Using Focus Group in Software Engineering: lessons learned on characterizing software technologies in academia and industry

Breno Bernard Nicolau de França, Talita Vieira Ribeiro, Paulo Sérgio Medeiros dos Santos and Guilherme Horta Travassos

COPPE / Universidade Federal do Rio de Janeiro
{bfranca,tvribeiro,pasemes,ght}@cos.ufrj.br

Abstract. CONTEXT. Focus Group (FG) has been adopted to support researches in different areas, offering instruments to simultaneously collect qualitative data from a group of people. Even peopleware representing an important perspective in software engineering (SE), FG is not largely used to support empirical studies in the field due to lack of understanding on its use. OBJECTIVE. To investigate the use of FG as supplementary technique to support the characterization of sociotechnical and cognitive aspects of software technologies into the context of SE research and practice activities. METHOD. To use FG to support the qualitative characterization of two distinct software technologies, in both academia and industry perspectives. From these two experiences, to identify commonalities and differences aiming at improving the understanding of FG usage when applied in SE studies. RESULTS. Focus Group was able to support both studies, however the object of study, environment, and participants influenced their design. Similarities and differences in terms of context, goals, design and threats to validity between the two experiences were observed, resulting in a set of lessons learned and possible improvements for future trials or new opportunities. CONCLUSIONS. Although some contexts can make hard the subjects' motivation and commitment, FG can promote interaction and raise important discussions among subjects therefore contributing to reveal information usually difficult to obtain when using only pure observations or feedback questionnaires. These are relevant characteristics when observing SE phenomena and can be used as an interesting tool to data triangulation. Despite its benefits, FG should be carefully planned to avoid validity threats regarding its execution and data collection.

Keywords: Focus Group · Empirical Software Engineering.

1 Introduction

Focus Group (FG) is a type of qualitative technique used to collect data through organizing group interviews. Having its roots in the Marketing research, FG has been widely and increasingly employed as source of primary or supplementary data for different types of research questions in most of social and health sciences disciplines.

This interest stems from FG distinctive interviewing features, which are characterized by its focus on a specific topic and its capitalization from the interaction that occurs within group settings mediated by a moderator or facilitator [1]. The group settings are one of the main methodological concerns of FG as they determine the facilitator role, questions and activities to be conducted, and most importantly, how they can encourage the group communication to reveal individual ideas, feelings, believes and differences in perspective about the addressed issues [2].

Comparing to other individual or group interviewing techniques, it is vital that FG studies concentrate on nondirective questions that can left room for the facilitator guidance and be used to elicit spontaneous expression among the participants [3]. There is abundant technical literature on FG regarding methods issues such as designing interview guides and structuring and moderating groups [3]. Even in Software Engineering (SE), where FG studies are relatively more recent and scarce, contextualized orientation for conducting FG in the area is already available [4]. Although the general FG's best practices are imperative for good quality and rigorous studies, SE still lacks specific discussions on how these general FG orientations can be exploited to investigate the use of SE technologies, involving aspects like team dynamics and individual engineering rationale, usually lost in other types of research methods.

It is hypothesized that SE particular blending of sociotechnical aspects affects how FG procedures can be applied. Besides, this application is interesting in terms of how practitioners and researchers can be stimulated to discuss about software technologies, how questions and activities should be structured to bring out unknown intricacies of their application, or even how the facilitator should proceed to drive the discussions to focus on typical SE research variables. In this paper, we analyze and discuss such aspects through the reflections on two FG experiences in the context of SE research and practice. The idea is not to make their comparison since they are un-related, but to abstract from them under a common FG process aiming at understanding how such research strategy can be used to characterize SE technologies in both environments. Such experiences indicate the FG feasibility as being a useful strategy to collect qualitative data regarding the application of SE technologies, regardless the intrinsic validity threats and possible mitigation actions over the studies' planning and designing.

The rest of this paper is organized in the following subsections. Section 2 presents the research method adopted for both experiences we conducted on SE research and practice. Section 3 presents the first FG we performed on a software company setting concerning guidelines for source coding standardization. Section 4 presents the second FG, concerning the evaluation of planning guidelines for simulation experiments in SE. Section 5 presents a discussion on the main methodological issues we faced in both experiences as well as the lessons learned. Section 6 presents the conclusions.

