experimental material.

We did not identify control actions for threat *"Only part of the subjects receive reward"* (INT-T34). Muller and Hofer (2007) mentioned this threat. Professionals might receive reward to participate in the experiment. Students, on the other hand, may participate in the experiment because it is a mandatory part of a course.

Threat INT-T35 is presented in Table 31. This threat occurs when subjects' personality influences the experiment (Salleh et al., 2009). For this threat, we identified two control actions.

Table 31. Threat INT-T35 and their control actions.

Threat description	Control action description
- INT-T35 – Differences among subjects related to personality.	- INT-C72 – To group subject according to personality profiles. - INT-C73 – To characterize subjects' personality through questionnaire.

Action INT-C72 aims at grouping subjects according to their personality profiles (blocking design principle) (Salleh et al., 2009). Action INT-C73 aims at getting information about subjects in order to form balanced groups (balancing design principle), as done by Salleh et al. (2009).

3.2. External Validity

In this section, we present threats to external validity identified through the SLR. At all, 9 external validity threats and 23 control actions were identified.

The first threat to external validity, EXT-T01, occurs when an experimental sample of subjects does not represent the intended population (e.g., in terms of knowledge, experience, ability, or size). Lucia et al. (2011) identified this threat in their experiment. We extracted six control actions for this threat (see Table 32).

Table 32. Threat EXT-T01 and its control actions.

Threat description	Control action description
- EXT-T01 – Subjects' representativeness.	- EXT-C01 – To select subjects from the real context. - EXT-C02 – To select subjects with similar knowledge of people from the real context. - EXT-C03 – To select subjects who will start career in the real context. - EXT-C04 – To select subjects with experience in the real context. - EXT-C13 – To provide training to all subjects. - EXT-C21 – To select heterogeneous subjects.

Actions EXT-C01, EXT-C02, EXT-C03, and EXT-C04 suggest that researcher select subjects who have some experience in the real context: (1) subjects involved in the real context (Dzidek et al., 2008); (2) subjects who already were involved in the real context (Damian et al., 2008); and (3) subjects with similar knowledge of people that take part in the real context (Lucia et al., 2011). Action EXT-C21 aims at selecting subjects with different knowledge levels. Arisholm and Sjoberg (2004), for instance, used professionals and students as subjects. Finally, action EXT-C13 aims at providing training to the subjects, so that they can get similar characteristics of people who are involved in the real context (Mouchawrab et al., 2011).

The second threat to external validity, EXT-T02 is presented in Table 33. This threat occurs when artifacts from the experimental context are poor representatives for real artifacts in terms of size and complexity, as presented by Perepletchikov and Ryan (2011). Five control actions were identified.

Action EXT-C14 suggests artifacts to be selected from literature (Mouchawrab et al., 2011). Artifacts presented in literature may be validated and representative. Researcher can also select artifacts from the real context (EXT-C06), as done by Biffl and Halling (2003). In addition, action EXT-C08 suggests identifying artifacts commonly used in the real context (Mouchawrab et al., 2011) to improve representativeness. Researcher can also formulate artifacts based on the real context (EXT-C09), as presented by

Lanubile et al. (2004). Finally, experts can check artifacts representativeness (EXT-C23) (Land et al., 2005).

Table 33. Threat EXT-T02 and its control actions.

Threat description	Control action description
- EXT-T02 – Artifacts representativeness.	- EXT-C06 – To select artifacts from the real context. - EXT-C08 – To select artifacts commonly used in the real context. - EXT-C09 – To formulate artifacts based on the real context. - EXT-C14 – To use artifacts from literature. - EXT-C23 – To send the artifacts to be analyzed by experts.

Threat EXT-T03 is presented is Table 34. This threat occurs when the domain of artifacts does not represent the domain of the real experiment context (Vokac et al., 2004). We identified only one action to address this threat.

Table 34. Threat EXT-T03 and its control action.

Threat description	Control action description
- EXT-T03 – Artifacts domain representativeness.	- EXT-C15 – To use artifacts from different domains.

Action EXT-C15 aims at using artifacts from different domains (e.g., in replications) (Vokac et al., 2004) to increase representativeness.

Table 35 presents threat EXT-T04, which occurs when the process used in the experiment does not represent the process in the real context (Biffl and Halling, 2003). For this threat, we identified three control actions.

Table 35. Threat EXT-T04 and its control actions.

Threat description	Control action description
- EXT-T04 – Process representativeness.	- EXT-C07 – To select similar process to that used in the real context. - EXT-C10 – To use process described in literature. - EXT-C22 – To explain activities performed previously.

Action EXT-C07 suggests that the process used in the experiment should be similar to that used in the real context (Biffl and Halling, 2003). If it is necessary to use only part of the process, previous activities should be explained to subjects (EXT-C22), as done by Hovsepyan et al. (2011). Finally, action EXT-C10 suggests to select the process to be used in the experiment from literature (Vitharana and Ramamurthy, 2003), which, especially if used in previous studies, may be representative of the real context.

Threat EXT-T05 occurs when experimental tasks do not properly represent tasks of the real context (Lucia et al., 2011). For this threat, we identified two control actions (see Table 36).

Table 36. Threat EXT-T05 and its control actions.

Threat description	Control action description
- EXT-T05 – Task representativeness.	- EXT-C16 – To select tasks that may occur in the real context, according to literature. - EXT-C17 – To select tasks from the real context.

Action EXT-C16 suggests that researcher should get information on tasks which usually are part of the real context. Lucia et al. (2011), for instance, stated that in their experiment the tasks were realistic enough. Another possibility, action EXT-C17, is selecting tasks directly from the real context (Dzidek et al., 2008).

Table 37 presents threat EXT-T06. This threat occurs when treatments applied do not properly represent treatments commonly used in the real context (Muller, 2005). We identified two control actions for this threat.

Table 37. Threat EXT-T06 and its control actions.

Threat description	Control action description
- EXT-T06 – Treatment representativeness.	- EXT-C05 – To select treatment usually applied in the real context. - EXT-C18 – To use different treatments from those used in previous studies, in case of replications.

The researcher should use different treatments to those used in previous studies, in case of a replication (EXT-C18) to increase representativeness, because this way a higher number of treatments would be compared (Anda and Sjoberg, 2005). In addition, researchers may identify treatments commonly used in the real context (EXT-C05), as done by Sinha and Smidts (2006).

Threat EXT-T07 is presented in Table 38. This threat is specifically related to experiments in which it is necessary seed faults in artifacts (e.g., inspection experiments, as presented by Mouchawrab et al., 2011). In these experiments, the seeded faults may not be representative. As control action for this threat we identified EXT-C12, concerning selecting fault types from literature, which may allow to represent the real context more accurately. Ruthruff et al. (2006) used this control action in their experiment.

Table 38. Threats EXT-T07 and EXT-T09 and their control actions.

Threat description	Control action description
- EXT-T07 – Faults representativeness.	- EXT-C12 – To select faults from literature.
- EXT-T09 – Requirements representativeness.	- EXT-C19 – To use requirements from previous studies.

Table 38 also presents threat EXT-T09, which occurs when a requirement document is used as experimental artifact. The requirement document may not be representative (Dag et al., 2006). For this threat, we identified action EXT-C19 concerning using requirement documents from previous studies (Dag et al., 2006).

Threat EXT-T08 is presented in Table 39. When experimental environment does not satisfy the environment of real context, this threat may arise (Mouchawrab et al., 2011). We identified two control actions for threat EXT-T08.

Table 39. Threat EXT-T08 and its control actions.

Threat description	Control action description
- EXT-T08 – Experimental environment representativeness.	- EXT-C11 – To set an environment as real as possible. - EXT-C20 – To conduct the experiment in the environment of the real context.

Action EXT-C20 suggests that the experiment should be conducted in the environment of the real context (Dzidek et al., 2008). As an alternative, researcher may set up an environment based on the real environment (EXT-C11), as done by Ferrari and Madhavji (2008).

3.3. Construct Validity

In this section, we present the threats to construct validity identified in the SLR. We extracted 10 threats to construct validity and 21 control actions.

The first threat to construct validity is COT-T01, which happens when a treatment does not well represent the cause. Prechelt (2011) mentioned lack of support to use a treatment as a threat to their experiment. We identified only one control action for this threat (see Table 40). Action COT-C20 suggests using treatments described in literature (Arisholm and Sjoberg, 2004).

Table 40. Threat COT-T01 and its control action.

Threat description	Control action description
- COT-T01 – Treatment does not represent the cause.	- COT-C20 – To use treatment from literature.

The second threat is COT-T02 (see Table 41). This threat occurs when the metrics do not well represent the effect. Lucia et al. (2011) identified this threat. For this threat, we extracted three control actions.

Table 41. Threat COT-T02 and its control actions.

Threat description	Control action description
- COT-T02 – Metrics do not represent the effect.	- COT-C01 – To use metrics applied in previous studies. - COT-C02 – To use metrics described in literature. - COT-C03 – To use several metrics.

Actions COT-C01 and COT-C02 suggests metrics to be selected from previous studies (Lucia et al., 2011) or from literature (Ruthruff et al., 2006). These metrics are well defined and, generally, already validated trough experiments. Another possible approach is to use several metrics to check the effect of results (COT-C03), as done by Damian et al. (2008).

Threat COT-T03 is presented in Table 42. This threat arises when subjects use more than one treatment in the experiment. In these situations, the first treatment used can influence subjects' performance in subsequent treatments, as mentioned by Sfetsos et al. (2009). For this threat, we identified four control actions.

Table 42. Threat COT-T03 and its control actions.

Threat description	Control action description
- COT-T03 – Treatment interaction.	- COT-C07 – To establish breaks between treatment applications. - COT-C15 – To assign only one treatment to each subject. - COT-C17 – To establish the order of treatment application. - COT-C18 – To use treatments that do not influence each other.

Action COT-C07 suggests that researcher provide a break between treatment applications, as done by Briand et al. (2005), avoiding one treatment to affect the following. This threat may also be addressed by assigning only one treatment to each subject (COT-C15) (Sfetsos et al., 2009). Action COT-C17 aims at establishing the order of treatment application (Canfora et al., 2006), assuring that treatments which cause less influence on others are applied first. Of course, if possible, the ideal case would be using treatments that do not cause influence each other (COT-C18) (Denger and Kolb, 2006).

Table 43 presents threat COT-T04. This threat concerns subjects having their performance affected when observed by other people. Carver et al. (2008) identified this threat. Four control actions addressing it were identified.

Action COT-C08 suggests to observe all subjects. Ferrari and Madhavji (2008) stated that, if this causes any effect, the entire sample would be affected equally. It is also possible to apply pretests to adapt subjects to the environment (COT-C13) (Heijstek et

al., 2011), avoiding them to feel constrained during the experiment. Heijstek et al. (2011) formulated challenging tasks in order to keep subjects concentrated (COT-C14), distracting them from observers. Action COT-C06 suggests evaluating subjects' participation instead of their performance (Lucia et al., 2009), minimizing the apprehension of being observed. Finally, action COT-C21 suggests selecting volunteer subjects. Sfetsos et al. (2009) stated that volunteer subjects were not afraid of being observed and evaluated.

Table 43. Threat COT-T04 and its control actions.

Threat description	Control action description
- COT-T04 – Subjects behave differently when being observed.	- COT-C06 – To evaluate subjects' participation, not subjects' performance. - COT-C08 – To observe all subjects, then effect is similar. - COT-C13 – To apply pretest in order to adapt subjects to the environment. - COT-C14 – To formulate challenging tasks, keeping subjects concentrated. - COT-C21 – To select volunteer subjects.

Threat COT-T05 is presented in Table 44. This threat occurs when the researcher unconsciously (or consciously) affects the experiment outcomes (Ferrari and Madhavji, 2008). Two control actions were identified for this threat.

Table 44. Threat COT-T05 and its control actions.

Threat description	Control action description
- COT-T05 – Researchers' influence.	- COT-C04 – To include independent researchers in the experiment. - COT-C11 – Not to allow interaction between interested researcher and subjects.

Action COT-C04 aims at including independent researchers in the experiment. Ferrari and Madhavji (2008) used multiple researchers and domain experts in their study process. Action COT-C11 suggests that there is no interaction between the interested researcher and subjects (Jedlitschka, 2010). This would assure that no information affecting subjects' performance is transmitted.

Table 45 presents threat COT-T06. This threat occurs when subjects try to figure out the goal of the experiment and base their behavior on assumptions, either positively or negatively. Perepletchikov and Ryan (2011) mentioned that providing information to subjects could influence their performance in the experiment. We identified three control actions for this threat.

Table 45. Threat COT-T06 and its control actions.

Threat description	Control action description
- COT-T06 – Hypothesis guessing.	- COT-C05 – To select subjects who do not have knowledge about the hypotheses. - COT-C09 – Not to allow communication among subjects. - COT-C10 – Not to inform experiment details.

Action COT-C05 suggests that the researcher selects subjects who do not have any knowledge about the experiment hypotheses. Sfetsos et al. (2009) selected subjects who were not aware of what the researchers intended to investigate. Additionally, researcher should not allow communication among the subjects (COT-C09). Petersen et al. (2008) did not allow subjects to communicate about the experiment, avoiding disseminating individual opinions. Finally, the researcher should also not inform experiment details

(COT-C10) (e.g., expected outcomes) to subjects, as done by Perepletchikov and Ryan (2011).

Threat COT-T07 is presented in Table 46. This threat may arise when only one metric is defined for measuring the results (Briand et al., 2005). Therefore, it will be difficult to detect if this metric is biased. We identified only one control action for this threat. Action COT-C03, which suggests using several metrics (Damian et al., 2008). By using more than one metric, results can be verified through cross-checking.

Table 46. Threat COT-T07 and its control action.

Threat description	Control action description
- COT-T07 – Using only one metric.	- COT-C03 – To use several metrics.

Table 47 presents threat COT-T08, related to the potential risk of not representing the construct of the cause properly when only one experimental artifact (or task, or treatment) is used. Erdogmus et al. (2005) used only one experimental task. We identified two control actions for this threat.

Table 47. Threat COT-T08 and its control actions.

Threat description	Control action description
- COT-T08 – Using only one experimental artifact, or task, or treatment.	- COT-C12 – To introduce more than one experimental artifact (or task, or treatment). - COT-C16 – To select variations of experimental artifacts (or tasks, or treatments).

Actions COT-C12 and COT-C16 suggest that researchers formulate more than one experimental artifact (Bernardez et al., 2004) and select variations (Michalik et al., 2011). This way, the cause construct may be better represented.

Table 48 presents threat COT-T09. This threat may arise when subject unconsciously (or consciously) fill the characterization forms incorrectly (Arisholm et al., 2007). For this threat, we identified one control action. Action COT-C19 aims at using historical data in order to get characterization form consistence (Lucia et al., 2009).

Table 48. Threat COT-T09 and its control action.

Threat description	Control action description
- COT-T09 – Subjects may fill the characterization form incorrectly.	- COT-C19 – To use subjects' historical data.

We did not identify any control actions for threat *"Subjects' characterization metrics may not be correct"* (COT-T10). This threat may arise when the characterization form has metrics that do not affect subjects' performance (Arisholm et al., 2007). Carver et al. (2004) mentioned misclassification of subjects as a potential experiment risk. They also cited the possibility of using an invalid subject grouping criterion.

3.4. Conclusion Validity

In this section, we present the threats to conclusion validity identified in the SLR. We extracted 8 threats to conclusion validity and 19 control actions.

The first threat, COS-T01, arises when measures used in the experiment are not reliable (Briand et al., 2005) and when data analysis is not clear. For this threat, we identified five control actions (see Table 49).

Action COS-C07 suggests that researcher selects measures from previous studies, as done by Lucia et al. (2011). It is also possible to define a data measuring method (COS-C02) (Cornelissen et al., 2011). In addition, researcher may conduct a study in order to

evaluate and to validate the defined method (COS-C16), as presented by Karlsson et al. (2007). By planning how data should be measured subjective evaluation is avoided. Action COS-C03 aims at using more than one person to analyze the collected data (Arisholm and Sjoberg, 2004). In case of disagreements in data analysis, meetings should be conducted to address them (COS-C05). Lucia et al. (2011) conducted meetings to discuss and resolve such discrepancies.

Table 49. Threat COS-T01 and its control actions.

Threat description	Control action description
- COS-T01 – Measure reliability.	- COS-C02 – To define a data measuring method. - COS-C03 – To assign different people to analyze collected data. - COS-C05 – To conduct meetings in order to address disagreements. - COS-C07 – To use measures from previous studies. - COS-C16 – To conduct studies to evaluate the data measuring method.

The second threat is COS-T02 (see Table 50). This threat occurs when an inadequate statistical test is applied on the data collected (Lucia et al., 2010). There are many factors which influence a statistical test (e.g., sample size, data distribution, and data homoscedasticity). We identified three control actions for this threat.

Table 50. Threat COS-T02 and its control actions.

Threat description	Control action description
- COS-T02 – Violated assumption of statistical test.	- COS-C01 – To check requirements of statistical test. - COS-C10 – To use statistical test from previous studies. - COS-C19 – To use a statistical test tool.

Action COS-C01 aims at verifying statistical test usage requirements (Damian et al., 2008), avoiding wrong decisions on which test to apply. Researchers may also use statistical test applied in similar previous studies (COS-C10) (Ricca et al., 2010) and tool support (COS-C19) (Madeyski, 2010).

