Overcoming Open Source Project Entry Barriers with a Portal for Newcomers

Igor Steinmacher
Department of Computing
Federal University of Technology – Paraná (UTFPR) – Brazil
igorfs@utfpr.edu.br

Tayana Uchoa Conte
Institute of Computing
Federal University of Amazonas (UFAM) – Brazil
tayana@icomp.ufam.br

Christoph Treude, Marco Aurélio Gerosa
Institute of Mathematics and Statistics
University of Sao Paulo (USP) – Brazil
{ctreude, gerosa}@ime.usp.br

ABSTRACT

Community-based Open Source Software (OSS) projects are usually self-organized and dynamic, receiving contributions from distributed volunteers. Newcomer are important to the survival, long-term success, and continuity of these communities. However, newcomers face many barriers when making their first contribution to an OSS project, leading in many cases to dropouts. Therefore, a major challenge for OSS projects is to provide ways to support newcomers during their first contribution. In this paper, we propose and evaluate FLOSScoach, a portal created to support newcomers to OSS projects. FLOSScoach was designed based on a conceptual model of barriers created in our previous work. To evaluate the portal, we conducted a study with 65 students, relying on qualitative data from diaries, self-efficacy questionnaires, and the Technology Acceptance Model. The results indicate that FLOSScoach played an important role in guiding newcomers and in lowering barriers related to the orientation and contribution process, whereas it was not effective in lowering technical barriers. We also found that FLOSScoach is useful, easy to use, and increased newcomers' confidence to contribute. Our results can help project maintainers on deciding the points that need more attention in order to help OSS project newcomers overcome entry barriers.

CCS Concepts

- **Information systems → Open source software**
- **Software and its engineering → Collaboration in software development**

Keywords

Newcomers; Newbies; Novices; Beginners; Open Source Software; Barriers; Obstacles; Onboarding; Joining Process

1 INTRODUCTION

Open Source Software (OSS) projects have risen to great prominence within the last several years [24]. Many OSS projects leverage contributions from geographically distributed volunteers and newcomers are important for their survival, long-term success, and continuity. According to Qureshi and Fang [33], it is essential to motivate, engage, and retain new developers in a project in order to promote a sustainable number of developers. Furthermore, newcomers are a source of innovation for new ideas and work procedures that the group needs [23].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICSE '16, May 14-22, 2016, Austin, TX, USA
© 2016 ACM. ISBN 978-1-4503-3900-1/16/05…$15.00
DOI: http://dx.doi.org/10.1145/2884781.2884806

However, newcomers face many difficulties when making their first contribution to a project. OSS project newcomers are usually expected to learn about the project on their own [37]. Dagenais et al. [10] compare them to explorers in a hostile environment who need to orient themselves. Thus, a major challenge for OSS projects is providing newcomer support.

Previous research related to the joining process of newcomers examined the dynamics that drive OSS contributors, mostly focusing on the motivations for becoming a member, roadmaps to becoming a core developer, or indicators of potential long-term commitment [16, 20, 38, 51, 52]. An understudied aspect of the OSS joining process is what happens during the period after a newcomer decides to participate and before their first code contribution is accepted and included in the shared repository. This period is particularly relevant to OSS projects, as many newcomers do not want to remain on the project, but only post a single contribution (e.g., a bug correction or a new feature) [32]. During this period, newcomers face barriers that can result in their decision to give up contributing. Thus, as Karl Fogel [14] states, *"if a project doesn't make a good first impression, newcomers may wait a long time before giving it a second chance."*

In our previous work [41, 43], we proposed a preliminary model to help identify and better understand the barriers faced by newcomers to OSS projects. The 58 barriers presented in the model are organized into six categories: cultural differences, newcomers' characteristics, reception issues, newcomers' orientation, technical hurdles, and documentation problems.