2 Research Method

For the analysis performed in this paper, we abstracted from both experiences under the same perspective: activities defined for the adopted FG process (Fig.1). This way, we analyzed how these activities were performed to understand how they supported

the studies w.r.t. the characterization purpose in their different settings. We executed specific FG activities adapted from [4], on both FG studies. Additional activities and steps were also included, once we understand the FG technique is a group dynamics strategy as well as the basis for a primary study. Thus, general tasks regarding primary studies such as object of study and goals definition, and activities such as planning assessment and information packaging presented in [5] were added in the FG process.

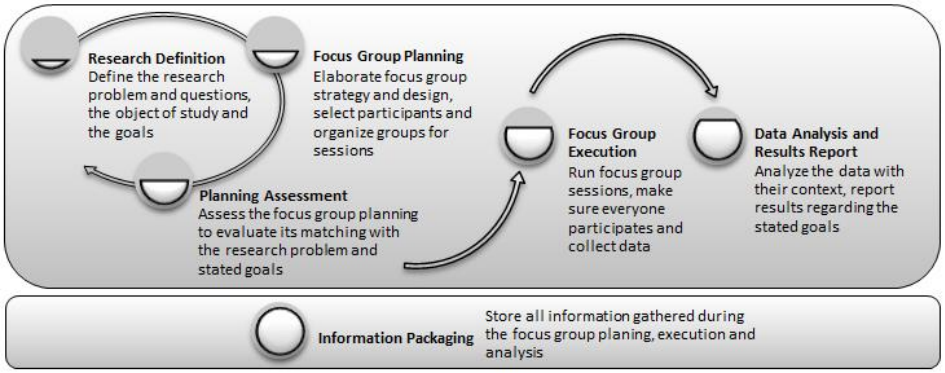

Fig.1. Focus group process adapted from activities and steps presented in [4] and [5]

The FG process has six major activities. During the research definition, the researcher has to identify the research problem, questions and context, and the reasons for choosing the FG technique so it can be possible to check whether it suites his needs for information and the environment in which the study will be conducted. Afterwards, in planning, the researcher has to define the FG strategy and design, that is, selection of participants and moderators, group settings and design of the participants' interactions. All information planned has to be assessed, so the strategy and design of the study remain in accordance with the research problem and questions. When necessary, these activities can be redone until the plan is ready to be executed.

The FG session should follow its plan and it is important for the moderator to ensure the participation of all involved in the study and takes notes of important contributions that may help answering the stated research questions. The final activities include the analysis, report and packaging of all collected data, including the session execution context and threats to validity of the study, so the researcher is able to combine data and the raised understanding concerning with the object of study.

Next two sections describe the experiences on using FG in the industry and academia. We decide on FG instead of feedback questionnaires for the evaluation of perceived usefulness and ease of use, using questions as in [6]. The justification lies on the possibility of getting more than quantitative responses, understanding real difficulties on both guidelines application and promoting discussion regarding the characteristics under focus and improvement opportunities. Still, time constraints forced us to reduce the scope of observation w.r.t. which coding or planning guidelines should be debated, since meetings usually last up to 3 hours to avoid participants' exhaustion [4].

3 Industry Experience: Applicability of Coding Guidelines

3.1 Contextual Information

The first FG study was planned to characterize the applicability of coding guidelines (CG) to a specific software company. The software company, for now on called Alpha, is a distributed company that develops embedded, desktop, web and mobile software solutions for tracking, ambient intelligence and information systems domains. Regarding the software construction, there is a massive reuse of code snippets from embedded software to another one within the company, especially in the tracking domain. As they do not have a well-established reuse process, some rework during reuse activities can be observed due to the conceptual misalignment of developers.

In order to find out the main cause for software development rework and identify how to mitigate it, we carried out a series of experimental studies, such as surveys, systematic literature reviews (SLR) and qualitative data analysis under the action-research methodology. The survey results have shown that most part of rework was due to the misalignment of developers' quality perspectives of source code. Thus, an alternative solution to overcome this issue can be the definition and use of CG. Thereby, we have extracted 24 guidelines from the SLR and the company context (qualitative analysis) related to the readability and understandability of source code [7]. Later, the characterization of these guidelines on their feasibility to support the alignment of source code quality perspectives among developers as well as their contribution to increase the source code quality in the company shall be accomplished.