Table 51 presents threat COS-T03, which occurs when the selected samples is heterogeneous, as presented by Lucia et al. (2011). In these situations, differences among subjects related to experience may exists and influence the experiment results. We identified only one control action for threat COS-T03. Action COS-C04 suggests selecting homogeneous subjects (with similar background), as done by Falessi et al. (2006).

Table 51. Threats COS-T03, COS-T04 and their control actions.

Threat description	Control action description
- COS-T03 – Selection of heterogeneous sample.	- COS-C04 – To select homogeneous subjects.
- COS-T04 – Small sample size.	- COS-C09 – To choose a design that increase number of data points.

Table 51 also presents threat COS-T04. This threat occurs when the sample is considered small (Mouchawrab et al., 2011). Consequently, there might be a lack of statistical power. In order to address threat COS-T04, researcher may choose a design that increases the number of data points (COS-C09), as done by Mouchawrab et al. (2011).

Threat COS-T05 is presented in Table 52. This threat concerns subjects applying the treatment incorrectly (Thelin et al., 2004a). This problem occurs due to lack of

knowledge, or by not following procedures reported by researchers, or even consciously. Seven control actions were identified for this threat.

Table 52. Threat COS-T05 and their control actions.

Threat description	Control action description
- COS-T05 – Reliability of treatment implementation.	- COS-C06 – To provide training for applying the treatment correctly. - COS-C11 – To define guidelines to support subjects during the experiment. - COS-C12 – To remind subjects that they must use treatment correctly. - COS-C13 – To introduce observers in order to check correctness of treatment implementation. - COS-C14 – To use unexperienced subjects. - COS-C15 – To use camera to monitor treatment implementation. - COS-C18 – To use tool support which automates the treatment.

Action COS-C11 aims at formulating guidelines to support subjects during the experiment, as done by Berling and Thelin (2004). It is also possible to explicitly remind subjects how the treatment should be applied (COS-C12) (El-Attar and Miller, 2009). Michalik et al. (2011) introduced observers in the experimental environment in order to check treatment implementation (COS-C13). If there are mistakes in treatment implementation, observers can interfere. Researcher can avoid threat COS-T05 if unexperienced subjects are selected (COS-C14) (El-Attar and Miller, 2009). People who have experience may have a stronger tendency to perform tasks in their own way. Another action to address threat COS-T05 is to use cameras to monitor treatment implementation (COS-C15), as done by Heijstek et al. (2011). Furthermore, Dzidek et al. (2008) conducted training in order to avoid unconformities during treatment implementation (COS-C06). Finally, tool support automating treatment application might also help to mitigate this risk (COS-C18) (Lanubile et al., 2004).

Threat *"Selecting data to satisfy hypothesis"* (COS-T06) may arise when researcher selects the best results among those collected in order to satisfy hypothesis (Bunse, 2006). We did not identify any control actions for this threat.

Table 53 presents threats COS-T07, and COS-T08. The first one may arise when problems happen in the experimental environment (Sfetsos et al., 2009). This fact may affect subjects' performance. The latter one concerns the confidence interval (Ferrari and Madhavji, 2008), if an inadequate confidence interval is used the experiment may lack statistical power. We identified only one action for each of those threats.

Table 53. Threats COS-T07, COS-T08 and their control actions.

Threat description	Control action description
- COS-T07 – Random facts in the experimental environment.	- COS-C08 – To assign subjects to the same experimental environment.
- COS-T08 – Inadequate confidence interval.	- COS-C17 – To use adequate confidence interval, according to literature.

To control threat COS-T07, action COS-C08 suggests assigning subjects to the same experimental environment (Karahasanovic et al., 2005). Therefore, facts which arise will affect all subjects equally. Finally, action COS-C17 suggests using confidence intervals described in literature (Lui et al., 2008).

In this section, we provided on overview on the identified threats and their control actions. However, there are relationships between threats and control actions and understanding these relationships might assist researchers in their experiment planning

activities (RQ2). Therefore, the next section explores these relations by presenting a conceptual model.

4. The Conceptual Model

From the list of threats to validity and control actions retrieved in the SLR, we developed a conceptual model. We believe that the best way to understand the relationships is through graphical form. The conceptual model was built by: (i) grouping threats to validity and their respective control actions, (ii) identifying threats and actions that have trade-off relationships (action used to address a threat that may cause another threat). These relationships not only arise among threats and actions of the same validity type, but also of different validity types.

Figure 1 presents relationships between threat INT-T18 and control actions related to conducting pretest. Pretest activities may change subjects' performance. These changes might be either positive or negative.

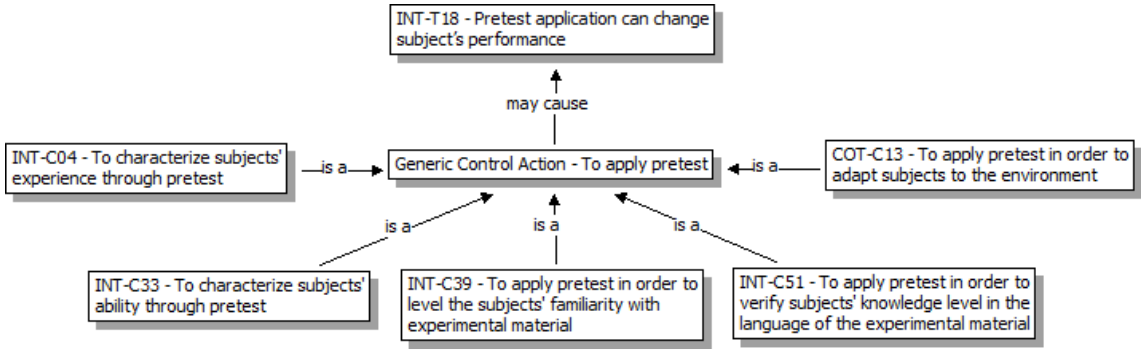

Figure 1. Relationships between threat INT-T18, and control actions INT-C04, INT-C33, INT-C39, INT-C51 and COT-C13.

Actions INT-C04, INT-C33, INT-C39, and INT-C51 are applied in order to address threats INT-T01, INT-T02, INT-T06, and INT-T16, respectively. Action COT-C13 is used in order to mitigate threat COT-T04. The researcher should keep in mind that when one of these actions is applied, the pretest can influence the subjects' performance. Thus, the pretest should be designed carefully.

Figure 1 characterizes a trade-off between construct validity and internal validity. In this case, action COT-C13, which is used to address threat COT-T04, may cause threat INT-T18. Thus, we increase construct validity, but decrease internal validity.

Relationships between threats COT-T09 and COT-T10, and control actions related to using characterization forms are presented in Figure 2.

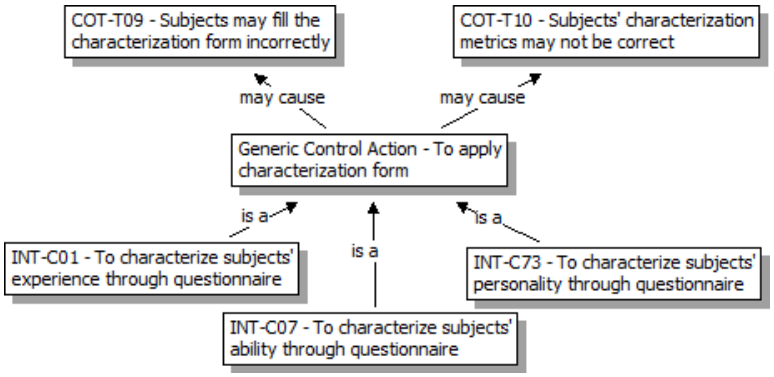

Figure 2. Relationships between threats COT-T09 and COT-T10, and control actions INT-C01, INT-C07 and INT-C73.

Figure 2 also presents a trade-off between construct validity and internal validity. In this case, all three control actions are used in order to mitigate threats to internal validity. However, they may cause threats to construct validity.

Figure 3 presents relationships between control actions related to discarding data and threats INT-T01, INT-T06, INT-T30, INT-T35, EXT-T01, COS-T02 and COS-T04. If any data is discarded, groups may become unbalanced (experience, ability, gender, personality). In addition, subjects' representativeness also is affected. Finally, there may be an impact on the statistical tests (e.g., violating assumptions related to the sample size or distribution).

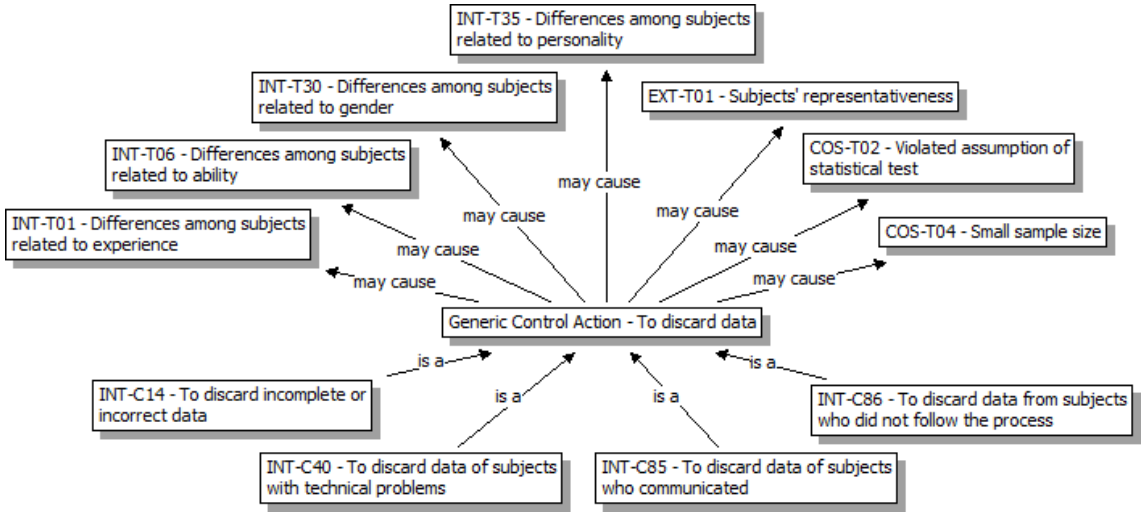

Figure 3. Relationships between control actions INT-C14, INT-C40, INT-C85 and INT-C86, and threats INT-T01, INT-T06, INT-T30, INT-T35, EXT-T01, COS-T02 and COS-T04.

Action INT-C14 aims at controlling threats INT-T07 and INT-T32. In addition, actions INT-C40, INT-C85, and INT-C86 are applied in order to address threats INT-T28, INT-T08, and INT-T25, respectively. By discarding data of subjects, researchers have to conduct data analysis in order to check the impact over the sample.

Two trade-offs are presented in Figure 3. There are actions to control threats to internal validity. Nevertheless, these actions may cause threats to external (EXT-T01) and conclusion validities (COS-T02 and COS-T05). Then, we increase internal validity, but may potentially decrease external and conclusion validities.

Figure 4 presents two relationships. The first one regards introducing observers in the experimental environment. By introducing observers in the experimental environment (INT-C56 and COS-C13), we increase internal and conclusion validities. However, we affect external validity (EXT-T08) and construct validity (COT-T04). Thus, there is trade-off between: (1) internal and external validities; (2) internal and construct validities; (3) conclusion and external validities; and (4) conclusion and construct validities.

Action INT-C56 is applied to address threats INT-T08 and INT-T25. Action COS-C13 aims at mitigating threat COS-T05. In normal working environments usually there is no person observing. Therefore, being observed subjects may not behave naturally.

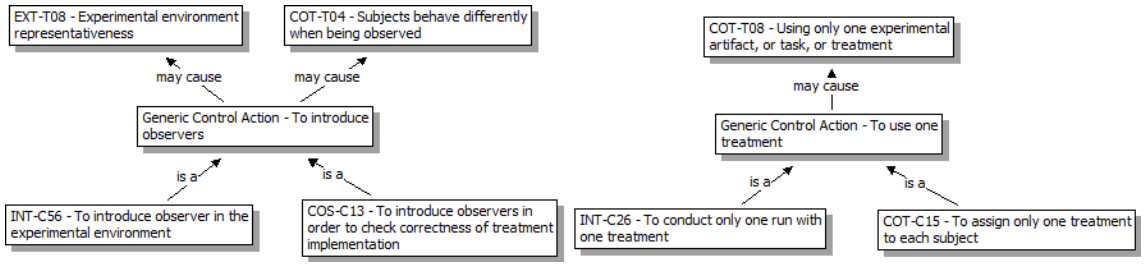

Figure 4. Relationships between control actions INT-C56 and COS-C13, and threats EXT-T08 and COT-T04; and control actions COT-C15 and INT-C26, and threat COT-T08.

The second relationship presented in Figure 4 is related to using only one treatment per subject. Wohlin et al. (2012) state that "*using a single independent variable (treatment) the experiment may under-represent the cause construct and thus not give the full picture of the theory*". These actions (COT-C15 and INT-C26) increase internal and construct validities. Nevertheless, they may cause threat COT-T08. Therefore, in this case, there is a trade-off between internal validity and construct validity.

Relationships between control actions that suggest training and threats to internal validity (INT-T15, INT-T26, INT-T29, and INT-T31) are presented in Figure 5. However, the provided training may be unbalanced (INT-T15). This occurs, for instance, when training material is different among groups. In addition, training may not provide all necessary knowledge for subjects to perform the experiment (INT-T26). Finally, according to specific conditions, trouble may arise both, when training is provided by different people (INT-T29) and when provided by only one person (INT-T31). The first one happens because different people express themselves differently, potentially emphasizing different parts of the training material. The second one, on the other hand, may occur because the person responsible for the training may consciously or unconsciously train a group better.

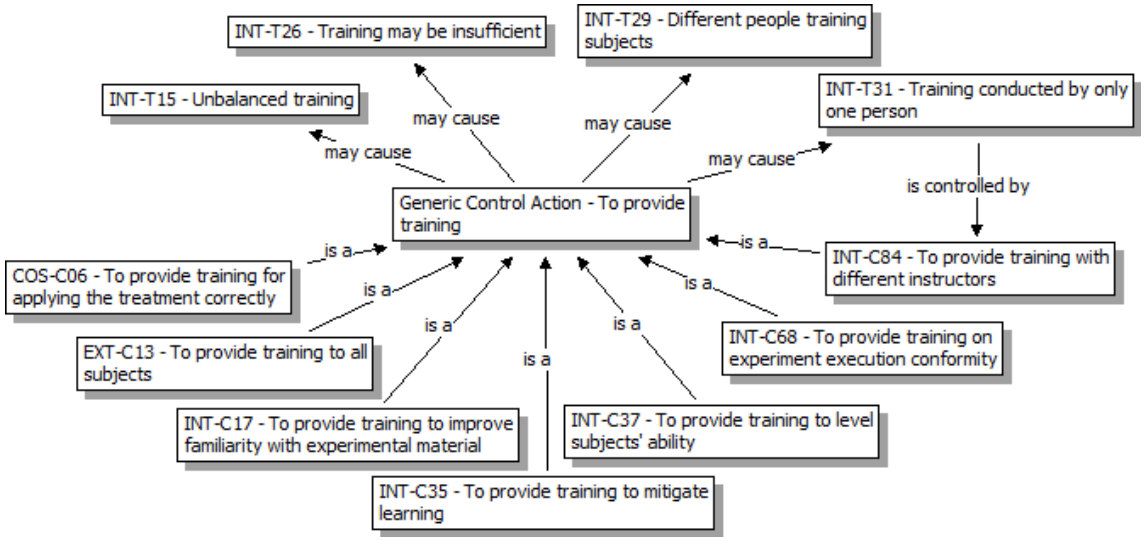

Figure 5. Relationships between control actions COS-C06, EXT-C13, INT-C17, INT-C35, INT-C37, INT-C68 and INT-C84, and threats INT-T15, INT-T26, INT-T29 and INT-T31.

Actions COS-C06 and EXT-C13 are applied in order to control threats COS-T05 and EXT-T01, while actions INT-C17, INT-C35, INT-C37, INT-C68, and INT-C84 aim at addressing threats INT-T02, INT-T03, INT-T06, INT-T25, and INT-T31, respectively.

The training has to be planned consistently in order to ensure fairness. However, we can observe in Figure 5 that there are hardly no threats if training is provided. Therefore, researchers should act carefully when training has to be provided during experiments.

The control actions presented in Figure 5 may cause threats to internal validity. Then, there is a trade-off between: (1) conclusion and internal validities; and (2) external and internal validities. Thus, albeit increasing conclusion and external validities, internal validity may be decreased.

Figure 6 presents relationships which are related to subject selection. Action COS-C04 suggests selection of homogeneous subjects in order to avoid that individual differences influence the experiment results to increase conclusion validity. Action EXT-C21, on the other hand, is related to selecting heterogeneous subjects in order to improve subjects' representativeness, increasing external validity. In this scenario, the trade-off between conclusion and external validity is evident.

