2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement

Describing What Experimental Software Engineering Experts Do When They Design their Experiments – A Qualitative Study

Liliane Fonseca, Sergio Soares
Informatics Center (CIn)
Federal University of Pernambuco (UFPE)
Recife, Brazil
{lss4, scbs}@cin.ufpe.br

Carolyn Seaman
Department of Information System
University of Maryland, Baltimore County
Baltimore, United States
cseaman@umbc.edu

***Abstract*— Background:** Although there has been a significant amount of research focused on designing and conducting controlled experiments, few studies report how experienced experimental software engineering researchers actually design and conduct their studies. **Aims:** This study aimed to offer a practical perspective from their viewpoint regarding controlled experiment planning. **Method:** We collected data through semi-structured interviews from 11 researchers, and we used qualitative analysis methods from the grounded theory approach to analyze them. **Result:** Although the complete study presents four research questions, in this paper, we answer the first one. As a result, we present a preliminary result about what these experts actually do when they design experiments. **Conclusions:** This work contributes to a better understanding of the practical performance of experimental software engineering.

Keywords—Qualitative Study; Experiments; Human Subjects; Software Engineering.

I. INTRODUCTION

Experimental planning is a process that specifies how experimenters will carry out their experiments. It involves determining under exactly what conditions the experiment is to be conducted, which variables can affect the experiment, who is going to participate in the study, how many times it is to be repeated, and so on. A well conducted experimental planning minimizes costs and bias [1]. Although the planning process iterates until a complete experiment design is ready [2], some researchers, especially beginners, are not aware when it is ready to run. Once an experiment is started, lack of important elements in the experimental plan and wrong decisions made by experimenters can affect the experiment results, thus, leading to false or imprecise results [3]. Although the literature specifies some guidelines to conduct experiments in software engineering [1], [2], [4], [5], [3], few focus on a practical perspective from experienced empirical software engineering researchers regarding how they actually plan their experiments.

Qualitative study methods came from social sciences to help researchers understand human behavior [6], and they have been used in software engineering through action research, case study research, ethnography, and grounded theory. Seaman [7] presents several research methods for collecting and analyzing qualitative data, and describes how these qualitative methods can be used in empirical studies of software engineering.

Because we wanted to extract information from experienced software engineering researchers that were interviewed only one time, conducting a qualitative study using semi-structured interviews is the appropriate study method because semi-structured interview provides reliable and comparable qualitative data through a clear set of instructions for investigators. We conducted 11 interviews, which were audio recorded and transcribed. To perform the data analysis, we chose to use open and axial coding from grounded theory because they allowed us to associate codes with quotes and then analyze the coded quotes through merging and identifying relationships between them [8]. The complete goal of this study is to understand what experienced software engineering researchers actually do when they design their experiments, including what kinds of problems or traps they fall into, how they currently learn about research methods, and what gaps they have in their knowledge. However, in this paper, we present the results from the first question, which is, what experienced software engineering researchers actually do when they design their experiments. In this work, we delimited the scope of this research considering only experiments using human subjects who are randomly assigned, which means, the participants were aware that questions should be answered regarding the context of controlled experiments.

In this paper, we described the preliminary results about qualitative interview study performed to understanding how experienced experimental software engineering researchers plan their experiments. It is organized into the following six sections: Section 2 reports the research methodology; Section 3 describes the preliminary results and principal findings; Section 4 presents the discussion; Section 5 presents the trustworthiness of this study, and Section 6 presents the conclusion of this work.

II. METHODOLOGY

In this research, the overall focus is to describe the experiment planning process according to experienced experimental software engineering researchers' point of view regarding how they actually plan their experiments in practice.

978-1-5090-4039-1/17 $31.00 © 2017 IEEE
DOI 10.1109/ESEM.2017.63

In order to do this, we used the GQM goal template [9] to structure the following general goal of the study:

Analyze the process of planning experiments,
for the purpose of describing how experts plan their experiments,
with respect to how experiment planning is done,
from the viewpoint of experimental experts,
in the context of software engineering.