Therefore, we decided to plan and execute a FG with a fourfold purpose. The first purpose was to characterize the applicability of ten guidelines with lack of evidence regarding their use in Alpha Company. Second, there was a need to start the internalization of the guidelines within the company. Third, we were expecting that the exchange of experience and knowledge among the developers could be of a great help in identifying specific contexts in which the code guidelines could be useful. Fourth, we were interested in gathering as much information as possible to enable us to reformulate the guidelines or create new ones in case of need.

3.2 Planning/Design

This FG has the **objective** of analyzing Coding Guidelines for readability and understandability for the purpose of characterization, with respect to their applicability and parameterization for the Alpha Company's projects, from the point of view of embedded, desktop, web and mobile software developers in the Alpha Company's development context. In order to reach this goal, we identified the following **research questions**: RQ1: *Are the coding guidelines for readability and understandability applicable to the Alpha company development context?* RQ2: *In which way the coding guidelines could be modified or parameterized so they could be applicable to the Alpha Company's development context?*

We selected ten out of 24 guidelines to be characterized as our **object of study**. This selection was based on evidence identified in the technical literature about their

low impact on readability and understandability of source code and on the lack of evidence on the data collected from the developers that can support the indication of their applicability in the company's projects.

For the FG session, we selected as **participants** ten developers that would be available to take part of the session and were engaged in the data collection for both qualitative analysis and guidelines formulation. As **moderators,** we selected two researchers, one master student and one doctor who had some experience in conducting group discussions. Two small **groups** of discussion were created, each one for a specific development domain: embedded systems for tracking domain (with 6 developers) and software web for information systems domain (with 4 developers).

The FG **planning was assessed *ad-hoc*** by one of the researchers conducting the study. Some changes were done specially concerning the **interaction strategy** among participants illustrated in Fig. 2 and described in sequence.

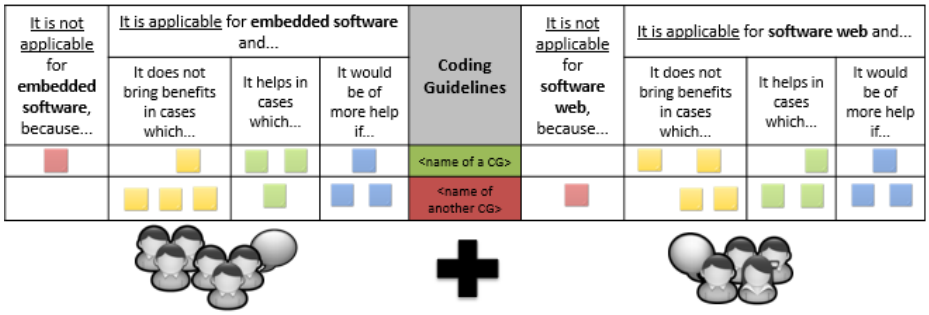

Fig.2. Focus Group board for the first study

The FG was scheduled to run in one session of about 3 hours in which each of the ten guidelines selected ought to be characterized by time. Each group must characterize the guidelines according to their applicability in the group domain context. In order to do this, two boards – hand written in colored paper – must be set up so that participants can place their opinions on appropriate spots.

Each guideline should be placed in the board only during its time of characterization, as described in the following steps: i) the moderator places a guideline in the board and displays its complete description on the screen; ii) the participants discuss about the guideline with their teammates to realize whether the guideline is applicable or not to their development context; iii) the two groups share experiences and knowledge regarding the applicability of the guideline to their context; iv) each participant writes down on a pink post-it why the guideline is not applicable to his development context, on a yellow post-it in which cases the guideline does not bring benefits to his context, on a green post-it in which cases the guideline is applicable to his context, and on a blue post-it in each way the guideline could be modified so it could be better applicable to his context; v) the participants place the post-its in their specified place on the board (see Fig.2). After the participants characterize 3 to 5 guidelines – depending on the ongoing discussions, the moderator should ask the participants to read the opinions raised in the same group so they can expose their agreement or disagreement within the group. The moderator must record all these information.

3.3 Execution and Results

The FG session lasted about two hours and a half. The room used for the session was large with movable chairs enabling the participants to move around, approach each other and look at the boards. Two groups were organized, however only four participants from the embedded systems for tracking domain and two participants from the web information systems domain were available to take part of the FG session.