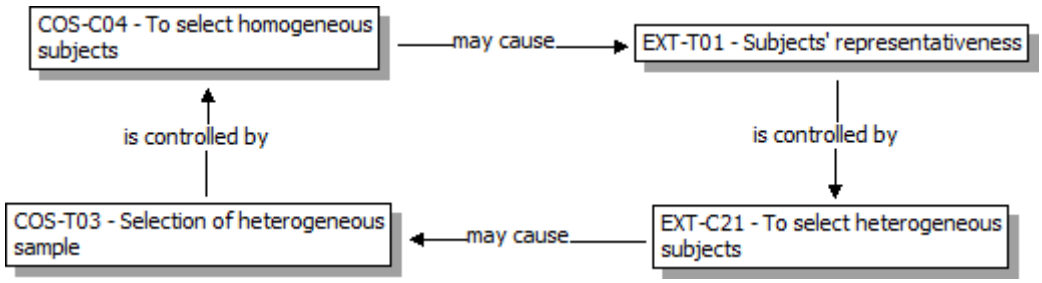

Figure 6. Relationships between control action COS-C04 and threat EXT-T01; and control action EXT-C21 and threat COS-T03.

Relationships between actions COT-C19 and INT-C27, and threat INT-T17 is presented in Figure 7. Historical data from subjects may be used to improve construct validity. This data can be used in order to check data from characterization questionnaires. Thus, if there are any inconsistencies, researcher can identify them. However, subject selection may be influenced, since researcher may have access to information on subjects' performance and potentially select the best subjects from the initial sample. This relationship presents a trade-off between construct and internal validity.

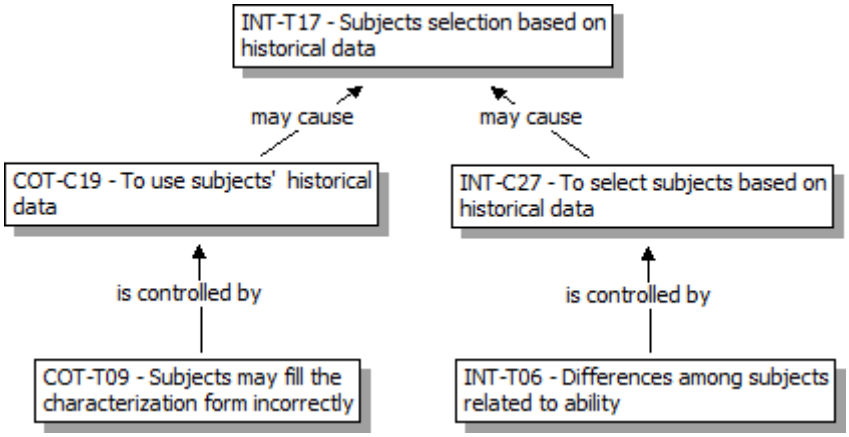

Figure 7. Relationships between control actions COT-C19 and INT-C27, and threat INT-T17.

Figure 8 presents the relationship between actions EXT-C01 and INT-C46, and threat COS-T05. These actions are related to selecting subjects from the real experiment target context. However, these subjects usually have their own way of performing tasks, i.e., do not depend on a specific treatment. Therefore, there is a threat when selecting

subjects from the real context. Thus, we increase the external and internal validities, but decrease the conclusion validity, i.e., we have a trade-off between external and conclusion validities, and internal and conclusion validities.

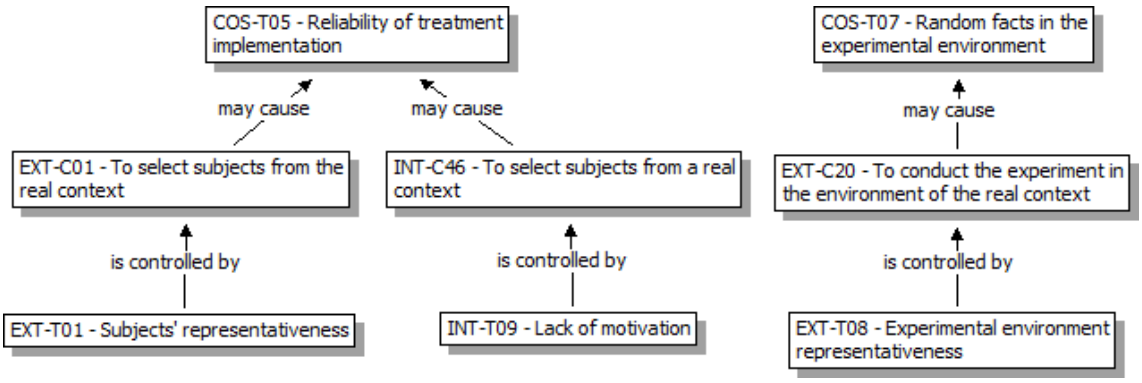

Figure 8. Relationships between control actions EXT-C01 and INT-C46, and threat COS-T05; and control action EXT-C20 and threat COS-T07.

The relationship between action EXT-C20 and threat COS-T07 is also presented in Figure 8. By conducting the experiment in real environment we can control the threat EXT-T08, i.e. we increase the external validity. However, real environments have other variables that may influence the experiment outcome, which may decrease conclusion validity. In this case, there is trade-off between external and conclusion validities.

Figure 9 presents relationships between actions INT-C28 and INT-C31, and threats INT-T13 and INT-T14. These actions suggest the researcher to allow breaks (short or long) during the experiment. The relationship is between control actions and threats from the same validity type, in this case internal validity. To avoid fatigue effects, researchers may apply actions INT-C28 or INT-C31. Nevertheless, in breaks subjects may exchange information about the experiment (INT-T14). Then, subject's performance might be influenced. In addition, everyday situations may occur, potentially affecting subjects' performance (INT-T13).

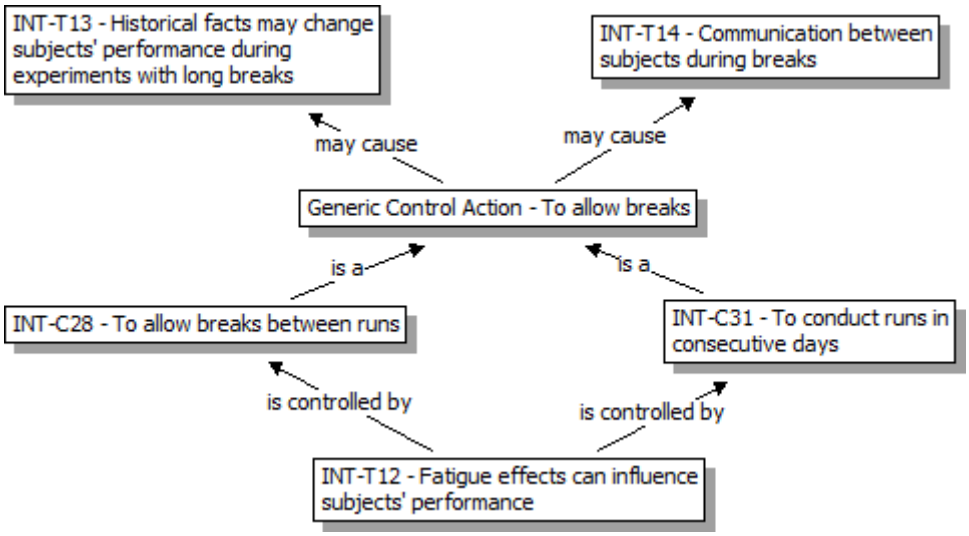

Figure 9. Relationships between control actions INT-C28 and INT-C31, and threats INT-T13 and INT-T14.

The relationship between action INT-C48 and threat COT-T06 is presented in Figure 10. Through questions, subjects can clarify their doubts, increasing internal validity. However, subjects may also get information that enables them to guess the experimental

hypotheses, potentially affecting construct validity. Therefore, in this relationship there is trade-off between internal and construct validities.

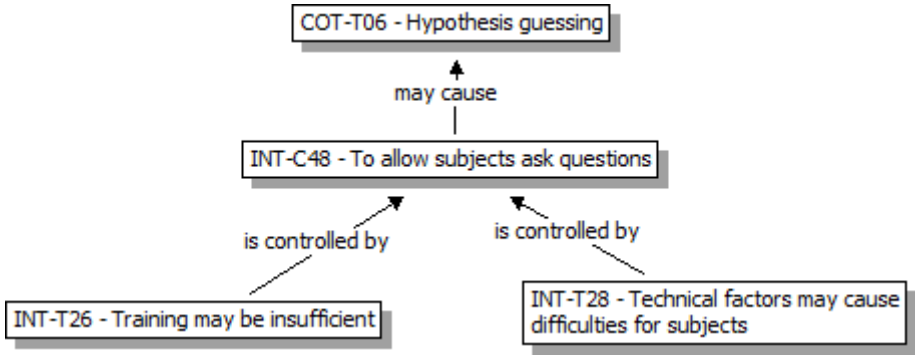

Figure 10. Relationships between control action INT-C48, and threat COT-T06.

In this section, we presented several trade-offs that exists between validity types. When increasing one type, another type may be decreased (Wohlin et al. 2012). Prioritizing the validity types is the responsibility of the researcher. When doing so he should keep in mind that, depending on the experiment goal, some threats are more critical than others (Andersson et al. 2003).

5. Summary

We presented the results of a SLR which aims at identifying threats to validity and control actions. We identified a total of 36 threats to internal validity with 86 control actions; 9 threats to external validity with 23 actions to control; 10 threats to construct validity with 21 control actions; and 8 threats to conclusion validity with 19 control actions. Furthermore, relationships found during the identification of the threats to validity and control actions were presented. These relationships occur when choosing an action to minimize or mitigate a threat enables the emergence of a new threat to validity. Thus, the experiments must be carefully planned, because there are trade-offs between the validity types.

References

S. Abrahão, N. Condori-Fernandez, L. Olsina, and O. Pastor, *"Defining and validating metrics for navigational models"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 200-210, 2003.

R. Ahmed, T. Hall, P. Wernick, and S. Robinson, *"Evaluating a rapid simulation modelling process (RSMP) through controlled experiments"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 325-334, 2005.

B. Anda, and D. I. K. Sjoberg, *"Investigating the role of use cases in the construction of class diagrams"*. Empirical Software Engineering (EMSE) 10 (3), pp. 285-309, 2005.

C. Andersson, T. Thelin, P. Runeson, and N. Dzamashvili, *"An experimental evaluation of inspection and testing for detection of design faults"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 174-184, 2003.

E. Arisholm, and D. I. K. Sjoberg, *"Evaluating the effect of a delegated versus centralized control style on the maintainability of object-oriented software"*. IEEE Transactions on Software Engineering (TSE) 30 (8), pp. 521-534, 2004.

E. Arisholm, L. C. Briand, S. E. Hove, and Y. Labiche, *"The impact of UML documentation on software maintenance – an experimental evaluation"*. IEEE Transactions on Software Engineering (TSE) 32 (6), pp. 365-381, 2006.

E. Arisholm, H. Gallis, T. Dyba, and D. I. K. Sjoberg, *"Evaluating pair programming with respect to system complexity and programmer expertise"*. IEEE Transactions on Software Engineering (TSE) 33 (2), pp. 65-86, 2007.

M. A. Babar, D. Winkler, and S. Biffl, *"Evaluating the usefulness and ease of use of a groupware tool for the software architecture evaluation process"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 430-439, 2007.

R. K. Bandi, V. K. Vaishnavi, and D. E. Turk, *"Predicting maintenance performance using object-oriented design complexity metrics"*. IEEE Transactions on Software Engineering (TSE) 29 (1), pp. 77-87, 2003.

T. Berling, and T. Thelin, *"A case study of reading techniques in a software company"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 229-238, 2004.

B. Bernardez, M. Genero, A. Duran, and M. Toro, *"A controlled experiment for evaluating a metric–based reading technique for requirements inspection"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 257-268, 2004.

S. Biffl, and M. Halling, *"Investigating the defect detection effectiveness and cost benefit of nominal inspection teams"*. IEEE Transactions on Software Engineering (TSE) 29 (5), pp. 385-397, 2003.

L. C. Briand, Y. Labiche, M. D. Penta, and H. Yan-Bondoc, *"An experimental investigation of formality in UML-based development"*. IEEE Transactions on Software Engineering (TSE) 31 (10), pp. 833-849, 2005.

L. C. Briand, Y. Labiche, and R. Madrazo-Rivera, *"An experimental evaluation of the impact of system sequence diagrams and system operation contracts on the quality of the domain model"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 157-166, 2011.

D. Budgen, A. J. Burn, and B. Kitchenham, *"Reporting computing projects through structured abstracts – a quasi-experiment"*. Empirical Software Engineering (EMSE) 16 (2), pp. 244-277, 2011.

C. Bunse, *"Using patterns for the refinement and translation of UML models – a controlled experiment"*. Empirical Software Engineering (EMSE) 11 (2), pp. 227-267, 2006.

G. Canfora, A. Cimitile, F. Garcia, M. Piattini, and C. A. Visaggio, *"Evaluating advantages of test driven development – a controlled experiment with professionals"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 364-371, 2006.

G. Canfora, A. Cimitile, F. Garcia, M. Piattini, and C. A. Visaggio, *"Evaluating performances of pair designing in industry"*. Journal of Systems and Software (JSS) 80 (8), pp. 1317-1327, 2007.

J. Carver, J. Vanvoorhis, and V. R. Basili, *"Understanding the impact of assumptions on experimental validity"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 251-260, 2004.

J. Carver, N. Nagappan, and A. Page, *"The impact of educational background on the effectiveness of requirements inspections – an empirical study"*. IEEE Transactions on Software Engineering (TSE) 34 (6), pp. 800-812, 2008.

D. Chatterji, J. Carver, B. Massengill, J. Oslin, and N. A. Kraft, *"Measuring the efficacy of code clone information in a bug localization task – an empirical study"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 20-29, 2011.

T. Conte, J. Massollar, E. Mendes, and G. H. Travassos, *"Usability evaluation based on web design perspectives"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 146-155, 2007.

T. Conte, *"Usability Inspection Technique Based on Web Design Perspectives"*. Doctoral Thesis, COPPE/UFRJ, 2009.

B. Cornelissen, A. Zaidman, and A. V. Deursen, *"A controlled experiment for program comprehension through trace visualization"*. IEEE Transactions on Software Engineering (TSE) 37 (3), pp. 341-355, 2011.

J. N. Dag, T. Thelin, and B. Regnell, *"An experiment on linguistic tool support for consolidation of requirements from multiple sources in market-driven product development"*. Empirical Software Engineering (EMSE) 11 (2), pp. 303-329, 2006.

D. Damian, F. Lanubile, and T. Mallardo, *"On the need for mixed media in distributed requirements negotiations"*. IEEE Transactions on Software Engineering (TSE) 34 (1), pp. 116-132, 2008.

I. Deligiannis, I. Stamelos, L. Angelis, M. Roumeliotis, and M. Shepperd, *"A controlled experiment investigation of an object-oriented design heuristic for maintainability"*. Journal of Systems and Software (JSS) 72 (2), pp. 129-143, 2004.

C. Denger, and R. Kolb, *"Testing and inspecting reusable product line components – first empirical results"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 184-193, 2006.

W. J. Dzidek, E. Arisholm, and L. C. Briand, *"A realistic empirical evaluation of the costs and benefits of UML in software maintenance"*. IEEE Transactions on Software Engineering (TSE) 34 (3), pp. 407-432, 2008.

M. El-Attar, and J. Miller, *"A subject-based empirical evaluation of SSUCD's performance in reducing inconsistencies in use case models"*. Empirical Software Engineering (EMSE) 14 (5), pp. 477-512, 2009.

H. Erdogmus, M. Morisio, and M. Torchiano, *"On the effectiveness of the test-first approach to programming"*. IEEE Transactions on Software Engineering (TSE) 31 (3), pp. 226-237, 2005.

D. Falessi, G. Cantone, and M. Becker, *"Documenting design decision rationale to improve individual and team design decision making – an experimental evaluation"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 134-143, 2006.

R. Ferrari, and N. H. Madhavji, *"Software architecting without requirements knowledge and experience – what are the repercussions?"*. Journal of Systems and Software (JSS) 81 (9), pp. 1470-1490, 2008.

M. Genero, E. Manso, C. A. Visaggio, G. Canfora, and M. Piattini, *"Building measure-based prediction models for UML class diagram maintainability"*. Empirical Software Engineering (EMSE) 12 (5), pp. 517-549, 2007.

A. Gupta, and P. Jalote, *"An experimental evaluation of the effectiveness and efficiency of the test driven development"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 285-294, 2007.

J. H. Hayes, and J. Offutt, *"Input validation analysis and testing"*. Empirical Software Engineering (EMSE) 11 (4), pp. 493-522, 2006.

W. Heijstek, T. Kuhne, and M. R. V. Chaudron, *"Experimental analysis of textual and graphical representations for software architecture design"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 167-176, 2011.

L. Hochstein, V. R. Basili, U. Vishkin, and J. Gilbert, *"A pilot study to compare programming effort for two parallel programming models"*. Journal of Systems and Software (JSS) 81 (11), pp. 1920-1930, 2008.

A. Hovsepyan, R. Scandariato, S. V. Baelen, W. Joosen, and S. Demeyer, *"Preserving aspects via automation – a maintainability study"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 315-324, 2011.

L. Huang, and M. Holcombe, *"Empirical investigation towards the effectiveness of test first programming"*. Information and Software Technology (IST) 51 (1), pp. 182-194, 2009.

J. Itkonen, M. V. Mantyla, and C. Lassenius, *"Defect detection efficiency – test case based vs. exploratory testing"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 61-70, 2007.

A. Jansen, P. Avgeriou, and J. S. Ven, *"Enriching software architecture documentation"*. Journal of Systems and Software (JSS) 82 (8), pp. 1232-1248, 2009.