A. Research Questions

From the general goal, we specified four sub goals, which are also our research questions (RQ), and we applied the GQM template to each one. We present the research questions as follows:

- **RQ1-** What experienced software engineering researchers actually do when they design their experiments?
- **RQ2-** What kinds of problem/traps do the experts fall into?
- **RQ3-** How do experts currently learn about experiment planning?
- **RQ4**- What gaps do experts have in their knowledge?

However, as previously mentioned, in this paper, we answer only the first research question. Soon, the remain questions will be analyzed, and the results will be reported.

B. Interview Design

The questions were developed, discussed with a researcher mentor, piloted, and adjusted until the final version of the questionnaire was created. Nine interview questions were generate for RQ-1. The complete questionnaire list contains 24 questions, which includes questions for the four research questions. The questions were organized in order to achieve a natural flow during the interview. At the end of the interview, the interviewees had the opportunity to express any additional thoughts, comments or lessons learned from their experience in experiment planning. The interview questions regarding the RQ1 are presented in Table I.

TABLE I. INTERVIEW QUESTIONS

ID	Interview Questions
1	From your perspective, what is an experimental plan and what are the main elements that should be contained in it?
2	Do you use any written experimental process or guidelines to plan experiments? If so, which ones.
3	What other guidelines for planning experiments do you know about?
4	What other types of guidance or tools would be useful to help inexperienced researchers to learn how to do experiment planning?
5	What are the characteristics of an experimental plan that you think are most important to help optimize the effectiveness of an experiment?
6	In your opinion what are the acceptance criteria for the quality of the experiment plan?
7	Does planning an experiment make the experiment more effective (as compared to doing an experiment with no planning)? Why?
8	Does planning an experiment make the experiment more effective (as compared to doing an experiment with no planning)? Why?
9	Does it make sense to track the amount of time it takes to plan an experiment? If yes, What is the average time required to plan an experiment?

The duration of each interview was designed to last around 60 minutes. This time was respected in all interviews, and it was not necessary to stop any interview because the time was up. We started the participant's selection by sending an invitation e-mail to experienced experimental software engineering researchers based on their experience in conducting experiments and their published experiments in software engineering. We also tried to select experts from different research groups around the world in order to identify distinct ways of planning experiments. Although snowball sampling was not done, we asked those initial researchers who could not participate in the study, to suggest other researchers who could participate in our study. As a result, 20 researchers from several countries, such as Brazil, the United States, Canada, Germany, England, and Spain were contacted. Eleven out of 20 experts were available to participate in this study. Three experts were from U.S.A, one from England, one from Spain, and six from Brazil. The six participants from Brazil were interviewed in Portuguese; the rest of the interviews were conducted in English. Ten interviews were remotely conducted via Skype while one interview was performed face to face at UMBC because of the availability of the interviewee. They each have relevant studies published in the empirical software engineering literature, with more than ten years of experience. Although the majority of subjects are academics, some of them are empirical researchers from industry.

A pilot study was performed in order to test materials and procedures for verifying potential problems and the flow of the interview. We performed two pilots. The first one was accomplished in Portuguese with a Ph.D. student in empirical software engineering from a Brazilian university. The second pilot was accomplished in English. We used a senior researcher in empirical software engineering from a U.S.A university. After the pilots, the participants gave feedback about the study related to the content of the questionnaire, flow, tone and body language of the interviewer. In addition, we verified the duration, record resources, and environment of the interview. We concluded that the study was designed suitable to the goals of the study.

C. Data Collection

The data collection was split into two phases: interviews with Brazilian experts and interviews with experts from other nationalities (U.S.A, England, and Spain). We performed 11 interviews from 20 invitations. The interviews were transcribed in the same language as was used during the interview. That means, interviews from phase 1 were transcribed in Portuguese and the interviews from phase 2 were transcribed in English. The interviews were conducted via Skype, and were audio recorded. Regarding the confidentiality of data a code was associated with each research participant to protect personal privacy. Only one table contains both names and codes, and this table is kept in electronic form only on a password protected computer. Analysis of the data used only codes, no names. The results are published without reference to any participants` names. The original recordings and data files were stored on a secured shared drive accessible only to the investigators.