The FG execution was performed as planned. Each guideline displayed in the board was discussed within each group according to their development context. There were interactions not only within the groups but also between them. The interaction between the groups often happened to clarify doubts regarding the use of some guideline and to indicate issues that had happened in some development projects within the company that could have been overcome if the CG had been used.

Almost all guidelines under characterization were identified as applicable to the Alpha Company's context. Only one guideline received a pink post-it. One of the participants from the software web group said that the CG 1 – "*Use filename extensions according to the file contents*" – was not applicable to the programming language they were used to work with, in that case PHP. The argument used to support this opinion was that PHP code is produced together with HTML and CSS codes, and there was no room for the guideline usage in their context. However, it is important to emphasize that this guideline intends to avoid the grouping of languages and types of contents, since other developers within the company believe this practice is detrimental to code readability and reuse (indication obtained from data analysis).

The colors distribution gave us a quick perception of the guideline applicability to the Alpha Company's context. Besides the only one pink post-it, the FG discussion generated 11 yellows, 40 greens and 22 blues post-its. Regarding the yellow post-its, for instance, some embedded system developers raised issues on using file extensions "inc" for table definition in C programs – which it is meant to be for (CG 1). The reason for not using it was based on the difficulties in finding information in this type of file while using the company's standardized IDE. Other example of guideline modification, now presented on blue post-its, is related to the CG 21 – "*Global constants and variables must be identified as such*". The developers mentioned this guideline could be applicable to other program elements, such as recursive functions, since there are many misuses of them in several available source codes.

3.4 Threats to Validity

The amount of participants that took part of the FG study was at least half than the total amount of the company developers. Most of the developers were not available to participate in the studies at the time they had been conducted. This is a common issue we can face when conducting studies in software companies, there is a real chance the study design has to be modified due to organization issues. Because of this scenario, even within the company we cannot generalize the results.

Another validity threat is related to the distribution of the guidelines to the participants before the FG. We planned not showing the CG until the time for the FG ses-

sion in other to get the first perception and discussion about the guidelines. However, the embedded software developers had access to the guidelines before the session and this might have influenced the interactions among the participants, since they were more likely to agree with each other, as if they had discussed about it before.

As additional threat, as moderators, we were more worried about bringing focus and interaction to the two groups and among them rather than taking notes of everything that was said. We did take notes of important information regarding the applicability of the CG, though, especially those that were not described on the post-its, but there is a chance we missed some information during the session.

4 Academia Experience: Planning Guidelines for Simulation Based Experiments in SE

4.1 Contextual Information

In this experience, we performed the FG as part of a larger observational study. The object regards a set of guidelines aiming at driving researchers through the elaboration of simulation based experiment plans, identifying *a priori* and eventually reducing the threats to validity, also promoting a coherent plan, in which the information is logically organized. Therefore, relevant information (like research context, problem and goals, as well as model specification and validation, experimental design, and output analysis) can be produced for the report. These guidelines are based on findings from a SLR [6] and consolidated technical literature on simulation and SE. Previous versions of these guidelines consisted only in reporting concerns [9]. At that stage, a set of evaluations was performed and indicated certain level of their correctness and completeness. As the guidelines evolved to the planning perspective, they need external evaluation regarding their effectiveness and perceived usefulness.

The evaluation study followed a qualitative approach, in which researchers observe subjects on the elaboration of a simulation experiment plan for an organizational scenario, describing the human resource allocation policies in software projects and the goal of understanding the effects of Brooks' Law on software project schedule.

The study's execution procedure started after two training sessions, one about simulation in the context of SE and another regarding System Dynamics, and it was organized in four sequential stages: 1st) subjects' recruitment and characterization; 2nd) preparation, the subjects receive the instruments (including the organizational scenario, simulation model specification and the executable version for the Vensim PLE tool, and the set of planning guidelines); 3rd) execution, subjects should read the scenario, elaborate the simulation experiment plan (optionally supported by the guidelines), and review the plan elaborated by other subject to identify possible issues, and; 4th) FG session, subjects discuss topics related to the guidelines application.