A. Jedlitschka, *"Evaluating a model of software managers' information needs – an experiment"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 177-186, 2010.

A. Karahasanovic, B. Anda, E. Arisholm, S. E. Hove, M. Jorgensen, D. I. K. Sjoberg, and R. Welland, *"Collecting feedback during software engineering experiments"*. Empirical Software Engineering (EMSE) 10 (2), pp. 113-147, 2005.

L. Karlsson, T. Thelin, B. Regnell, P. Berander, and C. Wohlin, *"Pair-wise comparisons versus planning game partitioning – experimentts on requirements prioritisation techniques"*. Empirical Software Engineering (EMSE) 12 (1), pp. 3-33, 2007.

B. Kitchenham, and S. Charters, *"Guidelines for performing systematic literature reviews in software engineering"*. EBSE Technical Report EBSE-2007-01, v. 2.3 2007.

J. Knodel, D. Muthig, and M. Naab, *"An experiment on the role of graphical elements in architecture visualization"*. Empirical Software Engineering (EMSE) 13 (6), pp. 693-726, 2008.

L. P. W. Land, B. C. Y. Tan, and L. Bin, *"Investigating training effects on software reviews – a controlled experiment"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 356-366, 2005.

F. Lanubile, T. Mallardo, F. Calefato, C. Denger, and M. Ciolkowski, *"Assessing the impact of active guidance for defect detection – a replicated experiment"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 269-278, 2004.

A. Lucia, R. Oliveto, and G. Tortora, *"Assessing IR-based traceability recovery tools through controlled experiments"*. Empirical Software Engineering (EMSE) 14 (1), pp. 57-92, 2009.

A. Lucia, C. Gravino, R. Oliveto, and G. Tortora, *"An experimental comparison of ER and UML class diagrams for data modelling"*. Empirical Software Engineering (EMSE) 15 (5), pp. 455-492, 2010.

A. Lucia, M. D. Penta, and R. Oliveto, *"Improving source code lexicon via traceability and information retrieval"*. IEEE Transactions on Software Engineering (TSE) 37 (2), pp. 205-227, 2011.

K. M. Lui, K. C. C. Chan, and J. T. Nosek, *"The effect of pairs in program design tasks"*. IEEE Transactions on Software Engineering (TSE) 34 (2), pp. 197-211, 2008.

L. Madeyski, *"The impact of test-first programming on branch coverage and mutation score indicator of unit tests – an experiment"*. Information and Software Technology (IST) 52 (2), pp. 169-184, 2010.

J. C. Maldonado, J. Carver, F. Shull, S. Fabbri, E. Dória, L. Martimiano, M. Mendonça, and V. R. Basili, *"Perspective-based reading – a replicated experiment focused on individual reviewer effectiveness"*. Empirical Software Engineering (EMSE) 11 (1), pp. 119-142, 2006

M. Mendonça, and M. C. F. Oliveira, *"Using observational pilot studies to test and improve lab packages"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 48-57, 2006.

B. Michalik, D. Weyns, N. Boucke, and A. Helleboogh, *"Supporting online updates of software product lines – a controlled experiment"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 187-196, 2011.

S. Mouchawrab, L. C. Briand, Y. Labiche, and M. D. Penta, *"Assessing, comparing, and combining state machine-based testing and structural testing – a series of experiments"*. IEEE Transactions on Software Engineerin (TSE) 37 (2), pp. 161-187, 2011.

M. M. Muller, *"Are reviews an alternative to pair programming?"*. Empirical Software Engineering (EMSE) 9 (4), pp. 335-351, 2004.

M. M. Muller, *"Two controlled experiments concerning the comparison of pair programming to peer review"*. Journal of Systems and Software (JSS) 78 (2), pp. 166-179, 2005.

M. M. Muller, and A. Hofer, *"The effect of experience on the test-driven development process"*. Empirical Software Engineering (EMSE) 12 (6), pp. 593-615, 2007.

M. Perepletchikov, and C. Ryan, *"A controlled experiment for evaluating the impact of coupling on the maintainability of service-oriented software"*. IEEE Transactions on Software Engineering (TSE) 37 (4), pp. 449-465, 2011.

K. Petersen, K. Ronkko, and C. Wohlin, *"The impact of time controlled reading on software inspection effectiveness and efficiency"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 139-148, 2008.

D. Pfahl, O. Laitenberger, J. Dorsch, and G. Ruhe, *"An externally replicated experiment for evaluating the learning effectiveness of using simulations in software project management education"*. Empirical Software Engineering (EMSE) 8 (4), pp. 367-395, 2003.

L. Prechelt, *"Plat_forms – a web development platform comparison by an exploratory experiment searching for emergent platform properties"*. IEEE Transactions on Software Engineering (TSE) 37 (1), pp. 95-108, 2011.

H. Qattous, P. Gray, and R. Welland, *"An empirical study of specification by example in a software engineering tool"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 147-156, 2010.

E. Ras, and J. Rech, *"Using wikis to support the net generation in improving knowledge acquisition in capstone projects"*. Journal of Systems and Software (JSS) 82 (4), pp. 553-562, 2009.

F. Ricca, G. Scanniello, M. Torchiano, G. Reggio, and E. Astesiano, *"On the effectiveness of screen mockups in requirements engineering – results from an internal replication"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2010.

M. P. Robillard, and G. C. Murphy, *"Representing concerns in source code"*. ACM Transactions Software Engineering and Methodology (TOSEM) 16 (1), pp. 3-41, 2007.

J. R. Ruthruff, M. Burnett, and G. Rothermel, *"Interactive fault localization techniques in a spreadsheet environment"*. IEEE Transactions on Software Engineering (TSE) 32 (4), pp. 213-239, 2006.

G. Sabaliauskaite, F. Matsukawa, S. Kusumoto, and K. Inoue, *"Further investigations of reading techniques for object-oriented design inspection"*. Information and Software Technology (IST) 45 (9), pp. 571-585, 2003.

N. Salleh, E. Mendes, J. Grundy, and G. St. J. Burch, *"An empirical study of the effects of personality in pair programming using the five-factor model"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 214-225, 2009.

N. Salleh, E. Mendes, J. Grundy, and G. St. J. Burch, *"An empirical study of the effects of conscientiousness in pair programming using the five-factor personality model"*. ACM International Conference on Software Engineering (ICSE), pp. 577-586, 2010.

K. Sandahl, O. Blomkvist, J. Karlsson, C. Krysander, M. Lindvall, and N. Ohlsson, *"An extended replication of an experiment for assessing methods for software requirements inspections"*. Empirical Software Engineering (EMSE) 3 (4), pp. 327-354, 1998.

P. Sfetsos, I. Stamelos, L. Angelis, and I. Deligiannis, *"An experimental investigation of personality types impact on pair effectiveness in pair programming"*. Empirical Software Engineering (EMSE) 14 (2), pp. 187-226, 2009.

A. Sinha, and C. Smidts, *"HOTTest: a model-based test design technique for enhanced testing of domain-specific applications"*. ACM Transactions on Software Engineering and Methodology (TOSEM) 15 (3), pp. 242-278, 2006.

D. I. K. Sjoberg, J. E. Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N. K. Liborg, and A. C. Rekdal, *"A survey of controlled experiments in software engineering"*. IEEE Transactions on Software Engineering (TSE) 31 (9), pp. 733-753, 2005.

M. Staron, L. Kuzniarz, and C. Wohlin, *"Empirical assessment of using stereotypes to improve comprehension of UML models – a set of experiments"*. Journal of Systems and Software (JSS) 79 (5), pp. 727-742, 2006.

M. Svahnberg, and C. Wohlin, *"An investigation of a method for identifying a software architecture candidate with respect to quality attributes"*. Empirical Software Engineering (EMSE) 10 (2), pp. 149-181, 2005.

T. Thelin, P. Runeson, and B. Regnell, *"Usage-based reading – an experiment to guide reviewers with use cases"*. Information and Software Technology (IST) 43 (15), pp. 925-938, 2001.

T. Thelin, P. Runeson, and C. Wohlin, *"An experimental comparison of usage-based and checklist-based reading"*. IEEE Transactions on Software Engineering (TSE) 29 (8), pp. 687-704, 2003.

T. Thelin, C. Andersson, P. Runeson, and N. Dzamashvili-Fogelstrom, *"A replicated experiment of usage-based and checklist-based reading"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 246-256, 2004a.

T. Thelin, P. Runeson, C. Wohlin, T. Olsson, and C. Andersson, *"Evaluation of usage-based reading – conclusions after three experiments"*. Empirical Software Engineering (EMSE) 9 (1-2), pp. 77-110, 2004b.

P. Tonella, and M. Ceccato, *"Refactoring the aspectizable interfaces – an empirical assessment"*. IEEE Transactions on Software Engineering (TSE) 31 (10), pp. 819-832, 2005.

P. Vitharana, and K. Ramamurthy, *"Computer-mediated group support, anonymity, and the software inspection process – an empirical investigation"*. IEEE Transactions on Software Engineering (TSE) 29 (2), pp. 167-180, 2003.

M. Vokac, W. Tichy, D. I. K. Sjoberg, E. Arisholm, and M. Aldrin, *"A controlled experiment comparing the maintainability of programs designed with and without design patterns – a replication in a real programming"*. Empirical Software Engineering (EMSE) 9 (3), pp. 149-195, 2004.

R. Wettel, M. Lanza, and R. Robbes, *"Software systems as cities – a controlled experiment"*. ACM International Conference on Software Engineering (ICSE), pp. 551-560, 2011.

C. Wohlin, P. Runeson, M. Host, M. C. Ohlsson, B. Regnell, and A. Wesslen, *"Experimentation in software engineering"*. Springer-Verlag Berlin Heidelberg, 2012.

Selected Papers (SP) of the TTV Review

[SP001] S. Abrahão, N. Condori-Fernandez, L. Olsina, and O. Pastor, *"Defining and validating metrics for navigational models"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 200-210, 2003.

[SP002] S. Abrahão, and G. Poels, *"Experimental evaluation of an object-oriented function point measurement procedure"*. Information and Software Technology (IST), vol. 49, no. 4, pp. 366-380, 2007.

[SP003] S. Acuna, M. Gomez, and N. Juristo, *"Towards understanding the relationship between team climate and software quality - a quasi-experimental study"*. Empirical Software Engineering (EMSE), vol. 13, no. 4, pp. 401-434, 2008.

[SP004] S. Acuna, M. Gomez, and N. Juristo, *"How do personality, team processes and task characteristics relate to job satisfaction and software quality?"*. Information and Software Technology (IST), vol. 51, no. 3, pp. 627-639, 2009.

[SP005] R. Ahmed, T. Hall, P. Wernick, and S. Robinson, *"Evaluating a rapid simulation modelling process (RSMP) through controlled experiments"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 325-334, 2005.

[SP006] B. Anda, and D. I. K. Sjoberg, *"Applying use cases to design versus validate class diagrams - a controlled experiment using a professional modelling tool"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 50-60, 2003.

[SP007] B. Anda, and D. I. K. Sjoberg, *"Investigating the role of use cases in the construction of class diagrams"*. Empirical Software Engineering (EMSE), vol. 10, no. 3, pp. 285-309, 2005.

[SP008] C. Andersson, T. Thelin, P. Runeson, and N. Dzamashvili, *"An experimental evaluation of inspection and testing for detection of design faults"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 174-184, 2003.

[SP009] E. Arisholm, D. I. K. Sjoberg, and M. Jorgensen, *"Assessing the changeability of two object-oriented design alternatives - a controlled experiment"*. Empirical Software Engineering (EMSE), vol. 6, no. 3, pp. 231-277, 2001.

[SP010] E. Arisholm, and D. I. K. Sjoberg, *"Evaluating the effect of a delegated versus centralized control style on the maintainability of object-oriented software"*. IEEE Transactions on Software Engineering (TSE), vol. 30, no. 8, pp. 521-534, 2004.

[SP011] E. Arisholm, L. C. Briand, S. E. Hove, and Y. Labiche, *"The impact of UML documentation on software maintenance - an experimental evaluation"*. IEEE Transactions on Software Engineering (TSE), vol. 32, no. 6, pp. 365-381, 2006.

[SP012] E. Arisholm, H. Gallis, T. Dyba, and D. I. K. Sjoberg, *"Evaluating pair programming with respect to system complexity and programmer expertise"*. IEEE Transactions on Software Engineering (TSE), vol. 33, no. 2, pp. 65-86, 2007.

[SP013] M. A. Babar, and S. Biffl, *"Eliciting better quality architecture evaluation scenarios: a controlled experiment on top-down vs. bottom-up"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 307-315, 2006.

[SP014] M. A. Babar, B. Kitchenham, L. Zhu, I. Gorton, and R. Jeffery, *"An empirical study of groupware support for distributed software architecture evaluation process"*. Journal of Systems and Software (JSS), vol. 79, no. 7, pp. 912-925, 2006.

[SP015] M. A. Babar, and B. Kitchenham, *"The impact of group size on software architecture evaluation: a controlled experiment"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 420-429, 2007.

[SP016] M. A. Babar, D. Winkler, and S. Biffl, *"Evaluating the usefulness and ease of use of a groupware tool for the software architecture evaluation process"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 430-439, 2007.

[SP017] M. A. Babar, B. Kitchenham, and R. Jeffery, *"Comparing distributed and face-to-face meetings for software architecture evaluation: a controlled experiment"*. Empirical Software Engineering (EMSE), vol. 13, no. 1, pp. 39-62, 2008.

[SP018] R. K. Bandi, V. K. Vaishnavi, and D. E. Turk, *"Predicting maintenance performance using object-oriented design complexity metrics"*. IEEE Transactions on Software Engineering (TSE), vol. 29, no. 1, pp. 77-87, 2003.

[SP019] V. R. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sorumgard, and M. V. Zelkowitz, *"The empirical investigation of perspective-based reading"*. Empirical Software Engineering (EMSE), vol. 1, no. 2, pp. 133-164, 1996.

[SP020] E. Bellini, G. Canfora, F. Garcia, M. Piattini, and C. A. Visaggio, *"Pair designing as practice for enforcing and diffusing design knowledge"*. Software Maintenance and Evolution: Research and Practice (SME), vol. 17, no. 6, pp. 401-423, 2005.

[SP021] T. Berling, and T. Thelin, *"A case study of reading techniques in a software company"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 229-238, 2004.

[SP022] B. Bernardez, M. Genero, A. Duran, and M. Toro, *"A controlled experiment for evaluating a metric–based reading technique for requirements inspection"*.

ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 257-268, 2004.

[SP023] A. Bianchi, F. Lanubile, and G. Visaggio, *"A controlled experiment to assess the effectiveness of inspection meetings"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 42-50, 2001.

[SP024] S. Biffl, and M. Halling, *"Investigating the influence of inspector capability factors with four inspection techniques on inspection performance"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 107-117, 2002.

[SP025] S. Biffl, *"Evaluating defect estimation models with major defects"*. Journal of Systems and Software (JSS), vol. 65, no. 1, pp. 13-29, 2003.

[SP026] S. Biffl, and M. Halling, *"Investigating the defect detection effectiveness and cost benefit of nominal inspection teams"*. IEEE Transactions on Software Engineering (TSE), vol. 29, no. 5, pp. 385-397, 2003.

[SP027] S. Biffl, S. Thurnher, G. Goluch, D. Winkler, W. Aigner, and S. Miksch, *"An empirical investigation on the visualization of temporal uncertainties in software engineering project planning"*. CM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 437-446, 2005.

[SP028] F. O. Bjornson, A. I. Wang, and E. Arisholm, *"Improving the effectiveness of root cause analysis in post mortem analysis: a controlled experiment"*. Information and Software Technology (IST), vol. 51, no. 1, pp. 150-161, 2009.

[SP029] L. C. Briand, C. Bunse, J. W. Daly, and C. Differding, *"An experimental comparison of the maintainability of object-oriented and structured design documents"*. Empirical Software Engineering (EMSE), vol. 2, no. 3, pp. 291-312, 1997.

[SP030] L. C. Briand, C. Bunse, and J. W. Daly, *"A controlled experiment for evaluating quality guidelines on the maintainability of object-oriented designs"*. IEEE Transactions on Software Engineering (TSE), vol. 27, no. 6, pp. 513-530, 2001.

[SP031] L. C. Briand, Y. Labiche, M. D. Penta, and H. Yan-Bondoc, *"An experimental investigation of formality in UML-based development"*. IEEE Transactions on Software Engineering (TSE), vol. 31, no. 10, pp. 833-849, 2005.

[SP032] L. C. Briand, Y. Labiche, and R. Madrazo-Rivera, *"An experimental evaluation of the impact of system sequence diagrams and system operation contracts on the quality of the domain model"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 157-166, 2011.

[SP033] D. Budgen, and M. Thomson, *"CASE tool evaluation: experiences from an empirical study"*. Journal of Systems and Software (JSS), vol. 67, no. 2, pp. 55-75, 2003.

[SP034] D. Budgen, B. Kitchenham, S. Charters, M. Turner, P. Brereton, and S. Linkman, *"Presenting software engineering results using structured abstracts: a randomised experiment"*. Empirical Software Engineering (EMSE), vol. 13, no. 4, pp. 435-468, 2008.

[SP035] D. Budgen, A. J. Burn, and B. Kitchenham, *"Reporting computing projects through structured abstracts - a quasi-experiment"*. Empirical Software Engineering (EMSE), vol. 16, no. 2, pp. 244-277, 2011.