D. Data Analysis

The data was transcribed by the principal investigator of this study, and all transcripts were double checked by two reviewers, one Portuguese native speaker researcher for Portuguese transcripts, and one English native speaker researcher for English transcripts. The reviewers were the advisors of this research. They randomly picked up pieces of the transcripts and checked them by listening to the audio and reading the transcriptions. Because the results of this research are published in English, we did a free translation of the quotes from Brazilian experts used in the results section. In order to assure credibility in the translations, they were checked by another researcher.

To perform the data analysis, we chose to use open and axial coding from Grounded Theory [8]. Because it was not our goal to create a theory from this study, we did not use selective coding. Open coding begins with an intensive reading of the transcripts. Some seed codes were created from brainstorming meetings, then relevant words, phrases, sentences, and sections were labeled with the codes previously created. However, during the open coding process, new codes were created from relevant ideas, for example, something that was repeated several times, something that surprised the coder, in case that the interviewee explicitly states it was important, or some other reason that the coders think it was relevant. During the open coding process, the list of codes increased. The open codes regarding the first research question of this study, (RQ1) What do experienced software engineering researchers actually do when they design their experiments?, are: what they do; experiment design process; support mechanisms used; importance of experiment planning; perception from empirical software engineering community; acceptance criteria, elements, time, experimental plan quality, experimental plan definition.

An Excel spreadsheet was used to facilitate the open coding phase. We performed the open coding analysis line-by-line highlighting what was important. We did a second round of open coding in order to ensure we had not previously missed any important categories. At the end of the open coding phase, we classified the transcripts, that is, we grouped some selected quotes into relevant factors for each question to be analyzed in the axial coding phase. In the axial coding phase, we identified relationships among the data that had been broken into pieces by the open coding process. We performed this phase within the Excel spreadsheet. Additionally, memos were an important part of analysis process. We wrote a series of them to capture thoughts and connections.

III. Results and principal Findings

In this section, the results from the interviews and principal findings about what experienced software engineering researchers actually do when they design their experiments are presented, and it is organized around the major quotes from our analysis. Additionally, we present a conceptual model about what experts actually do when they design their experiments. The participants described several distinct pathways for experiment planning. However, the activities that they practiced are almost the same. We grouped these activities and we developed a conceptual model that captures all the variations the participants said they do when they design their experiments. These activities are: brainstorming about experiment rationales; using support mechanisms; writing and updating an experimental plan; revising the experimental plan; meetings with external researchers; running pilots; and discussing results from pilots. An overview of a conceptual model depicting these activities is provided in Figure 1.

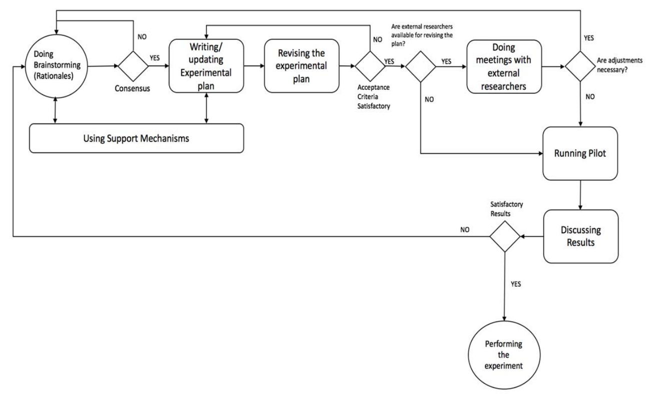

Fig. 1. Conceptual model about what experts actually do when they design their experiments

A. Brainstorming about experiment rationales

For experts, brainstorming about experiment rationales is the first important step in planning experiments. Before they start to plan, they have multiple meetings to discuss all the design decisions that they will be making. These decisions are described as follows:

"We discuss all the design decisions that we will be making like who should we contact, who should we talk to, what data shall we collect and what problems will there be? ...That may take multiple meetings but, usually, we try to reach the point where everybody involved in the research project is happy with the plan and the assignment."