4.2 Focus Group Planning

The **research goal** for this FG session is to analyze the planning guidelines for simulation experiments for the purpose of characterization, with respect to perceived usefulness (opinions whether the guidelines effectively support the plan elaboration) and ease of use (explicitness, understanding and application), from the point of view of SE graduate (master and doctoral) students. It was conducted in the context of the E Software Engineering course at COPPE-UFRJ. Based on this goal, the derived two **research questions**: RQ1: *Do the planning guidelines for simulation experiments effectively support the elaboration of the study plan?* RQ2: *Are the planning guidelines for simulation experiments clear, ease of understand and use?*

The **participants' selection** for the FG session followed the sample from the study. We designed the FG in two subgroups of four subjects each **organized** based on their characterization (level of instruction, both SE and simulation experience) and their performance on the planning tasks.

From the set of 33 planning guidelines, we selected ten guidelines for discussion due time constraint: the nine less used in the elaborated plans and the most used one was selected as a control for the analysis, concerning the amount of discussion.

This FG **design** organizes discussions in two phases. First, the groups follow a role-play design using "*lovers" and "haters"* roles, in which lovers should argue in favor of the guidelines and the haters against them. The second phase consisted on the identification of improvement opportunities, but playing no role. Additionally, we adopted a technique using a board where subjects post each individual argument. The board is divided into four sections where subjects are encouraged to use post-its of different colors (Fig.3) for their arguments. Two sections for subjects playing the lovers and two more for the haters, each of them concerning perceived usefulness and ease of use. Each row concerns with just one guideline under discussion.

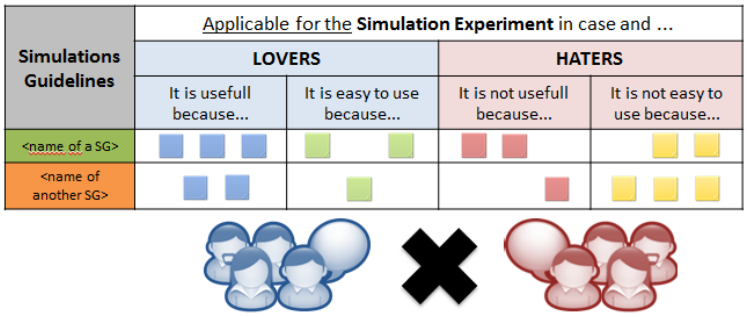

Fig.3. Focus Group board for the second study

This way, the **FG strategy** consisted in presenting to subjects how the FG would work, followed by presentation slides containing the descriptions of each planning guideline under evaluation. Then, the subgroups (lovers and haters) discuss internally and post their arguments in the board, according to their roles. Later, the subgroups were encouraged to read their arguments and discuss them one against the other. After discussing each guideline, subjects used the post-its to present their opinions regarding improvement opportunities w.r.t. perceived usefulness and ease of use.

Three researchers were involved in conducting this FG. One mediator drove the discussions to keep it on focus, encouraging the subjects to post their arguments in favor or against the usefulness or easiness of use of each guideline, according to their roles, and then they exposed their arguments and discussed their point of view. In parallel, two additional researchers were taking notes: one concerned to the discussions and arguments regarding the usefulness and ease of use for each guideline; other was responsible for capturing subjects' behaviors regarding the FG dynamics and the role-play design. Thus, the third researcher captured behaviors like ironic arguments, laughs, change of mindset, consensus reaching, and other behaviors that could reveal how strong their arguments are in favor or not the planning guidelines.

For **data analysis**, we took the answers and discussions in the FGs, according to the notes capturing the multiple perspective (post-its and notes from two observers) to analyze the accomplishment of the two research goals. An additional researcher **assessed** the planning information by reviewing and checking for threats to validity.

4.3 Execution and Results

During the **execution**, the meeting lasted three hours and a half long. The FG meeting took place in a typical classroom, with a data show for presenting slides and fixed seat rows, which is not an ideal setting for the planned dynamics. However, it was possible to accommodate the subjects and make the dynamics work.

From both groups, we observed one subject not engaging the discussions in the beginning. Thus, the moderator intervened, by asking this subject if he could contribute with his experiences, whether he has similar experiences when compared to other subjects' report. This sort of intervention was performed every time any subject was perceived to not contribute with the discussions.

Overall, eight planning guidelines were considered useful and two as out of scope for this study. Six of them were considered useful as checklist; since some types of validation are often overlooked. None was considered ease to use: some of them lack orientation on how to perform specific procedures, others lack of expertise to support their use with domain knowledge. The first five evaluated guidelines motivated many discussions regarding their usefulness and ease of use, explaining scenarios from the simulation experiment and possible alternative ways of applying them.