[SP036] C. Bunse, *"Using patterns for the refinement and translation of UML models - a controlled experiment"*. Empirical Software Engineering (EMSE), vol. 11, no. 2, pp. 227-267, 2006.

[SP037] C. Calero, M. Piattini, and M. Genero, *"Empirical validation of referential integrity metrics"*. Information and Software Technology (IST), vol. 43, no. 15, pp. 949-957, 2001.

[SP038] G. Canfora, A. Cimitile, F. Garcia, M. Piattini, and C. A. Visaggio, *"Evaluating advantages of test driven development - a controlled experiment with professionals"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 364-371, 2006.

[SP039] G. Canfora, A. Cimitile, F. Garcia, M. Piattini, and C. A. Visaggio, *"Evaluating performances of pair designing in industry"*. Journal of Systems and Software (JSS), vol. 80, no. 8, pp. 1317-1327, 2007.

[SP040] J. Carver, M. Fisher II, and G. Rothermel, *"An empirical evaluation of a testing and debbuging methodology for Excel"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 278-287, 2006a.

[SP041] J. Carver, F. Shull, and V. R. Basili, *"Can observational techniques help novices overcome the software inspection learning curve? An empirical investigation"*. Empirical Software Engineering (EMSE), vol. 11, no. 4, pp. 523-539, 2006b.

[SP042] J. Carver, N. Nagappan, and A. Page, *"The impact of educational background on the effectiveness of requirements inspections - an empirical study"*. IEEE Transactions on Software Engineering (TSE), vol. 34, no. 6, pp. 800-812, 2008.

[SP043] D. Chatterji, J. Carver, B. Massengill, J. Oslin, and N. A. Kraft, *"Measuring the efficacy of code clone information in a bug localization task - an empirical study"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 20-29, 2011.

[SP044] T. Y. Chen, P. Poon, S. Tang, and T. H. Tse, *"On the identification of categories and choices for specification-based test case generation"*. Information and Software Technology (IST), vol. 46, no. 13, pp. 887-898, 2004.

[SP045] R. W. Collins, A. R. Hevner, G. H. Walton, and R. C. Linger, *"The impacts of function extraction technology on program comprehension: a controlled experiment"*. Information and Software Technology (IST), vol. 50, no. 11, pp. 1165-1179, 2008.

[SP046] M. Colosimo, A. Lucia, G. Scanniello, G. Tortora, *"Evaluating legacy system migration technologies through empirical studies"*. Information and Software Technology (IST), vol. 51, no. 2, pp. 433-447, 2009.

[SP047] T. Conte, J. Massollar, E. Mendes, and G. H. Travassos, *"Usability evaluation based on web design perspectives"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 146-155, 2007.

[SP048] B. Cornelissen, A. Zaidman, and A. V. Deursen, *"A controlled experiment for program comprehension through trace visualization"*. IEEE Transactions on Software Engineering (TSE), vol. 37, no. 3, pp. 341-355, 2011.

[SP049] J. A. Cruz-Lemus, M. Genero, M. E. Manso, S. Morasca, and M. Piattini, *"Assessing the understandability of UML statechart diagrams with composite states - a family of empirical studies"*. Empirical Software Engineering (EMSE), vol. 14, no. 6, pp. 685-719, 2009.

[SP050] J. N. Dag, T. Thelin, and B. Regnell, *"An experiment on linguistic tool support for consolidation of requirements from multiple sources in market-driven product development"*. Empirical Software Engineering (EMSE), vol. 11, no. 2, pp. 303-329, 2006.

[SP051] D. Damian, F. Lanubile, and T. Mallardo, *"On the need for mixed media in distributed requirements negotiations"*. IEEE Transactions on Software Engineering (TSE), vol. 34, no. 1, pp. 116-132, 2008.

[SP052] I. Deligiannis, M. Shepperd, M. Roumeliotis, and I. Stamelos, "An empirical investigation of an object-oriented design heuristic for maintainability". Journal of Systems and Software (JSS), vol. 65, no. 2, pp. 127-139, 2003.

[SP053] I. Deligiannis, I. Stamelos, L. Angelis, M. Roumeliotis, and M. Shepperd, *"A controlled experiment investigation of an object-oriented design heuristic for maintainability"*. Journal of Systems and Software (JSS), vol. 72, no. 2, pp. 129-143, 2004.

[SP054] C. Denger, M. Ciolkowski, and F. Lanubile, *"Investigating the active guidance factor in reading techniques for defect detection"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 219-228, 2004.

[SP055] C. Denger, and R. Kolb, *"Testing and inspecting reusable product line components - first empirical results"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 184-193, 2006.

[SP056] G. Du, J. McElroy, and G. Ruhe, *"A family of empirical studies to compare informal and optimization-based planning of software releases"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 212-221, 2006.

[SP057] A. Dunsmore, M. Roper, and M. Wood, *"Systematic object-oriented inspection - an empirical study"*. ACM International Conference on Software Engineering (ICSE), pp. 135-144, 2001.

[SP058] A. Dunsmore, M. Roper, and M. Wood, *"Further investigations into the development and evaluation of reading techniques for object-oriented code inspection"*. ACM International Conference on Software Engineering (ICSE), pp. 47-57, 2002.

[SP059] W. J. Dzidek, E. Arisholm, and L. C. Briand, *"A realistic empirical evaluation of the costs and benefits of UML in software maintenance"*. IEEE Transactions on Software Engineering (TSE), vol. 34, no. 3, pp. 407-432, 2008.

[SP060] H. Eichelberger, and K. Schmid, *"Guidelines on the aesthetic quality of UML class diagrams"*. Information and Software Technology (IST), vol. 51, no. 12, pp. 1686-1698, 2009.

[SP061] M. El-Attar, and J. Miller, *"A subject-based empirical evaluation of SSUCD's performance in reducing inconsistencies in use case models"*. Empirical Software Engineering (EMSE), vol. 14, no. 5, pp. 477-512, 2009.

[SP062] H. Erdogmus, M. Morisio, and M. Torchiano, *"On the effectiveness of the test-first approach to programming"*. IEEE Transactions on Software Engineering (TSE), vol. 31, no. 3, pp. 226-237, 2005.

[SP063] D. Falessi, G. Cantone, and M. Becker, *"Documenting design decision rationale to improve individual and team design decision making - an experimental evaluation"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 134-143, 2006.

[SP064] R. Ferrari, and N. H. Madhavji, *"Software architecting without requirements knowledge and experience - what are the repercussions?"*. Journal of Systems and Software (JSS), vol. 81, no. 9, pp. 1470-1490, 2008.

[SP065] B. Freimut, O. Laitenberger, and S. Biffl, *"Investigating the impact of reading techniques on the accuracy of different defect detection content estimation techniques"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 51-62, 2001.

[SP066] P. Fusaro, F. Lanubile, and G. Visaggio, *"A replicated experiment to assess requirements inspection techniques"*. Empirical Software Engineering (EMSE), vol. 2, no. 1, pp. 39-57, 1997.

[SP067] M. Genero, M. Piattini, E. Manso, and G. Cantone, *"Building UML class diagrams maintainability prediction models based on early metrics"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 263-275, 2003.

[SP068] M. Genero, E. Manso, C. A. Visaggio, G. Canfora, and M. Piattini, *"Building measure-based prediction models for UML class diagram maintainability"*. Empirical Software Engineering (EMSE), vol. 12, no. 5, pp. 517-549, 2007.

[SP069] B. George, and L. Williams, *"A structured experiment of test-driven development"*. Information and Software Technology (IST), vol. 46, no. 5, pp. 337-342, 2004.

[SP070] A. Geras, M. Smith, and J. Miller, *"A prototype empirical evaluation of test driven development"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 405-416, 2004.

[SP071] G. Giraudo, and P. Tonella, *"Designing and conducting an empirical study on test management automation"*. Empirical Software Engineering (EMSE), vol. 8, no. 1, pp. 59-81, 2003.

[SP072] C. Glezer, M. Last, E. Nachmany, and P. Shoval, *"Quality and comprehension of UML interaction diagrams - an experimental comparison"*. Information and Software Technology (IST), vol. 47, no. 10, pp. 675-692, 2005.

[SP073] C. K. Gonzales, and G. Leroy, *"Eliciting user requirements using appreciative inquiry"*. Empirical Software Engineering (EMSE), vol. 16, no. 6, pp. 733-772, 2011.

[SP074] G. Gulesir, K. Berg, L. Bergmans, and M. Aksit, *"Experimental evaluation of a tool for the verification and transformation of source code in event-driven systems"*. Empirical Software Engineering (EMSE), vol. 14, no. 6, pp. 720-777, 2009.

[SP075] A. Gupta, and P. Jalote, *"An experimental evaluation of the effectiveness and efficiency of the test driven development"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 285-294, 2007a.

[SP076] A. Gupta, and P. Jalote, *"Test inspected unit or inspect unit tested code?"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 51-60, 2007b.

[SP077] R. Harrison, S. Counsell, and R. Nithi, *"Experimental assessment of the effect of inheritance on the maintainability of object-oriented systems"*. Journal of Systems and Software (JSS), vol. 52, no. 2-3, pp. 173-179, 2000.

[SP078] J. H. Hayes, and J. Offutt, *"Input validation analysis and testing"*. Empirical Software Engineering (EMSE), vol. 11, no. 4, pp. 493-522, 2006.

[SP079] L. He, and J. Carver, *"PBR vs. checklist: a replication in the n-fold inspection context"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 95-104, 2006.

[SP080] W. Heijstek, T. Kuhne, and M. R. V. Chaudron, *"Experimental analysis of textual and graphical representations for software architecture design"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 167-176, 2011.

[SP081] A. Herrmann, and B. Paech, *"Practical challenges of requirements prioritization based on risk estimation"*. Empirical Software Engineering (EMSE), vol. 14, no. 6, pp. 644-684, 2009.

[SP082] E. Hill, L. Pollock, and K. Vijay-Shanker, *"Automatically capturing source code context of NL-queries for software maintenance and reuse"*. ACM International Conference on Software Engineering (ICSE), pp. 232-242, 2009.

[SP083] L. Hochstein, V. R. Basili, U. Vishkin, and J. Gilbert, *"A pilot study to compare programming effort for two parallel programming models"*. Journal of Systems and Software (JSS), vol. 81, no. 11, pp. 1920-1930, 2008.

[SP084] R. Hofman, *"Behavioral economics in software quality engineering"*. Empirical Software Engineering (EMSE), vol. 16, no. 2, pp. 278-293, 2011.

[SP085] M. Host, and C. Wohlin, *"An experimental study of individual subjective effort estimations and combinations of the estimates"*. ACM International Conference on Software Engineering (ICSE), pp. 332-339, 1998.

[SP086] M. Host, and C. Johansson, *"Evaluation of code review methods through interviews and experimentation"*. Journal of Systems and Software (JSS), vol. 52, no. 2-3, pp. 113-120, 2000.

[SP087] M. Host, B. Regnell, and C. Wohlin, *"Using students as subjects - a comparative study of students and professionals in lead-time impact assessment"*. Empirical Software Engineering (EMSE), vol. 5, no. 3, pp. 201-214, 2000.

[SP088] A. Hovsepyan, R. Scandariato, S. V. Baelen, W. Joosen, and S. Demeyer, *"Preserving aspects via automation - a maintainability study"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 315-324, 2011.

[SP089] L. Huang, and M. Holcombe, *"Empirical investigation towards the effectiveness of test first programming"*. Information and Software Technology (IST), vol. 51, no. 1, pp. 182-194, 2009.

[SP090] J. Itkonen, M. V. Mantyla, and C. Lassenius, *"Defect detection efficiency - test case based vs. exploratory testing"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 61-70, 2007.

[SP091] A. Jansen, P. Avgeriou, and J. S. Ven, *"Enriching software architecture documentation"*. Journal of Systems and Software (JSS), vol. 82, no. 8, pp. 1232-1248, 2009.

[SP092] A. Jedlitschka, *"Evaluating a model of software managers' information needs – an experiment"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 177-186, 2010.

[SP093] P. M. Johnson, and D. Tjahjono, *"Does every inspection really need a meeting?"*. Empirical Software Engineering (EMSE), vol. 3, no. 1, pp. 9-35, 1998.

[SP094] M. Jorgensen, and D. I. K. Sjoberg, *"Impact of effort estimates on software project work"*. Information and Software Technology (IST), vol. 43, no. 15, pp. 939-948, 2001.

[SP095] N. Juristo, and S. Vegas, *"The role of non-exact replications in software engineering experiments"*. Empirical Software Engineering (EMSE), vol. 16, no. 3, pp. 295-324, 2011.

[SP096] J. Kabeli, and P. Shoval, *"Comprehension and quality of analysis specifications - a comparison of FOOM and OPM methodologies"*. Information and Software Technology (IST), vol. 47, no. 4, pp. 271-290, 2005.

[SP097] E. Kamsties, A. Knethen, and R. Reussner, *"A controlled experiment to evaluate how styles affect the understandability of requirements specifications"*. Information and Software Technology (IST), vol. 45, no. 14, pp. 955-965, 2003.

[SP098] A. Karahasanovic, B. Anda, E. Arisholm, S. E. Hove, M. Jorgensen, D. I. K. Sjoberg, and R. Welland, *"Collecting feedback during software engineering experiments"*. Empirical Software Engineering (EMSE), vol. 10, no. 2, pp. 113-147, 2005.

[SP099] L. Karlsson, T. Thelin, B. Regnell, P. Berander, and C. Wohlin, *"Pair-wise comparisons versus planning game partitioning - experiments on requirements prioritisation techniques"*. Empirical Software Engineering (EMSE), vol. 12, no. 1, pp. 3-33, 2007.

[SP100] M. Keil, L. Li, L. Mathiassen, and G. Zheng, *"The influence of checklists and roles on software practitioner risk perception and decision-making"*. Journal of Systems and Software (JSS), vol. 81, no. 6, pp. 908-919, 2008.

[SP101] J. D. Kiper, B. Auernheimer, and C. K. Ames, *"Visual depiction of decision statements: what is best for programmers and non-programmers?"*. Empirical Software Engineering (EMSE), vol. 2, no. 4, pp. 361-379, 1997.

[SP102] J. Knodel, D. Muthig, and M. Naab, *"An experiment on the role of graphical elements in architecture visualization"*. Empirical Software Engineering (EMSE), vol. 13, no. 6, pp. 693-726, 2008.

[SP103] A. J. Ko, and B. A. Myers, *"Extracting and answering why and why not questions about java program output"*. ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 20, no. 2, pp. 4-36, 2010.

[SP104] O. Laitenberger, and J. DeBaud, *"Perspective-based reading of code documents at Robert Bosch GmbH"*. Information and Software Technology (IST), vol. 39, no. 11, pp. 781-791, 1997.

[SP105] O. Laitenberger, C. Atkinson, M. Schlich, and K. E. Emam, *"An experimental comparison of reading techniques for defect detection in UML design documents"*. Journal of Systems and Software (JSS), vol. 53, no. 2, pp. 183-204, 2000.

[SP106] O. Laitenberger, K. E. Emam, and T. G. Harbich, *"An internally replicated quasi-experimental comparison of checklist and perspective-based reading of code documents"*. IEEE Transactions on Software Engineering (TSE), vol. 27, no. 5, pp. 387-421, 2001.

[SP107] L. P. W. Land, C. Sauer, and R. Jeffery, *"The use of procedural roles in code inspections: an experimental study"*. Empirical Software Engineering (EMSE), vol. 5, no. 1, pp. 11-34, 2000.

[SP108] L. P. W. Land, B. C. Y. Tan, and L. Bin, *"Investigating training effects on software reviews - a controlled experiment"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 356-366, 2005.

[SP109] F. Lanubile, T. Mallardo, F. Calefato, C. Denger, and M. Ciolkowski, *"Assessing the impact of active guidance for defect detection - a replicated experiment"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 269-278, 2004.

[SP110] C. M. Lott, *"A controlled experiment to evaluate on-line process guidance"*. Empirical Software Engineering (EMSE), vol. 2, no. 3, pp. 269-289, 1997.

[SP111] A. Lucia, R. Francese, I. Passero, and G. Tortora, *"Development and evaluation of a system enhancing Second Life to support synchronous role-based collaborative learning"*. Software: Practice and Experience (SP&E), vol. 39, no. 12, pp. 1025-1054, 2009a.

[SP112] A. Lucia, R. Oliveto, and G. Tortora, *"Assessing IR-based traceability recovery tools through controlled experiments"*. Empirical Software Engineering (EMSE), vol. 14, no. 1, pp. 57-92, 2009b.

[SP113] A. Lucia, C. Gravino, R. Oliveto, and G. Tortora, *"An experimental comparison of ER and UML class diagrams for data modelling"*. Empirical Software Engineering (EMSE), vol. 15, no. 5, pp. 455-492, 2010.

[SP114] A. Lucia, M. D. Penta, and R. Oliveto, *"Improving source code lexicon via traceability and information retrieval"*. IEEE Transactions on Software Engineering (TSE) vol. 37, no. 2, pp. 205-227, 2011.

[SP115] K. M. Lui, K. C. C. Chan, and J. T. Nosek, *"The effect of pairs in program design tasks"*. IEEE Transactions on Software Engineering (TSE), vol. 34, no. 2, pp. 197-211, 2008.