Another participant provided more detail about the specific topics that need to be addressed in these initial brainstorming meetings, such as potential confounding factors and risks. The participants also indicated that this initial brainstorming contributes to important elements in the experimental plans:

"...what the goal is to accomplish, what question that I`m trying to answer. ... thinking about the type of data that I want to gather and collect... what people are we going to talk to, what software project are we going to look at... some issues about the sample, what type of treatments we`re going to use, what type of metrics we`re going to calculate, what type of statistical tests we`re going to use, and..."

During the meetings, the experts also ask themselves about some important elements in the experiment. Some of these elements are presented at goo.gl/qz5Qay.

B. Writing and updating experimental plans

According to the interviewees, it is hard to achieve a relevant result without planning an experiment, in fact that the plan is "*critical*". However, there is no unanimity regarding how they document experimental plans. Some participants follow a sequence of activities suggested by empirical literature. While other experts do not follow such guidelines.

"I'm aware a lot of those guidelines that exist but I don't really follow a set process as prescribed by them."

However, all of participants agreed that the rationales of the experiment plan discussed in the meetings should be documented. It avoids the experimenter becoming "lost in the middle" of experiment and being surprised with issues that can appear during the experiment.

C. Revising the experimental plan

Revising experimental plans was considered for several participants an important activity to find mistakes in the experimental plan. One participant said that in this activity he reviews different kind of issues including risks, participants, training, among others.

" there is all sort of risks we can look at, you can look at , do you have the right participants? What knowledge do these participants need to know to be able to do tasks you have, are there including criteria that you need to sort of reduce those risks, how much training are the participants going to need to be able to use the tools that you are working with, does that training sort of address the issues that they might potentially have in using the tool."

Most of the participants pointed out completeness as one of the acceptance criteria for the quality of the experimental plan.

"So the quality is the completeness, I think, as long as everything is in there, I think the quality of the experimental plan should be good."

"Its completeness, its coverage regarding the phenomena which is being observed."

They described completeness as a set of information that should be contained in the experimental plan. Another factor that they frequently review was the rationale of the decisions made.

"I think going through each part of the research plan and making sure that there`s a reason why we are taking each step, so there`s some justification, some rationale, for each step in the experimental plan."

D. Using Support Mechanisms

Some of the participants reported that they use support mechanisms for writing experimental plans and planning the statistical analysis.

"I use literature when I do experimental planning."

Although not all interviewed experts plan their experiments using the empirical literature, because some of them rely on their own personal experience, all of them have agreed that empirical literature is an important support, especially for beginner researchers. One participant said that usually he follows guidelines to report or replicate experiments but he does not follow guidelines to plan his experiments.

"When we are reporting experiments, we follow Jedlitschka's guidelines, if you are reporting for example replication, we use Jeffrey Carver guidelines, but when we are doing the plan itself we do not follow any kind of guidelines. "

Another interesting finding is that most of the experts reported that they read similar studies, especially the lessons learned sections to try not to make the same mistakes reported in the previous studies.

"I will look at how other researchers made their experiments, how they were organized, what was going wrong and right, the lessons learned section."

As mentioned before, not all of experts use support mechanisms for guidance for their experiments, they do the experimental activities based on their experience. However, in cases where they do not use any support mechanisms anymore, we asked them to suggest good experimental materials useful for beginner researchers. Therefore, in the following, we describe a mix of support mechanisms that the interviewees have the habit of using and mechanisms that they have used in the past and have suggested for their students and interns.