One example of guideline classified as useful, but difficult to use states, "*Make use of Face Validity procedure (involving domain experts) to assess the plausibility of both conceptual and executable models and simulation outcomes, using proper diagrams and statistical charts as instruments respectively*"**.** Its usefulness appears on the opportunity of an expert possibly recognize what is being simulated and realize reference behaviors for comparison, as well as to identify eventual inconsistencies in model behavior. Among the hindrances, it is possible to highlight the tradeoff between the effort to perform Face Validity and the return w.r.t. the model validity. Besides, the experts' availability to perform it can be difficult to dispose.

The guideline selected as control (regarding the experimental design) was also an attempt to motivate the subjects to join the discussions since it was the most used one. For this reason, it was the first to be discussed. In terms of intensity of discussions,

the previous contact with each guideline does not seem to influence. However, the last guidelines had fewer discussions due to subjects' exhaustion to the long session.

In addition, we identified general contributions to facilitate the guidelines use in the second stage. Two subjects suggested making explicit whether the guidelines apply to the simulation model development or experimentation. Such discussion makes sense because simulation-based studies' lifecycle encompasses both stages. However, the planning guidelines intend to be specific for experimentation and not the whole lifecycle. Besides, we agree that guidelines related to the study definition, like research context, problem, goals and questions are applicable for both situations.

In some cases, they mentioned that a kind of glossary could improve understanding. Furthermore, they claimed for examples on how to apply each guideline, due to the level of abstraction they are discussed. There is explicit mention to V&V procedures and experimental designs for simulation in these cases.

4.4 Threats to Validity

Two subjects did not use/read the planning guidelines before the FG meeting, having lowest participation in the discussions, since they have no previous experience on the application of the guidelines. Hence, they just realize what would be the application.

The existence of a published journal paper influenced the subjects' opinion regarding the model validity. It limited our capacity of observing guidelines related to validity threats, since they assumed these threats are not applicable to the model and study.

From the construct perspective, our main threat is the possibilities of surrogate measures not represent perceived usefulness and ease of use. As the perceived usefulness and the ease of use have a subjective meaning, we adopted the strategy of triangulate the data collected during the FG, using post-its and notes from two observers.

5 Discussion

5.1 Contextual Information: What is FG good for?

FG supports the gathering of cognitive and social aspects regarding an activity execution and it is a worthy alternative to get subjects more committed to the activity. The experiences described in sections 3 and 4 share common motivations for taking FG as a research method. Both researches were interested in a procedure for data collection that could get the subjects more committed on providing feedback to the study than using questionnaires. It does help to explain patterns, trends and unexpected behaviors when analyzing data from the studies. For the first experience, this method worked as a way of spreading them over and exchanging experiences among teams at the organization. For the second, FG provided information that is more detailed, including the reasoning for the subjects judge the guidelines as applicable, useful, or ease of use.

Even showing different contexts (i.e., industry and academia), we considered the FG method effective for our purposes since we were trying to understand how those SE technologies can be applied. The understanding encompasses both cognitive issues, such as people's reasoning on why and how to use the guidelines in each case,

and social issues influencing their use, such as the organizational culture. For instance, the elaboration of both simulation experiment plans and source code requires cognitive effort due to abstractions and decision-making. This reasoning configures an important knowledge, but it is usually omitted after completing the activities.

5.2 Focus Group Planning

Why would I use FG? FG sessions can be used to overcome complex, time-consuming or risky characterization activities presented in other types of research methods. Apart from the motivations and contextual issues, we also identified characteristics inherent of the object of study and instruments that influenced our decision in favor of the FG method as investigation approach. Both experiences had guidelines as objects of study, while domain, complexity and organization are quite different. However, we understand they share common features in terms of application. The sets are large enough to hinder a comprehensive evaluation in detail. Besides, the tasks involved on the application of the whole set or a reasonable subset are time-consuming for *in sito* observation. It explains the two-week period to deliver the simulation plans.