[SP116] F. MacDonald, and J. Miller, *"A comparison of tool-based and paper-based software inspection"*. Empirical Software Engineering (EMSE), vol. 3, no. 3, pp. 233-253, 1998.

[SP117] Madeyski, L., *"The impact of test-first programming on branch coverage and mutation score indicator of unit tests - an experiment"*. Information and Software Technology (IST), vol. 52, no. 2, pp. 169-184, 2010.

[SP118] J. C. Maldonado, J. Carver, F. Shull, S. Fabbri, E. Dória, L. Martimiano, M. Mendonça, and V. R. Basili, *"Perspective-based reading - a replicated experiment focused on individual reviewer effectiveness"*. Empirical Software Engineering (EMSE), vol. 11, no. 1, pp. 119-142, 2006.

[SP119] A. Martens, H. Koziolek, L. Prechelt, and R. Reussner, *"From monolithic to component-based performance evaluation of software architectures - a series of experiments analysing accuracy and effort"*. Empirical Software Engineering (EMSE), vol. 16, no. 5, pp. 587-622, 2011.

[SP120] P. McCarthy, A. A. Porter, H. Siy, and L. G. Votta Jr., *"An experiment to assess cost-benefits of inspection meetings and their alternatives: a pilot study"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 100-110, 1996.

[SP121] M. Mendonça, and M. C. F. Oliveira, *"Using observational pilot studies to test and improve lab packages"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 48-57, 2006.

[SP122] B. Michalik, D. Weyns, N. Boucke, and A. Helleboogh, *"Supporting online updates of software product lines - a controlled experiment"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 187-196, 2011.

[SP123] J. Miller, M. Wood, and M. Roper, *"Further experiences with scenarios and checklists"*. Empirical Software Engineering (EMSE), vol. 3, no. 1, pp. 37-64, 1998.

[SP124] K. Molokken-Ostvold, and M. Jorgensen, *"Expert estimation of web-development projects: are software professionals in technical roles more optimistic than those in non-technical roles?"*. Empirical Software Engineering (EMSE), vol. 10, no. 1, pp. 7-29, 2005.

[SP125] S. Mouchawrab, L. C. Briand, Y. Labiche, and M. D. Penta, *"Assessing, comparing, and combining state machine-based testing and structural testing - a series of experiments"*. IEEE Transactions on Software Engineerin (TSE), vol. 37, no. 2, pp. 161-187, 2011.

[SP126] M. M. Muller, *"Are reviews an alternative to pair programming?"*. Empirical Software Engineering (EMSE), vol. 9, no. 4, pp. 335-351, 2004.

[SP127] M. M. Muller, and F. Padberg, *"An empirical study about the feelgood factor in pair programming"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 151-158, 2004.

[SP128] M. M. Muller, *"Two controlled experiments concerning the comparison of pair programming to peer review"*. Journal of Systems and Software (JSS), vol. 78, no. 2, pp. 166-179, 2005.

[SP129] M. M. Muller, and A. Hofer, *"The effect of experience on the test-driven development process"*. Empirical Software Engineering (EMSE), vol. 12, no. 6, pp. 593-615, 2007.

[SP130] S. Nanz, F. Torshizi, M. Pedroni, and B. Meyer, *"Design of an empirical study for comparing the usability of concurrent programming languages"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 325-334, 2011.

[SP131] V. Nguyen, B. Boehm, and P. Danphitsanuphan, *"A controlled experiment in assessing and estimating software maintenance tasks"*. Information and Software Technology (IST), vol. 53, no. 6, pp. 682-691, 2011.

[SP132] A. Nugroho, *"Level of detail in UML models and its impact on model comprehension: a controlled experiment"*. Information and Software Engineering (IST), vol. 51, no. 12, pp. 1670-1685, 2009.

[SP133] M. Ochodek, B. Alchimowicz, J. Jurkiewicz, and J. Nawrocki, *"Improving the reliability of transaction identification in use cases"*. Information and Software Engineering (IST), vol. 53, no. 8, pp. 885-897, 2011.

[SP134] M. C. Ohlsson, C. Wohlin, and B. Regnell, *"A project effort estimation study"*. Information and Software Engineering (IST), vol. 40, no. 14, pp. 831-839, 1998.

[SP135] A. L. Opdahl, and G. Sindre, *"Experimental comparison of attack trees and misuse cases for security threat identification"*. Information and Software Engineering (IST), vol. 51, no. 5, pp. 916-932, 2009.

[SP136] M. C. Otero, and J. J. Dolado, *"An initial experimental assessment of the dynamic modelling in UML"*. Empirical Software Engineering (EMSE), vol. 7, no. 1, pp. 27-47, 2002.

[SP137] M. C. Otero, and J. J. Dolado, *"Evaluation of the comprehension of the dynamic modeling in UML"*. Information and Software Engineering (IST), vol. 46, no. 1, pp. 35-53, 2004.

[SP138] M. C. Otero, and J. J. Dolado, *"An empirical comparison of the dynamic modeling in OML and UML"*. Journal of Systems and Software (JSS), vol. 77, no. 2, pp. 91-102, 2005.

[SP139] M. Pancur, and M. Ciglaric, *"Impact of test-driven development on productivity, code and tests: a controlled experiment"*. Information and Software Engineering (IST), vol. 53, no. 6, pp. 557-573, 2011.

[SP140] M. Perepletchikov, and C. Ryan, *"A controlled experiment for evaluating the impact of coupling on the maintainability of service-oriented software"*. IEEE Transactions on Software Engineering (TSE), vol. 37, no. 4, pp. 449-465, 2011.

[SP141] A. Perini, F. Ricca, and A. Susi, *"Tool-supported requirements prioritization: comparing the AHP and CBRank methods"*. Information and Software Engineering (IST), vol. 51, no. 6, pp. 1021-1032, 2009.

[SP142] K. Petersen, K. Ronkko, and C. Wohlin, *"The impact of time controlled reading on software inspection effectiveness and efficiency"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 139-148, 2008.

[SP143] D. Pfahl, N. Koval, and G. Ruhe, *"An experiment for evaluating the effectiveness of using a system dynamics simulation model in software project management education"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 97-109, 2001.

[SP144] D. Pfahl, O. Laitenberger, J. Dorsch, and G. Ruhe, *"An externally replicated experiment for evaluating the learning effectiveness of using simulations in software project management education"*. Empirical Software Engineering (EMSE), vol. 8, no. 4, pp. 367-395, 2003.

[SP145] D. Pfahl, O. Laitenberger, G. Ruhe, J. Dorsch, and T. Krivobokova, *"Evaluating the learning effectiveness of using simulations in software project management education: results from a twice replicated experiment"*. Information and Software Engineering (IST), vol. 46, no. 2, pp. 127-147, 2004.

[SP146] M. Phongpaibul, and B. Boehm, *"An empirical comparison between pair development and software inspection in Thailand"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 85-94, 2006.

[SP147] G. C. Porras, and Y. Gueheneuc, *"An empirical study on the efficiency of different design pattern representations in UML class diagrams"*. Empirical Software Engineering (EMSE), vol. 15, no. 5, pp. 493-522, 2010.

[SP148] A. A. Porter, and L. G. Votta Jr., *"An experiment to assess different defect detection methods for software requirements inspections"*. ACM International Conference on Software Engineering (ICSE), pp. 103-112, 1994.

[SP149] A. A. Porter, L. G. Votta Jr., and V. R. Basili, *"Comparing detection methods for software requirements inspections: a replicated experiment"*. IEEE Transactions on Software Engineering (TSE), vol. 21, no. 6, pp. 563-575, 1995.

[SP150] A. A. Porter, and P. M. Johnson, *"Assessing software review meetings: results of a comparative analysis of two experimental studies"*. IEEE Transactions on Software Engineering (TSE), vol. 23, no. 3, pp. 129-145, 1997.

[SP151] A. A. Porter, and L. G. Votta Jr., *"Comparing detection methods for software requirements inspections: a replication using professional subjects"*. Empirical Software Engineering (EMSE), vol. 3, no. 4, pp. 355-379, 1998.

[SP152] L. Prechelt, and W. Tichy, *"A controlled experiment to assess the benefits of procedure argument type checking"*. IEEE Transactions on Software Engineering (TSE), vol. 24, no. 4, pp. 302-312, 1998.

[SP153] L. Prechelt, and B. Unger, *"An experiment measuring the effects of personal software process (PSP) training"*. IEEE Transactions on Software Engineering (TSE), vol. 27, no. 5, pp. 465-472, 2000.

[SP154] L. Prechelt, B. Unger, W. Tichy, P. Brossler, and L. G. Votta Jr., *"A controlled experiment in maintenance comparing design patterns to simpler solutions"*. IEEE Transactions on Software Engineering (TSE), vol. 27, no. 12, pp. 1134-1144, 2001.

[SP155] L. Prechelt, B. Unger, M. Philippsen, and W. Tichy, *"Two controlled experiments assessing the usefulness of design pattern documentation in program maintenance"*. IEEE Transactions on Software Engineering (TSE), vol. 28, no. 6, pp. 595-606, 2002.

[SP156] L. Prechelt, B. Unger, M. Philippsen, and W. Tichy, *"A controlled experiment on inheritance depth as a cost factor for code maintenance"*. Journal of Systems and Software (JSS), vol. 65, no. 2, pp. 115-126, 2003.

[SP157] L. Prechelt, *"Plat_forms - a web development platform comparison by an exploratory experiment searching for emergent platform properties"*. IEEE Transactions on Software Engineering (TSE), vol. 37, no. 1, pp. 95-108, 2011.

[SP158] H. Qattous, P. Gray, and R. Welland, *"An empirical study of specification by example in a software engineering tool"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 147-156, 2010.

[SP159] S. Ramanujan, R. W. Scamell, and J. R. Shah, *"An experimental investigation of the impact of individual, program, and organizational characteristics on software*

maintenance effort". Journal of Systems and Software (JSS), vol. 54, no. 2, pp. 137-157, 2000.

[SP160] E. Ras, and J. Rech, *"Using wikis to support the net generation in improving knowledge acquisition in capstone projects"*. Journal of Systems and Software (JSS), vol. 82, no. 4, pp. 553-562, 2009.

[SP161] B. Regnell, P. Runeson, and T. Thelin, *"Are the perspectives really different? - further experimentation on scenario-based reading of requirements"*. Empirical Software Engineering (EMSE), vol. 5, no. 4, pp. 331-356, 2000.

[SP162] I. Reinhartz-Berger, and D. Dori, *"OPM vs. UML - experimenting with comprehension and construction of web application models"*. Empirical Software Engineering (EMSE), vol. 10, no. 1, pp. 57-79, 2005.

[SP163] F. Ricca, M. D. Penta, M. Torchiano, P. Tonella, and M. Ceccato, *"How design notations affect the comprehension of web applications"*. Software Maintenance and Evolution: Research and Practice (SME), vol. 19, no. 5, pp. 339-359, 2007.

[SP164] F. Ricca, M. Torchiano, M. D. Penta, M. Ceccato, and P. Tonella, *"Using acceptance tests as a support for clarifying requirements: a series of experiments"*. Information and Software Technology (IST), vol. 51, no. 2, pp. 270-283, 2009.

[SP165] F. Ricca, G. Scanniello, M. Torchiano, G. Reggio, and E. Astesiano, *"On the effectiveness of screen mockups in requirements engineering - results from an internal replication"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2010.

[SP166] M. P. Robillard, and G. C. Murphy, *"Representing concerns in source code"*. ACM Transactions Software Engineering and Methodology (TOSEM), vol. 16, no. 1, pp. 3-41, 2007.

[SP167] G. Robiolo, *"How simple is it to measure software size and complexity for an IT practitioner?"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 40-48, 2011.

[SP168] M. Roper, M. Wood, and J. Miller, *"An empirical evaluation of defect detection techniques"*. Information and Software Technology (IST), vol. 39, no. 11, pp. 763-775, 1997.

[SP169] K. J. Rothermel, C. R. Cook, M. M. Burnett, J. Schonfeld, T. R. G. Green, and G. Rothermel, *"WYSIWYT testing in the spreadsheet paradigm: an empirical evaluation"*. ACM International Conference on Software Engineering (ICSE), pp. 230-239, 2000.

[SP170] J. R. Ruthruff, M. Burnett, and G. Rothermel, *"An empirical study of fault localization for end-user programmers"*. ACM International Conference on Software Engineering (ICSE), pp. 352-361, 2005.

[SP171] J. R. Ruthruff, M. Burnett, and G. Rothermel, *"Interactive fault localization techniques in a spreadsheet environment"*. IEEE Transactions on Software Engineering (TSE), vol. 32, no. 4, pp. 213-239, 2006.

[SP172] G. Sabaliauskaite, F. Matsukawa, S. Kusumoto, and K. Inoue, *"An experimental comparison of checklist-based reading and perspective-based reading for UML design document inspection"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 148-157, 2002.

[SP173] G. Sabaliauskaite, F. Matsukawa, S. Kusumoto, and K. Inoue, *"Further investigations of reading techniques for object-oriented design inspection"*. Information and Software Technology (IST), vol. 45, no. 9, pp. 571-585, 2003.

[SP174] G. Sabaliauskaite, S. Kusumoto, and K. Inoue, *"Assessing defect detection performance of interacting teams in object-oriented design inspection"*. Information and Software Technology (IST), vol. 46, no. 13, pp. 875-886, 2004.

[SP175] N. Salleh, E. Mendes, J. Grundy, and G. St. J. Burch, *"An empirical study of the effects of personality in pair programming using the five-factor model"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 214-225, 2009.

[SP176] N. Salleh, E. Mendes, J. Grundy, and G. St. J. Burch, *"An empirical study of the effects of conscientiousness in pair programming using the five-factor personality model"*. ACM International Conference on Software Engineering (ICSE), pp. 577-586, 2010a.

[SP177] N. Salleh, E. Mendes, J. Grundy, and G. St. J. Burch, *"The effects of neuroticism on pair programming: an empirical study in the higher education context"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), 2010b.

[SP178] K. Sandahl, O. Blomkvist, J. Karlsson, C. Krysander, M. Lindvall, and N. Ohlsson, *"An extended replication of an experiment for assessing methods for software requirements inspections"*. Empirical Software Engineering (EMSE), vol. 3, no. 4, pp. 327-354, 1998.

[SP179] P. Sfetsos, I. Stamelos, L. Angelis, and I. Deligiannis, *"An experimental investigation of personality types impact on pair effectiveness in pair programming"*. Empirical Software Engineering (EMSE), vol. 14, no. 2, pp. 187-226, 2009.

[SP180] S. E. Sim, M. Umarji, S. Ratanotayanon, and C. V. Lopes, *"How well do search engines support code retrieval on the web?"*. ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 21, no. 1, pp. 4-25, 2011.

[SP181] A. Sinha, and C. Smidts, *"An experimental evaluation of a higher-ordered-typed-functional specification-based test-generation technique"*. Empirical Software Engineering (EMSE), vol. 11, no. 2, pp. 173-202, 2006a.

[SP182] A. Sinha, and C. Smidts, *"HOTTest: a model-based test design technique for enhanced testing of domain-specific applications"*. ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 15, no. 3, pp. 242-278, 2006b.

[SP183] C. F. Snook, and R. Harrison, *"Experimental comparison of the comprehensibility of a Z specification and its implementation in Java"*. Information and Software Technology (IST), vol. 46, no. 14, pp. 955-971, 2004.

[SP184] M. Staron, L. Kuzniarz, and C. Wohlin, *"Empirical assessment of using stereotypes to improve comprehension of UML models - a set of experiments"*. Journal of Systems and Software (JSS), vol. 79, no. 5, pp. 727-742, 2006.

[SP185] M. Svahnberg, and C. Wohlin, *"An investigation of a method for identifying a software architecture candidate with respect to quality attributes"*. Empirical Software Engineering (EMSE), vol. 10, no. 2, pp. 149-181, 2005.

[SP186] S. Syed-Abdullah, M. Holcombe, and M. Gheorge, *"The impact of an agile methodology on the well being of development teams"*. Empirical Software Engineering (EMSE), vol. 11, no. 1, pp. 143-167, 2006.

[SP187] T. Thelin, P. Runeson, and B. Regnell, *"Usage-based reading - an experiment to guide reviewers with use cases"*. Information and Software Technology (IST), vol. 43, no. 15, pp. 925-938, 2001.

[SP188] T. Thelin, P. Runeson, C. Wohlin, T. Olsson, and C. Andersson, *"How much information is needed for usage-based reading? - a series of experiments"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 127-138, 2002.

[SP189] T. Thelin, P. Runeson, and C. Wohlin, *"An experimental comparison of usage-based and checklist-based reading"*. IEEE Transactions on Software Engineering (TSE), vol. 29, no. 8, pp. 687-704, 2003.

[SP190] T. Thelin, C. Andersson, P. Runeson, and N. Dzamashvili-Fogelstrom, *"A replicated experiment of usage-based and checklist-based reading"*. ACM/IEEE International Symposium on Software Metrics (METRICS), pp. 246-256, 2004a.

[SP191] T. Thelin, P. Runeson, C. Wohlin, T. Olsson, and C. Andersson, *"Evaluation of usage-based reading - conclusions after three experiments"*. Empirical Software Engineering (EMSE), vol. 9, no. 1-2, pp. 77-110, 2004b.

[SP192] P. Tonella, and M. Ceccato, *"Refactoring the aspectizable interfaces - an empirical assessment"*. IEEE Transactions on Software Engineering (TSE), vol. 31, no. 10, pp. 819-832, 2005.