Our goal in this paper is to propose and evaluate FLOSScoach, a web portal created to support the first contributions of newcomers to OSS projects. FLOSScoach has been designed and structured to reflect the categories identified in the preliminary barriers model. Each category was mapped onto a portal section, which contains information and strategies aimed at supporting newcomers in overcoming the identified barriers. To populate the portal, we collected already-existing strategies and manually fed the portal sections. The structure and a screenshot of FLOSScoach are presented in Figure 1.

To guide our research, we defined the following three research questions:

Q1. How do newcomers use the web portal to overcome the contribution barriers?

Q2. Does the use of the web portal impact newcomers' self-efficacy?

Q3. What is this web portal's perceived usefulness, ease of use, and likely future use?

We conducted a study with students, relying on qualitative analysis of diaries, a self-efficacy questionnaire [3], and the Technology Acceptance Model (TAM) [12]. The results enabled us to evaluate FLOSScoach and improve the barriers model according to the feedback received.

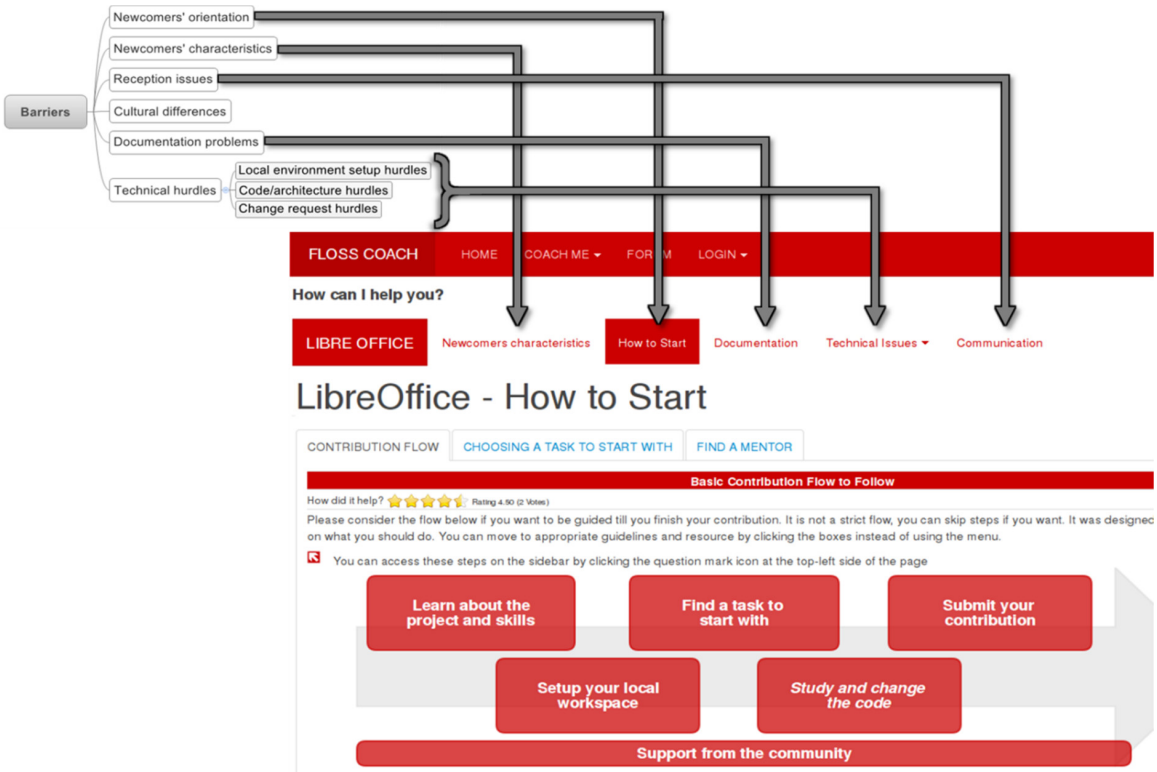

Figure 1. Barrier categories mapped to FLOSScoach ("How to Start" section)

The contributions of this paper include: (i) empirical validation of the model of barriers for newcomers to OSS by means of a web portal built according to the categories proposed in the model; (ii) refinement of the model according to the experience obtained building the web portal and the feedback received; (iii) empirical evidence for an initial set of features that help newcomers overcome project entry barriers; and (iv) an empirical evaluation of the newcomer web portal from three perspectives. We hope that OSS communities and researchers will take advantage of this work to improve their support for newcomers, ultimately leading to more contributions to OSS projects.