Wohlin et al. [2], Juristo and Moreno [1], Kitchenham et al. [4] are the books most cited as a guideline by the experts interviewed. They reported that Kitchenham et al. [4] provides more of a checklist of things that should be carried out in the experiments than the first two books. Shull et al. [12] is a book cited as a resource that covers the common difficulties and challenges faced by software engineering researchers who want to design their empirical studies. It provides guidance and information for research methods and techniques. Pfleeger [10], Zelkowitz [11], Travassos [15] are books cited by experts that provide general guidance for researchers. Basili et al. [9] is a template largely used by experts to structure the goal of an experiment, and Seaman [7] is cited as a support in qualitative analysis. Shadish et al. [13] and Cook and Campbell [14] guide them on the design of experiments and quasi-experiments. Ardelin Neto and Conte [16] is a tool suggested to support the identification of threats to validity and actions to control them in the planning stages of controlled experiments. Lopes and Travassos [17], Freire et al. [18], Travassos et al. [19], and Arisholm et al. [20] are proposed environments to support conducting experiments. Inc. Minitab[1], SAS Institute Inc.[2], R Development Core Team [22], IBM corp.[3] are statistics tools used to analyze quantitative data. Jedlitschka et al. [5] is a guideline to report experiments in software engineering, and Carver and Tuscaloosa [21] is cited as a guide for experiment replications.

[1] Inc. Minitab. Minitab 17 statistical software, 2010.

[2] SAS Institute Inc. Jmp statistics tool, 2007.

[3] IBM corp. Ibm spss statistics, 2015.

E. Meetings with external researchers

The experts reported the practice of having their experimental plans reviewed by experienced external researchers who are not involved in the experiment planning. They also presented their planning in workshops with graduate students who are able to think critically and make relevant comments. The goal of these meetings is to *"try to identify threats to validity, and things that could go wrong"*. As one participant noted, the advantage of this practice is simply *"more people thinking"*.

F. Running pilots

Running pilots, even using a small scale, was cited as a good practice performed by the experts interviewed. They think that running a pilot is essential, especially for controlled experiments using human subjects.

"So I think piloting probably is the biggest and the most important step for me."

Also, during the pilot, experimenters can see whether those risks that they imagined in the initial meetings are actually going to be a big issue or not. One expert said that he observes how everything behaves.

"Given what I want to learn, here is what I want the participants to be doing. I`ll run a pilot study. It is that actually sort of, what`s happening, and based on what they are actually doing. I`m actually sort of seeing the activities that they want to see that will help me sort of address the research questions."

When asked to give an example of mistakes that they committed, and caught during a pilot, one expert answered:

"About mistakes: Not documenting properly the design, that was an important mistake that we made. Another one was, that I mentioned about which rule we were discussing or the design we have two choices. First we thought ok, this is the best option but then we thought we could not make that choice because we need too many subjects, too many experimental subjects, so we decided very fast we change the design and because we run a pilot we discover that the change we made was wrong."

This example illustrated the importance of writing a good experimental plan and running a pilot during experiment planning. Because experiment planning is an iterative process, it is difficult to say that an experimenter has a complete experimental plan until they run the experiment. However, running a pilot comes as a good practice for revealing many problems including confusing experimental materials, procedures, tasks, questionnaires, and instructions. It is important to run the pilot as close as possible to conditions in which the experiment will be conducted, including the environment, participants, materials, among others.

"Although there is never really a point where a researcher can say I have a complete plan for the experiment formulated until I actually run the experiment because is kind of always this iterative process so for me a lot of it is doing things like fine-tuning the tasks."

G. Discussing results from pilot

The experts reported that they discuss the results from pilots with their research group because at this stage they are already aware of most of the problems that might arise. In this activity, depending on the pilot results, they decide what they must do. In some cases they have brainstorming meetings, in other cases, they just update their experimental plan and run their experiment, thus assuming they already know about most of the threats to validity that their experiment is exposed to. Also, they said that in this stage because they already know about the trade-offs in running the experiment, it is easier that they make changes to the experimental plan than in the experiment execution.