Each simulation guideline tells the reader what to do, presenting reasoning for that matter and examples on how it can be applied. However, there is no documentation on how to use them, allowing several ways of using them. Furthermore, as the guidelines application was an individual experience, it is likely subjects have distinct perceptions regarding their usefulness and ease of use. In the CG experience, the quality focus can be captured qualitatively. Using controlled experiments for this purpose may impose risks on the *in vivo* context.

What do I need for conducting FG? It is necessary to make sure the participants have the appropriate knowledge to take part in a FG session, or the discussion can be compromised. To make FG useful for these two characterization studies, we needed to be sure subjects had minimum knowledge or experience on source coding and simulation experiments planning activities. For that, the experiences had different approaches since the contexts were quite different. Since the CG aim at standardizing source code, the subjects needed to have knowledge or experience on programming and to understand the concepts mentioned in the guidelines. That is, their previous experience in the organization was enough to judge the guidelines as applicable or not. Instead, to judge the simulation guidelines as useful and ease to use, the graduate students needed to experience the elaboration of a simulation experiment plan. This way, FG was performed after the elaboration and review of their simulation plans.

How can I arrange FG? You can benefit from role-based designs that enable to collect information regarding the object of study from the roles' point of view, as well as groups whose participants can freely express their support or disagreements during the discussion without predetermined roles, according to the study context. We presented two different group settings. On both we had to analyze characteristics of the participants and research questions to set up the groups. In the industrial case, we were dealing with different contexts so it would not be reasonable to mix participants from different contexts to have them discussing about the applicability of the CG in their specific context. If we did it so, we could end up with no interaction among

them. Instead, to have different groups participating in a discussion about the same topic in the same room could help both groups raise discussions with knowledge and experiences that could help them criticize better the guidelines under characterization.

The second FG experience had a different setting. All participants were into the same context – a course on experimentation, so we had to use another characteristic to help us set up the groups. The participants had different personalities and level of instruction, some were more talkative and some had more experience in SE. We had to organize a group setting so they could rise with both benefits and harms regarding the use of the planning guidelines for simulation experiments. The lovers and haters strategy could help us achieve this goal. As we would like to obtain different opinions, especially critics regarding the guidelines usefulness and ease of use, we decided to penalize the technology and organize the groups so the participants that can better contribute would be placed to give critical opinions, that is, they would be haters.

Besides the group settings, the questions formulated and the use of post-its also contributed to a proper FG strategy. As we were analyzing many guidelines, we had to state questions that can be directly and easily kept in mind to have the participants internalizing and answering them quickly. The post-its were used not only to capture information, but also to make the participants think harder – possibly with a partner – about questions and the object of study so they could make a reasonable answer.

5.3 Execution

I am in! What about now? During FG execution, it is important to organize and guarantee an appropriate and natural environment to allow group interaction. On both experiences, we used colored papers and handwritten in order to create a casual environment, therefore to support the interaction among participants and moderators. Regarding the environment, in the first experience, we used a large room with space for the subjects to move and post their opinions regarding the guidelines on the wall. There was also a white board in the room that participants had used to describe coding problems that shall be overcome with a CG. The participants where comfortable enough to move around, talk to each other and even write things on the board without being asked to. This helped us collect even more information that we were intending to with the post-it. Although in the second experience there was also a white board available, it was not used; and with respect to the environment, it was not as suitable for FG interaction as the first one. We cannot guarantee these characteristics had huge impact on the interactions, since there are some confound factors that might have affected the interaction between the participants. For instance, the differences between the two objects of study, the participants from the industrial case were familiar with the meeting place, and they knew each other more time than the participants from the academic experience.

Lots of registered information. Do I still need to take notes? A wide source of information can help on the qualitative data triangulation. Regarding data collection, both studies were slightly different from each other. In the first one, the data encompasses the post-its, the data registered by the moderators and the information written on the white board for data analysis – all information related to the research questions.

In the second study, besides the post-its and data registered by the moderators regarding the research questions, it was also collected information regarding the subjects' conduct on the FG, that is, the information about the participants and moderators behaviors. This type of information helped us to properly analyze data in such a way that ironic and forced arguments were used combined with answers to the research questions in order to indicate whether the answers can or not be used during data analysis. This data collection is particularly important for those types of studies on which the subjects are "forced" to assume a position that they might not agree with.