[SP193] S. Vegas, and V. R. Basili, *"A characterization schema for software testing techniques"*. Empirical Software Engineering (EMSE), vol. 10, no. 4, pp. 437-466, 2005.

[SP194] J. Verelst, *"The influence of the level of abstraction on the evolvability of conceptual models of information systems"*. Empirical Software Engineering (EMSE), vol. 10, no. 4, pp. 467-494, 2005.

[SP195] P. Vitharana, and K. Ramamurthy, *"Computer-mediated group support, anonymity, and the software inspection process - an empirical investigation"*. IEEE Transactions on Software Engineering (TSE), vol. 29, no. 2, pp. 167-180, 2003.

[SP196] M. Vokac, W. Tichy, D. I. K. Sjoberg, E. Arisholm, and M. Aldrin, *"A controlled experiment comparing the maintainability of programs designed with and without design patterns - a replication in a real programming"*. Empirical Software Engineering (EMSE), vol. 9, no. 3, pp. 149-195, 2004.

[SP197] G. S. Walia, J. Carver, and T. Philip, *"Requirement error abstraction and classification: an empirical study"*. ACM/IEEE International Symposium on Empirical Software Engineering (ISESE), pp. 336-345, 2006.

[SP198] G. S. Walia, and J. Carver, *"Evaluation of capture-recapture models for estimating the abundance of naturally-occurring defects"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 158-167, 2008.

[SP199] T. Walle, and J. E. Hannay, *"Personality and the nature of collaboration in pair programming"*. ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM), pp. 203-213, 2009.

[SP200] A. I. Wang, and E. Arisholm, *"The effect of task order on the maintainability of object-oriented software"*. Information and Software Technology (IST), vol. 51, no. 2, pp. 293-305, 2009.

[SP201] C. G. Wangenheim, M. Thiry, and D. Kochanski, *"Empirical evaluation of an educational game on software measurement"*. Empirical Software Engineering (EMSE), vol. 14, no. 4, pp. 418-452, 2009.

[SP202] R. Wettel, M. Lanza, and R. Robbes, *"Software systems as cities - a controlled experiment"*. ACM International Conference on Software Engineering (ICSE), pp. 551-560, 2011.

[SP203] M. A. Wojcicki, and P. Strooper, *"Maximising the information gained from a study of static analysis technologies for concurrent software"*. Empirical Software Engineering (EMSE), vol. 12, no. 6, pp. 617-645, 2007.

[SP204] A. Zendler, T. Pfeiffer, M. Eicks, and F. Lehner, *"Experimental comparison of coarse-grained concepts in UML, OML, and TOS"*. Journal of Systems and Software (JSS), vol. 57, no. 1, pp. 21-30, 2001.

[SP205] J. Zettel, *"Methodology support in CASE tools and its impact on individual acceptance and use: a controlled experiment"*. Empirical Software Engineering (EMSE), vol. 10, no. 3, pp. 367-394, 2005.

[SP206] Z. Zhang, V. R. Basili, and B. Shneiderman, *"Perspective-based usability inspection: an empirical validation of efficacy"*. Empirical Software Engineering (EMSE), vol. 4, no. 1, pp. 43-69, 1999.

APPENDIX A. TTVs and CAs and their References

Table A1 presents each evidence (TTV or CA) with its respective references. In other words, which paper describes the TTV or CA extracted during the review. The IDs have the following standard: <validityType>-<evidenceType><sequentialNumber>. For each field, the possible values are:

- <validityType>:
  - INT – internal validity;
  - EXT – external validity;
  - COT – construct validity; and
  - COS – conclusion validity.
- <evidenceType>:
  - T – threat to validity; and
  - C – control action.
- <sequentialNumber>: this number is incremented as an evidence of the same validity type is identified.

Table A1. Identified TTVs and CAs and their references.

TTV/CA	References (Selected Papers – SP)
INT-T01	SP001, SP003, SP004, SP008, SP009, SP011, SP019, SP020, SP024, SP035, SP037, SP044, SP046, SP055, SP058, SP059, SP064, SP065, SP069, SP071, SP077, SP079, SP083, SP084, SP092, SP102, SP107, SP109, SP119, SP127, SP126, SP129, SP130, SP131, SP133, SP138, SP137, SP140, SP142, SP143, SP144, SP145, SP148, SP153, SP155, SP156, SP159, SP166, SP179, SP172, SP180, SP185, SP186, SP190, SP189, SP192, SP193, SP195, SP201, SP206
INT-T02	SP001, SP019, SP021, SP030, SP039, SP043, SP112, SP068, SP071, SP088, SP096, SP102, SP121, SP127, SP130, SP131, SP145, SP152, SP162, SP169, SP179, SP180, SP187, SP188, SP192, SP196, SP205, SP206
INT-T03	SP014, SP017, SP008, SP018, SP019, SP020, SP021, SP022, SP023, SP027, SP029, SP030, SP031, SP032, SP034, SP036, SP037, SP038, SP039, SP040, SP041, SP044, SP046, SP050, SP054, SP057, SP058, SP055, SP060, SP061, SP063, SP068, SP072, SP074, SP075, SP076, SP080, SP081, SP087, SP088, SP090, SP092, SP093, SP095, SP097, SP099, SP102, SP104, SP105, SP108, SP109, SP110, SP111, SP112, SP113, SP114, SP115, SP116, SP117, SP118, SP119, SP120, SP121, SP122, SP123, SP125, SP131, SP135, SP136, SP137, SP138, SP140, SP141, SP143, SP144, SP145, SP147, SP148, SP149, SP151, SP152, SP159, SP161, SP163, SP165, SP166, SP168, SP171, SP172, SP173, SP178, SP179, SP180, SP181, SP182, SP184, SP185, SP187, SP193, SP195, SP197, SP201, SP204, SP205, SP206
INT-T04	SP003, SP004, SP008, SP010, SP012, SP014, SP017, SP022, SP029, SP030, SP036, SP037, SP039, SP051, SP053, SP054, SP055, SP063, SP064, SP072, SP074, SP077, SP079, SP081, SP087, SP091, SP093, SP096, SP097, SP105, SP107, SP108, SP109, SP111, SP116, SP120, SP121, SP122, SP123, SP138, SP137, SP140, SP143, SP144, SP148, SP149, SP150, SP151, SP156, SP159, SP160, SP164, SP167, SP168, SP172, SP173, SP174, SP178, SP183, SP190, SP194, SP195, SP201, SP202
INT-T05	SP001, SP003, SP008, SP011, SP018, SP021, SP026, SP037, SP038, SP047, SP068, SP103, SP123, SP131, SP134, SP141, SP152, SP154, SP165, SP179, SP181, SP190, SP191, SP205
INT-T06	SP001, SP003, SP004, SP008, SP009, SP012, SP013, SP014, SP015, SP017, SP019, SP022, SP024, SP025, SP026, SP028, SP029, SP031, SP032, SP033, SP034, SP035, SP036, SP038, SP039, SP041, SP042, SP044, SP045, SP046, SP048, SP051, SP053, SP054, SP059, SP060, SP061, SP064, SP068, SP072, SP075, SP077, SP079, SP084, SP091, SP093, SP107, SP108, SP109, SP110, SP111, SP114, SP115, SP116, SP119, SP120, SP121, SP123, SP125, SP129, SP130, SP132, SP134, SP135, SP136, SP137, SP138, SP140, SP143, SP144, SP145, SP148, SP149, SP150, SP151, SP154, SP155, SP156, SP157, SP160, SP163, SP168, SP169, SP170, SP172, SP173, SP174, SP177, SP178, SP179, SP180, SP181, SP183, SP184, SP186, SP187, SP189, SP194, SP195, SP196, SP200, SP201, SP202, SP203, SP204, SP205

TTV/CA	References (Selected Papers – SP)
INT-T07	SP003, SP004, SP008, SP029, SP031, SP040, SP046, SP050, SP061, SP072, SP086, SP087, SP099, SP104, SP112, SP137, SP139, SP143, SP144, SP145, SP155, SP159, SP163, SP173, SP175, SP177, SP179, SP182, SP187, SP201, SP205
INT-T08	SP027, SP031, SP032, SP037, SP038, SP050, SP068, SP072, SP074, SP075, SP076, SP084, SP089, SP108, SP109, SP111, SP113, SP114, SP117, SP121, SP123, SP125, SP127, SP135, SP139, SP142, SP179, SP187, SP188, SP191, SP193
INT-T09	SP001, SP002, SP003, SP020, SP021, SP022, SP036, SP037, SP038, SP039, SP042, SP048, SP050, SP053, SP058, SP061, SP063, SP064, SP068, SP075, SP076, SP083, SP111, SP115, SP127, SP135, SP161, SP165, SP172, SP173, SP174, SP178, SP179, SP182, SP187, SP188, SP189, SP190, SP191, SP193, SP201, SP203
INT-T10	SP003, SP007, SP008, SP009, SP012, SP021, SP027, SP046, SP048, SP054, SP056, SP061, SP072, SP074, SP076, SP078, SP080, SP086, SP092, SP097, SP101, SP108, SP112, SP132, SP133, SP142, SP147, SP161, SP163, SP170, SP182, SP187, SP195, SP198, SP201, SP204, SP206
INT-T11	SP004, SP087, SP102, SP111
INT-T12	SP001, SP008, SP017, SP019, SP020, SP021, SP022, SP023, SP027, SP032, SP037, SP038, SP039, SP040, SP050, SP061, SP068, SP072, SP078, SP085, SP087, SP088, SP095, SP098, SP099, SP104, SP117, SP121, SP125, SP135, SP136, SP140, SP141, SP143, SP147, SP159, SP168, SP173, SP179, SP181, SP182, SP187, SP193, SP205, SP206
INT-T13	SP005, SP021, SP084, SP104, SP182, SP185, SP187
INT-T14	SP005, SP008, SP019, SP023, SP024, SP032, SP072, SP080, SP081, SP095, SP101, SP104, SP116, SP125, SP126, SP141, SP147, SP159, SP166, SP174, SP185, SP189, SP201
INT-T15	SP002, SP028, SP045, SP083, SP118, SP135, SP150, SP162, SP169, SP170, SP202
INT-T16	SP050, SP056, SP068, SP118, SP158, SP178, SP193
INT-T17	SP125, SP159, SP205
INT-T18	SP019, SP057, SP087, SP091, SP159, SP182, SP205
INT-T19	SP020, SP118, SP123, SP149, SP151
INT-T20	SP002, SP036, SP048, SP076, SP078, SP099, SP102, SP113, SP121, SP123, SP125, SP135, SP145, SP155, SP158, SP170, SP171, SP187, SP192, SP202
INT-T21	SP029, SP030, SP036, SP038, SP040, SP046, SP053, SP082, SP131, SP140, SP164, SP166, SP202
INT-T22	SP050, SP089, SP117, SP139, SP179, SP187, SP188, SP189
INT-T23	SP001, SP020, SP037, SP038, SP039, SP068, SP135, SP179, SP192
INT-T24	SP014, SP017, SP116, SP120, SP123, SP136, SP137, SP138, SP150, SP151, SP160
INT-T25	SP003, SP004, SP010, SP011, SP024, SP025, SP026, SP028, SP030, SP036, SP054, SP055, SP056, SP062, SP065, SP070, SP089, SP090, SP092, SP109, SP117, SP118, SP121, SP126, SP128, SP158, SP173, SP174, SP193, SP206
INT-T26	SP066, SP117, SP135, SP167, SP184
INT-T27	SP050, SP165, SP189, SP191
INT-T28	SP001, SP004, SP005, SP043, SP048, SP062, SP098, SP170, SP171, SP180, SP185, SP193, SP196, SP200
INT-T29	SP113, SP126, SP164, SP203
INT-T30	SP176, SP177
INT-T31	SP047, SP108, SP128
INT-T32	SP008, SP036, SP038, SP048, SP190
INT-T33	SP043
INT-T34	SP129

TTV/CA	References (Selected Papers – SP)
INT-T35	SP003, SP175, SP179
INT-T36	SP102, SP165
INT-C01	SP008, SP011, SP055, SP059, SP064, SP071, SP079, SP083, SP092, SP102, SP107, SP127, SP130, SP131, SP133, SP137, SP140, SP142, SP143, SP156, SP185, SP190, SP193, SP195
INT-C02	SP003, SP004, SP008, SP010, SP014, SP038, SP054, SP055, SP068, SP072, SP081, SP107, SP108, SP109, SP114, SP116, SP120, SP123, SP138, SP148, SP149, SP150, SP151, SP178, SP184, SP202
INT-C03	SP003, SP004, SP008, SP011, SP013, SP014, SP015, SP017, SP019, SP026, SP028, SP029, SP031, SP032, SP034, SP035, SP036, SP039, SP041, SP045, SP046, SP053, SP055, SP060, SP061, SP072, SP077, SP079, SP084, SP092, SP093, SP102, SP108, SP109, SP111, SP114, SP115, SP119, SP120, SP123, SP125, SP135, SP137, SP138, SP142, SP143, SP144, SP145, SP148, SP150, SP154, SP155, SP156, SP160, SP163, SP170, SP172, SP173, SP174, SP179, SP181, SP183, SP184, SP187, SP189, SP190, SP193, SP194, SP195, SP196, SP201, SP202, SP205
INT-C04	SP037, SP058, SP140, SP144, SP145, SP166, SP192
INT-C05	SP143, SP144, SP145, SP201
INT-C06	SP022, SP050, SP161, SP172, SP173, SP174, SP178, SP187, SP188, SP191
INT-C07	SP008, SP009, SP022, SP031, SP033, SP036, SP041, SP048, SP059, SP079, SP107, SP121, SP130, SP137, SP140, SP154, SP156, SP157, SP160, SP163, SP184, SP194, SP195, SP196, SP202
INT-C08	SP012, SP022, SP024, SP025, SP026, SP031, SP036, SP041, SP048, SP093, SP111, SP114, SP116, SP125, SP137, SP148, SP150, SP154, SP160, SP163, SP181, SP184, SP189, SP196, SP202
INT-C09	SP019, SP032, SP046, SP072, SP075, SP076, SP110, SP125, SP137, SP138, SP163, SP193
INT-C10	SP017, SP019, SP036, SP081, SP105, SP112, SP113, SP152, SP165, SP193
INT-C11	SP008, SP022, SP027, SP029, SP030, SP036, SP051, SP053, SP077, SP087, SP093, SP097, SP105, SP150, SP167, SP170, SP172, SP173, SP174, SP183, SP190, SP195, SP201
INT-C12	SP029, SP030, SP036, SP082, SP131, SP202
INT-C13	SP001, SP020, SP022, SP037, SP038, SP039, SP050, SP061, SP068, SP072, SP087, SP121, SP141, SP147, SP159, SP173, SP179, SP187, SP206
INT-C14	SP008, SP050, SP104, SP129, SP137, SP190
INT-C15	SP053, SP054, SP063, SP064, SP072, SP074, SP078, SP092, SP097, SP108, SP160, SP163, SP190, SP201, SP206
INT-C16	SP014, SP026, SP028, SP041, SP109, SP120, SP142, SP149, SP151, SP159, SP179, SP192, SP202
INT-C17	SP019, SP096, SP112, SP145, SP179
INT-C18	SP001, SP020, SP037, SP038, SP039, SP068, SP135, SP179
INT-C19	SP008, SP014, SP017, SP034, SP037, SP038, SP054, SP060, SP061, SP063, SP080, SP093, SP113, SP120, SP123, SP125, SP135, SP136, SP137, SP140, SP141, SP148, SP151, SP149, SP159, SP168, SP178, SP184, SP185, SP195, SP197, SP205, SP206
INT-C20	SP017, SP120, SP123, SP136, SP137, SP138, SP150, SP151, SP160
INT-C21	SP003, SP080, SP204
INT-C22	SP085, SP125, SP135
INT-C23	SP023, SP024, SP027, SP032, SP031, SP080, SP108, SP116, SP121, SP125, SP143, SP145
INT-C24	SP008, SP009, SP027, SP072, SP074, SP076, SP108, SP121, SP122, SP132
INT-C25	SP011, SP024, SP055, SP058, SP065, SP079, SP131, SP137, SP142, SP148, SP190, SP192
INT-C26	SP020, SP022, SP058, SP063, SP081, SP084, SP092, SP117, SP121, SP161, SP172, SP187