IV. Discussion

We identified that experts design their experiments by doing six activities: (1) brainstorming experiment rationales, (2) writing and updating the experimental plan, (3) revising the experimental plan, (4) holding meetings with external researchers, (5) running pilots, and (6) discussing results. From that, we developed a conceptual model, Figure 1 to illustrate the experiment planning activities that are performed by experts in order to help inexperienced researchers to plan their experiments. In the empirical software engineering literature, the experiment planning process follows a set of important steps such as context and variables definition, sample selection, hypothesis formulation, choice of design type, instrumentation, and validity evaluation [2]. Our results show that although the experts include these steps in their experiment planning, they carry out those activities that allow them to critically and efficiently think, write, review, run (even in a small scale), and discuss the experiment that they are developing. Therefore, the findings are an addition to the experiment planning process found in the empirical literature because these activities can help researchers to plan their controlled experiments efficiently.

V. Trustworthiness of this Study

We addressed some criteria for ensuring trustworthiness in this qualitative interview study including, credibility, dependability, and transferability.

Regarding credibility in terms of research bias, we were not interested in fitting our data into any particular conclusion. In contrast, we wanted to conduct a well designed qualitative study which captures how experienced experimental software engineering researchers actually do their experiments in practice, and present this result to the empirical software engineering community. The participants were selected based on their experiences in conducting experiments in software engineering and their relevant publications in the experimental software engineering field. After that, they showed their willingness to be interviewed. We were careful to prepare the qualitative interview questions using the GQM template in order to take maximum advantage of this interview opportunity. We were also careful to avoid leading questions that might cause embarrassment. In terms of data collection and data checking, all interviews were audio-recorded and transcribed in order to capture the participants' contributions accurately. All transcripts were checked by another researcher who was a speaker of the language that the participants used in the interviews, that is, Portuguese interviews were checked by a Brazilian researcher, and English interviews by an American researcher. In terms of additional peer review, the quotes used in the results from Brazilian experts were translated and reviewed by another researcher who has language skills both in Portuguese and English. Regarding dependability in terms of consistency, we recorded each step on spreadsheets in case someone else wants to audit the findings, which can be used as audit trails by external reviewers. In terms of coding and recoding, we coded and recoded the data several times both during the open coding and axial coding phases, in which we compared the various sets for completeness and consistency of data. Regarding transferability, we provided descriptions of the context in which the research was performed.

VI. Conclusions

This study aimed to answer the first research question of this research in order to offer insights into how experienced software engineering researchers actually plan their experiments. Interviews were conducted to understand how they deal with experimental issues in a practical way. As a result, we developed a conceptual model about how they actually plan experiments, including the common activities which help them to achieve satisfactory results. The qualitative study has two limitations. First, although the conceptual model about how experts actually plan their experiments is a strong result, these findings should be complemented with an empirical investigation to determine the extent to which these activities extend to experimental planning performed by inexperienced researchers. Second, more interviews should be conducted with other researchers from other companies and universities in order to try to understand how the empirical research communities in software engineering, as well as in other fields, carry out experiments in practice.

In future works, we will provide answers about the remain research questions, which are, an overview of ways in which experts learn about empirical studies, the list of common problems that they have faced, mistakes that they have already experienced, gaps both in their knowledge and in the empirical literature as well as things that they would like to have in the empirical software engineering field. Additionally, we will present some experts' opinions of experiment planning in empirical software engineering.

Acknowledgment

This work is partially supported by INES subsidy CNPq 141705/2015-9 and 465614/2014-0, FACEPE/APQ/0388-1.03/14. Liliane Fonseca is supported by CNPQ and FACEPE. Sergio Soares is partially supported by CNPq– 304499/2016-1.

References

[1] N. Juristo and A. M. Moreno. "Basics of Software Engineering Experimentation," (1st ed.). Springer Publishing Company, Incorporated. 2010.

[2] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. "Experimentation in Software Engineering". Springer, 2012.