5.4 Threats to Validity

As any research method, FG is vulnerable to validity threats. Just considering its qualitative research inheritance, it is possible to identify many issues related to external, internal and construct validity. Arguably, the most mentioned issue in such studies is associated to generalizability, since they are usually inductive from specific subjective contextualized data along with the use of purposive population sampling instead of random sampling. Both studies demonstrated concerns with this regard. The point here is to try to provide explanatory descriptions that are detailed enough to identify comparable situations in which they can be applied. It is possible to say that the study which has a research technique as object of study is favored in this respect as academic experience are more broadly well-defined and understood. However, in the case of software technologies, especially within an industry environment, it is much less straightforward to identify if the situations are similar enough. To that end, it is important to try to distinguish organizational and personal (i.e., worldview) aspects from the more technical and reasoning issues associated with software development.

In qualitative research, construct and internal validity are usually associated to the degree that findings are considered to be interpreted correctly. In FG studies, this can manifest in various ways. With discussions and activities as the only surrogates to 'measure' (i.e., interpret) the variables of interest, it is important to triangulate the data collected during FG session. To mitigate this threat in evaluating software technologies with FG, the focus should not be oriented to the technologies' characteristics themselves (e.g., architecture or usability), but rather to how they are affected by the individuals and teams. This is the case of both studies when evaluating the use of guidelines. Conversely, internal validity concerns with how experiences were elicited from discussions and if they really represent the facts. In the two FG studies, this aspect was present in the possible participants' avoidance in criticize the guidelines in the presence of moderators which are also the guidelines creators. This could have inhibited spontaneous discussions, but it was tried to be circumvented with FG activities and design – as it was the case of lovers and haters in the second study.

6 Conclusions

This paper described two experiences on using FG techniques for characterization purposes regarding SE technologies. The experiences shared some commonalities as

the research method and characteristics of the objects of study, and in terms of design, they have many differences. However, we can still compare the studies and observe that, besides the adopted strategy, both studies contributed for the characterization of these technologies in their distinct contexts. We consider the results as positive, also indicating that FG is useful as an alternative tool to support data collection in SE research. Particularly, we observed some advantages on exploring strategies to stimulate participants on performing the activities involved on the FG dynamics. The most significant ones are their commitment to the study and the detailed data regarding the participants' perceptions about how the technologies under evaluation can be applied.

Most part of the successful results we attribute to the systematic approach adopted to conduct the FG, with planning activities and the concerns with design aspects aligned to the object of study, the goals and the participants. We are aware of potential threats to the studies validity, including some we could not anticipate. However, we understand that most of these threats are applicable for other settings, also involving qualitative methods. We hope all of these discussions and lessons can be useful for researchers and practitioners interested on using FG to support their observations.

Acknowledgements

Authors thank CNPq (grants 141152/2010-9 and 304795/2010-0) and CAPES for supporting this research.

References

1. Sim, J.: Collecting and analysing qualitative data: issues raised by the focus group. Journal of Advanced Nursing 2(28), 345-352 (1998)
2. Colucci, E.: "Focus Group Can Be Fun": The Use of Activity-Oriented Questions in Focus Group Discussions. Qualitative Health Research 17(10), 1422-1433 (2007)
3. Kidd, P., Parshall, M.: Getting the Focus and the Group: Enhancing Analytical Rigor in Focus Group Research. Qualitative Health Research 10(3), 293-308 (2000)
4. Kontio, J., Bragge, J., Lehtola, L.: The Focus Group Method as an Empirical Tool in Software Engineering. In : Guide to Advanced Empirical Software Engineering. Springer-Verlag, London (2008) 93-116
5. Mian, P., Travassos, G., Rocha, A., Natali, A.: Towards a Computerized Infrastructure for Managing Experimental. In : JIISIC, Madrid, vol. 2, pp.475-487 (2004)
6. Davis, F.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly 13(3), 319-340 (1989)
7. Ribeiro, T.: Alinhando Perspectivas de Qualidade em Código Fonte a partir de Estudos Experimentais – Um Caso na Indústria. Master Dissertation, COPPE/UFRJ (2014)
8. de França, B., Travassos, G.: Are We Prepared for Simulation Based Studies in Software Engineering Yet? CLEI Electronic Journal 16(1), Paper 8 (2013)
9. de França, B., Travassos, G.: Reporting guidelines for simulation-based studies in software engineering. In : Proc. of the 16th EASE, Ciudad Real, Spain, pp.156-160 (2012)