TTV/CA	References (Selected Papers – SP)
INT-C27	SP064, SP093, SP125, SP150, SP168, SP172, SP173, SP177
INT-C28	SP019, SP027, SP031, SP032, SP138, SP143, SP187
INT-C29	SP005, SP019, SP080, SP104, SP147, SP166
INT-C30	SP001, SP021, SP068, SP131, SP170, SP187
INT-C31	SP008, SP038, SP136, SP179
INT-C32	SP038, SP058, SP179
INT-C33	SP009, SP012, SP026, SP053, SP140, SP143, SP145, SP156, SP200
INT-C34	SP187, SP189
INT-C35	SP023, SP074, SP093, SP105, SP109, SP112, SP147, SP163, SP166, SP179
INT-C36	SP178
INT-C37	SP003, SP004, SP032, SP038, SP075, SP174, SP203
INT-C38	SP001, SP038, SP141, SP181, SP205
INT-C39	SP039, SP088, SP196, SP205
INT-C40	SP062, SP098, SP200
INT-C41	SP088, SP119, SP122, SP147
INT-C42	SP185
INT-C43	SP007, SP048, SP133
INT-C44	SP018, SP053, SP083, SP089, SP135, SP173, SP179, SP193
INT-C45	SP020, SP036, SP039, SP048, SP061, SP075, SP076, SP111, SP115, SP127, SP179, SP193, SP203
INT-C46	SP042
INT-C47	SP118
INT-C48	SP004, SP048, SP180, SP184
INT-C49	SP041, SP118
INT-C50	SP026, SP141, SP181
INT-C51	SP050
INT-C52	SP078
INT-C53	SP053, SP114
INT-C54	SP108, SP135
INT-C55	SP068
INT-C56	SP003, SP004, SP010, SP026, SP027, SP028, SP032, SP038, SP068, SP072, SP074, SP109, SP111, SP113, SP114, SP117, SP125, SP128, SP158
INT-C57	SP002, SP028, SP045, SP047, SP083, SP108, SP118, SP135, SP164, SP170, SP203
INT-C58	SP003, SP004, SP050, SP072, SP165, SP189, SP191
INT-C59	SP054, SP089, SP090, SP092, SP117, SP126, SP128, SP173, SP174
INT-C60	SP102
INT-C61	SP048, SP102, SP121, SP158, SP202
INT-C62	SP040, SP102, SP140

TTV/CA	References (Selected Papers – SP)
INT-C63	SP102
INT-C64	SP061
INT-C65	SP054, SP061, SP064, SP097, SP190
INT-C66	SP031, SP075, SP121, SP142
INT-C67	SP075, SP089
INT-C68	SP090, SP117, SP121
INT-C69	SP016
INT-C70	SP047
INT-C71	SP008, SP092, SP142
INT-C72	SP175, SP179
INT-C73	SP175
INT-C74	SP158
INT-C75	SP165
INT-C76	SP043
INT-C77	SP202
INT-C78	SP021
INT-C79	SP001, SP020
INT-C80	SP038
INT-C81	SP038, SP131
INT-C82	SP121
INT-C83	SP121
INT-C84	SP128
INT-C85	SP031
INT-C86	SP011, SP062, SP089
EXT-T01	SP001, SP002, SP004, SP005, SP006, SP008, SP009, SP010, SP011, SP012, SP013, SP014, SP015, SP016, SP017, SP018, SP019, SP020, SP021, SP022, SP023, SP024, SP025, SP026, SP027, SP028, SP029, SP030, SP031, SP032, SP034, SP035, SP036, SP037, SP038, SP039, SP041, SP043, SP044, SP045, SP046, SP047, SP048, SP049, SP050, SP051, SP052, SP053, SP054, SP055, SP056, SP057, SP058, SP059, SP060, SP061, SP062, SP063, SP064, SP065, SP067, SP068, SP069, SP070, SP071, SP072, SP074, SP075, SP076, SP077, SP078, SP080, SP081, SP083, SP084, SP085, SP086, SP087, SP088, SP089, SP090, SP092, SP093, SP094, SP096, SP097, SP099, SP100, SP102, SP104, SP105, SP106, SP107, SP108, SP109, SP111, SP112, SP113, SP114, SP115, SP116, SP117, SP118, SP119, SP120, SP121, SP123, SP124, SP125, SP128, SP130, SP131, SP132, SP135, SP138, SP139, SP140, SP141, SP142, SP143, SP144, SP145, SP146, SP147, SP148, SP149, SP150, SP151, SP152, SP153, SP154, SP155, SP156, SP159, SP160, SP161, SP163, SP164, SP165, SP167, SP168, SP169, SP172, SP173, SP174, SP178, SP179, SP180, SP181, SP183, SP184, SP185, SP187, SP188, SP189, SP190, SP191, SP192, SP193, SP194, SP195, SP196, SP197, SP199, SP200, SP201, SP202, SP203, SP204, SP205

TTV/CA	References (Selected Papers – SP)
EXT-T02	SP001, SP009, SP013, SP007, SP008, SP011, SP012, SP015, SP017, SP018, SP019, SP020, SP022, SP023, SP024, SP025, SP026, SP027, SP028, SP029, SP030, SP031, SP032, SP033, SP035, SP036, SP037, SP038, SP039, SP042, SP045, SP046, SP047, SP049, SP051, SP052, SP053, SP054, SP055, SP056, SP057, SP058, SP059, SP060, SP061, SP063, SP065, SP067, SP068, SP069, SP070, SP072, SP074, SP075, SP077, SP078, SP080, SP081, SP083, SP086, SP087, SP088, SP091, SP092, SP093, SP094, SP096, SP097, SP099, SP104, SP105, SP106, SP108, SP109, SP110, SP112, SP113, SP116, SP117, SP119, SP120, SP122, SP123, SP125, SP132, SP138, SP139, SP140, SP141, SP142, SP143, SP144, SP146, SP147, SP148, SP149, SP150, SP151, SP154, SP155, SP156, SP157, SP158, SP160, SP161, SP163, SP164, SP165, SP167, SP168, SP169, SP171, SP172, SP173, SP174, SP178, SP181, SP183, SP184, SP185, SP188, SP189, SP190, SP191, SP192, SP193, SP194, SP195, SP196, SP197, SP198, SP200, SP201, SP202, SP203, SP204
EXT-T03	SP001, SP064, SP091, SP106, SP138, SP152, SP196, SP205
EXT-T04	SP005, SP006, SP008, SP013, SP015, SP016, SP017, SP021, SP022, SP023, SP025, SP026, SP043, SP050, SP051, SP058, SP064, SP065, SP088, SP090, SP092, SP093, SP101, SP103, SP105, SP106, SP109, SP110, SP112, SP116, SP120, SP122, SP123, SP124, SP129, SP131, SP135, SP148, SP149, SP150, SP151, SP154, SP157, SP161, SP170, SP174, SP178, SP194, SP195, SP203
EXT-T05	SP001, SP007, SP009, SP010, SP012, SP020, SP034, SP037, SP038, SP039, SP040, SP043, SP048, SP053, SP059, SP062, SP063, SP067, SP068, SP082, SP085, SP088, SP101, SP102, SP114, SP115, SP126, SP128, SP135, SP140, SP153, SP155, SP156, SP157, SP165, SP179, SP192, SP196, SP200, SP201, SP202, SP203, SP205
EXT-T06	SP005, SP007, SP009, SP057, SP058, SP088, SP090, SP092, SP100, SP128, SP135, SP161, SP163, SP168, SP181, SP183, SP193, SP205
EXT-T07	SP026, SP031, SP036, SP058, SP074, SP076, SP078, SP108, SP125, SP168, SP170, SP171, SP174, SP198, SP203
EXT-T08	SP004, SP009, SP010, SP011, SP012, SP016, SP019, SP021, SP027, SP033, SP038, SP040, SP047, SP052, SP053, SP054, SP059, SP063, SP064, SP069, SP070, SP074, SP076, SP083, SP086, SP089, SP092, SP101, SP107, SP111, SP114, SP117, SP121, SP125, SP127, SP130, SP134, SP140, SP146, SP153, SP155, SP159, SP163, SP170, SP171, SP179, SP197, SP199, SP200, SP201
EXT-T09	SP050
EXT-C01	SP010, SP019, SP021, SP038, SP039, SP043, SP059, SP060, SP063, SP069, SP070, SP071, SP076, SP078, SP084, SP086, SP087, SP092, SP100, SP104, SP105, SP106, SP113, SP124, SP151, SP167, SP190, SP201, SP202
EXT-C02	SP005, SP014, SP032, SP036, SP046, SP064, SP067, SP083, SP096, SP112, SP113, SP114, SP132, SP141, SP145, SP147, SP152, SP165
EXT-C03	SP002, SP008, SP022, SP050, SP072, SP143, SP144, SP159, SP160, SP161, SP163, SP164, SP172, SP173, SP174, SP191, SP187, SP202
EXT-C04	SP028, SP045, SP051, SP052, SP080, SP085, SP090, SP112, SP115, SP120, SP131, SP142, SP149, SP160, SP178, SP181, SP189, SP200
EXT-C05	SP168, SP181
EXT-C06	SP001, SP008, SP015, SP019, SP027, SP025, SP026, SP028, SP032, SP035, SP037, SP046, SP047, SP052, SP053, SP055, SP059, SP063, SP065, SP067, SP070, SP074, SP078, SP086, SP104, SP106, SP117, SP122, SP138, SP139, SP141, SP163, SP164, SP165, SP171, SP172, SP173, SP174, SP183, SP189, SP195, SP197, SP198
EXT-C07	SP013, SP015, SP016, SP017, SP023, SP025, SP026, SP064, SP065, SP093, SP106, SP105, SP116, SP123, SP150, SP151, SP149, SP161, SP178
EXT-C08	SP024, SP125, SP147, SP202
EXT-C09	SP039, SP077, SP080, SP109, SP125, SP132, SP142
EXT-C10	SP148, SP195
EXT-C11	SP064, SP070, SP117, SP140, SP134, SP159
EXT-C12	SP058, SP125, SP171

TTV/CA	References (Selected Papers – SP)
EXT-C13	SP011, SP019, SP023, SP025, SP045, SP051, SP112, SP117, SP125, SP139, SP163, SP165, SP183
EXT-C14	SP113, SP119, SP125, SP160
EXT-C15	SP163, SP196
EXT-C16	SP012, SP034, SP048, SP053, SP114, SP128, SP196
EXT-C17	SP001, SP010, SP040, SP043, SP059, SP063, SP067, SP082, SP088, SP126
EXT-C18	SP007
EXT-C19	SP050
EXT-C20	SP021, SP038, SP059, SP069, SP076
EXT-C21	SP010, SP034, SP043, SP071, SP102, SP202
EXT-C22	SP088, SP174
EXT-C23	SP097, SP108, SP174
COT-T01	SP010, SP012, SP031, SP040, SP048, SP096, SP097, SP106, SP123, SP125, SP130, SP135, SP143, SP144, SP145, SP153, SP157, SP161, SP171, SP179, SP189
COT-T02	SP001, SP005, SP006, SP007, SP010, SP011, SP012, SP014, SP015, SP018, SP020, SP021, SP022, SP028, SP029, SP030, SP031, SP032, SP034, SP036, SP037, SP039, SP045, SP046, SP047, SP049, SP051, SP052, SP053, SP056, SP059, SP060, SP064, SP067, SP068, SP076, SP081, SP087, SP089, SP091, SP092, SP096, SP098, SP102, SP111, SP112, SP113, SP114, SP117, SP119, SP132, SP135, SP139, SP140, SP141, SP143, SP145, SP147, SP159, SP160, SP163, SP164, SP165, SP171, SP176, SP179, SP181, SP183, SP186, SP188, SP191, SP200, SP201, SP203
COT-T03	SP008, SP021, SP031, SP038, SP048, SP054, SP055, SP063, SP076, SP112, SP113, SP118, SP121, SP149, SP163, SP179, SP184
COT-T04	SP042, SP050, SP063, SP064, SP080, SP112, SP113, SP121, SP129, SP139, SP141, SP147, SP159, SP163, SP164, SP166, SP168, SP179, SP185, SP205, SP206
COT-T05	SP002, SP010, SP027, SP031, SP048, SP050, SP061, SP064, SP074, SP075, SP076, SP089, SP092, SP096, SP117, SP119, SP123, SP130, SP135, SP139, SP142, SP158, SP159, SP179, SP184, SP190, SP202
COT-T06	SP032, SP040, SP046, SP048, SP056, SP063, SP075, SP076, SP088, SP089, SP092, SP099, SP112, SP113, SP117, SP125, SP139, SP140, SP141, SP142, SP147, SP159, SP161, SP163, SP164, SP165, SP172, SP173, SP179
COT-T07	SP010, SP021, SP031, SP063, SP068, SP099, SP112, SP163
COT-T08	SP022, SP050, SP062, SP089, SP099, SP102, SP117, SP122, SP124, SP139, SP147, SP179, SP190
COT-T09	SP010, SP012, SP043, SP047, SP175, SP179
COT-T10	SP012, SP074, SP112, SP113, SP117
COT-C01	SP021, SP022, SP029, SP030, SP032, SP034, SP047, SP076, SP096, SP113, SP114, SP141, SP143, SP145, SP165, SP164, SP188, SP191, SP201, SP203
COT-C02	SP005, SP006, SP007, SP015, SP028, SP029, SP030, SP046, SP047, SP049, SP052, SP053, SP064, SP068, SP111, SP112, SP139, SP140, SP141, SP160, SP163, SP165, SP171, SP179, SP183
COT-C03	SP001, SP021, SP029, SP030, SP031, SP032, SP036, SP045, SP051, SP052, SP053, SP063, SP067, SP081, SP092, SP096, SP099, SP112, SP117, SP130, SP132, SP147, SP160, SP171, SP179
COT-C04	SP010, SP027, SP050, SP061, SP064, SP074, SP075, SP076, SP089, SP096, SP142, SP159
COT-C05	SP161, SP165, SP172, SP179
COT-C06	SP112, SP113, SP141, SP159, SP164
COT-C07	SP031
COT-C08	SP064, SP139

TTV/CA	References (Selected Papers – SP)
COT-C09	SP088, SP142
COT-C10	SP032, SP046, SP048, SP063, SP089, SP092, SP112, SP113, SP125, SP139, SP140, SP141, SP147, SP164, SP174, SP173, SP179
COT-C11	SP089, SP092
COT-C12	SP022, SP089, SP102, SP139, SP179
COT-C13	SP080
COT-C14	SP080
COT-C15	SP063, SP121, SP179
COT-C16	SP122
COT-C17	SP008, SP038, SP113, SP054, SP076, SP112, SP184
COT-C18	SP055, SP118
COT-C19	SP047, SP112
COT-C20	SP010
COT-C21	SP179
COS-T01	SP001, SP006, SP007, SP010, SP012, SP013, SP014, SP015, SP017, SP019, SP020, SP021, SP031, SP038, SP039, SP040, SP045, SP047, SP048, SP052, SP053, SP056, SP059, SP063, SP064, SP072, SP091, SP096, SP097, SP099, SP114, SP123, SP129, SP132, SP134, SP135, SP136, SP139, SP141, SP142, SP144, SP145, SP147, SP157, SP162, SP163, SP165, SP179, SP181, SP183, SP185, SP186, SP187, SP189, SP190, SP194, SP201, SP202, SP203, SP205
COS-T02	SP008, SP011, SP012, SP021, SP022, SP028, SP031, SP032, SP036, SP046, SP050, SP051, SP059, SP064, SP067, SP068, SP099, SP111, SP112, SP113, SP114, SP115, SP117, SP125, SP132, SP139, SP141, SP147, SP163, SP164, SP165, SP179, SP183, SP187, SP188, SP189, SP190, SP191, SP201, SP205
COS-T03	SP026, SP031, SP050, SP061, SP063, SP092, SP112, SP114, SP117, SP142, SP147, SP173, SP174, SP179, SP188, SP191, SP200, SP203
COS-T04	SP001, SP007, SP012, SP019, SP022, SP029, SP035, SP041, SP047, SP048, SP049, SP052, SP056, SP059, SP067, SP068, SP069, SP073, SP074, SP078, SP079, SP084, SP087, SP094, SP099, SP104, SP115, SP118, SP125, SP127, SP135, SP139, SP141, SP146, SP161, SP163, SP164, SP166, SP167, SP175, SP176, SP184, SP186, SP199, SP201, SP205
COS-T05	SP019, SP021, SP051, SP059, SP061, SP063, SP074, SP078, SP080, SP098, SP109, SP112, SP117, SP122, SP135, SP139, SP158, SP179, SP189, SP190, SP206
COS-T06	SP036, SP179, SP185
COS-T07	SP003, SP021, SP082, SP098, SP139, SP147, SP179
COS-T08	SP012, SP064, SP115
COS-C01	SP011, SP012, SP022, SP028, SP031, SP032, SP036, SP046, SP050, SP051, SP059, SP064, SP067, SP068, SP099, SP111, SP112, SP113, SP114, SP115, SP117, SP125, SP132, SP139, SP141, SP147, SP163, SP164, SP165, SP183, SP187, SP188, SP189, SP190, SP191, SP205
COS-C02	SP001, SP010, SP031, SP038, SP040, SP048, SP059, SP063, SP072, SP096, SP097, SP114, SP132, SP135, SP139, SP162, SP163, SP165, SP183, SP187, SP190, SP202
COS-C03	SP006, SP007, SP010, SP013, SP015, SP017, SP019, SP021, SP039, SP048, SP063, SP091, SP129, SP142, SP187, SP202
COS-C04	SP050, SP061, SP063, SP117, SP173, SP174, SP179, SP188, SP191, SP203
COS-C05	SP013, SP019, SP039, SP114, SP129, SP187, SP202
COS-C06	SP019, SP059, SP112, SP135, SP139, SP190
COS-C07	SP114, SP141, SP144, SP145, SP147, SP165, SP179, SP203

TTV/CA	References (Selected Papers – SP)
COS-C08	SP098, SP139, SP147
COS-C09	SP125
COS-C10	SP165
COS-C11	SP021, SP061, SP074, SP117, SP122
COS-C12	SP059, SP061
COS-C13	SP122
COS-C14	SP061
COS-C15	SP080
COS-C16	SP013, SP014, SP064, SP099
COS-C17	SP012, SP064, SP115
COS-C18	SP109
COS-C19	SP117, SP139