[3] A. J. Ko, T. D. LaToza, and M. M. Burnett, "A practical guide to controlled experiments of software engineering tools with human participants". Empirical Software Engineering, 20(1):110–141, 2015.

[4] B. A. Kitchenham, S. L. Pfleeger, L. M. Pickard, P. W. Jones, D. C. Hoaglin, K. E. Emam, and J. Rosenberg. "Preliminary guidelines for empirical research in software engineering." IEEE Trans. Softw. Eng., 28(8):721–734, aug 2002.

[5] A. Jedlitschka, M. Ciolkowski, and D. Pfahl. "Reporting guidelines for controlled experiments in software engineering", 2007.

[6] N. K. Denzin and Y.S. Lincoln. "The SAGE Handbook of Qualitative Research". SAGE. Publications, 2011.

[7] C. B. Seaman. "Qualitative methods in empirical studies of software engineering." *IEEE Trans. Softw. Eng.*, 25(4):557–572, July 1999.

[8] A. L. Strauss and J. Corbin. "Basics of qualitative research : techniques and procedures for developing grounded theory." Sage, Thousand Oaks, 2nd edition, 1998.

[9] V. R. Basili, G. Caldiera, and H. D. Rombach. "Goal question metric paradigm." In *Encyclopedia of Software Engineering*, pages 1: 528–532. John Wiley and Sons Inc., 1994.

[10] S. L. Pfleeger. "Experimental design and analysis in software engineering." *Annals of Software Engineering*, 1:219–253, 1995.

[11] M. V. Zelkowitz, D. R. Wallace, and D. W. Binkley. "Experimental validation of new software technology". In Natalia Juristo and Ana M. Moreno, editors, *Lecture Notes on Empirical Software Engineering*, pages 229–263. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 2003.

[12] F. Shull, J. Singer, and D.I.K. Sjøberg. "*Guide to Advanced Empirical Software Engineering*." Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[13] W. R. Shadish, T. D. Cook, and D. T. Campbell. "*Experimental and Quasi-Experimental Designs for Generalized Causal Inference"*. Houghton Mifflin, Boston, 2002.

[14] T. D. Cook and D. T. Campbell. "*Quasi- Experimentation: Design & Analysis Issues for Field Settings"*. Houghton Mifflin Company, Boston, 1979.

[15] G. Travassos, D. Gurov, and E. Amaral. "Experimental software engineering - an introduction(inportuguese).", SETechnicalReport(ES-590/02-Abril),COPPE/UFRJ,Brazil, 2002.

[16] A. Ardelin Neto and T.U. Conte. "Identifying threats to validity and control actions in the planning stages of controlled experiments." In *Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering*, SEKE 2014, pages 256–261, 2014.

[17] V. P. Lopes and G. H. Travassos. "Knowledge repository structure of an experimental software engineering environment." *Brazilian Symposium on Software Engineering*, 2009.

[18] M. A. Freire, P. Accioly, G. Sizilio, E. Campos, U. Kulesza, E. Aranha, and P. Borba. A model-driven approach to specifying and monitoring controlled experiments in software engineering. In *Product-Focused Software Process Improvement - 14th International Conference, PROFES 2013, Paphos, Cyprus, June 12-14, 2013. Proceedings*, pages 65–79, 2013.

[19] G. H. Travassos, P. S. M. S. Santos, P. G. Mian, A. C. D. Neto, and J. Biolchini. "An environment to support large scale experimentation in software engineering." In *Proceedings of the 13th IEEE International Conference on Engineering of Complex Computer Systems*, pages 193–202. EEE Computer Society, 2008.

[20] E. Arisholm, D. I. K. Sjøberg, G. J. Carelius, and Y. Lindsjørn. "A web-based support environment for software engineering experiments. " *Nordic J. of Computing*, 9(3):231–247, September 2002.

[21] J. C. Carver and A. Tuscaloosa. "Towards reporting guidelines for experimental replications: A proposal", 2010.

[22] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.