2 RELATED WORK

The onboarding of newcomers is not an issue exclusively faced by OSS. Many studies in the literature deal with the joining process of newcomers in collective production communities, including Wikipedia [15, 49] and OSS projects [5, 19, 25, 27, 50]. Dagenais et al. [10] and Begel and Simon [4] presented studies on newcomers' joining process focusing on closed source industry settings. Many other studies focus on newcomers onboarding to OSS projects [44].

Some studies report scripts and cases of developers successfully joining OSS projects, presenting this data as a joining process. Von Krogh et al. [25], for example, proposed a joining script for developers who want to take part in a project based on historical analysis. Nakakoji et al. [28] studied OSS projects and presented eight possible roles for the community members and structured them into a model composed of concentric layers. Although these papers deal with the evolution of members' participation in OSS communities, they focus on newcomers after the onboarding.

Other researchers tried to understand the barriers that influence the retention of newcomers. Zhou and Mockus [52] identified the newcomers who are more likely to remain in a project in order to offer active support for them to become long-term contributors. Jensen et al. [19] analyzed whether emails sent by newcomers are quickly answered, if gender and nationality influence the kind of answer received, and if the reception of newcomers is different in user lists and developer lists. Similarly, in our previous work [45], we studied how the answers to the first emails of newcomers influenced their retention.

There are also some papers presenting tools to support newcomers' first steps. Čubranić et al. [8] presented Hipikat, a tool that supports newcomers by building a group memory and recommending source code, mail messages, and bug reports to support newcomers. Wang and Sarma [50] presented a Tesseract extension to enable newcomers to identify bugs of interest and resources related to those bugs, and to explore the appropriate socio-technical dependencies for a bug in an interactive manner. Park and Jensen [31] showed that visualization tools support the first steps of newcomers to OSS projects, helping them to find information more quickly. Mentoring was also explored as a way to support newcomers. Malheiros et al. [27] and Canfora et al. [5] proposed different approaches to identify and recommend mentors to OSS newcomers.

3 FLOSScoach: A PORTAL TO SUPPORT NEWCOMERS TO OSS PROJECTS

We developed the FLOSScoach portal based on the barriers model that we proposed previously [41]. The model comprises 58 barriers organized into six categories presented in Figure 2.

In Figure 1 we show how the barriers model was mapped onto the portal sections. The mapping is also presented in the following items:

- **Newcomers' orientation**. Newcomers often face rugged and unfamiliar landscapes when onboarding to an OSS project. They need proper orientation to find their way and correctly make their contributions. Examples of barriers under this category include difficulty to find a mentor and poor "How to contribute" documentation available. Figure 1 depicts the FLOSScoach

section, which aims at addressing this category of barriers. The category was mapped to the "How to start" section, which contains a clickable step-by-step flow, a list of easy tasks (filtered tasks based on tags provided by the project), and an indication of mentors listed at openhatch.com.

- **Newcomers' characteristics**. This category comprises the barriers related to the experience and behavior of the newcomers and the way they interact when joining a project. This category was mapped to a section called "Newcomer characteristics," which presents a lists of skills needed and expected to contribute to the project, links to tutorials provided by the community, and widgets (from openhub.net) displaying facts about the project.
- **Reception Issues.** This category comprises the barriers related to the interaction that occurs between newcomers and the community. A breakdown during these social interactions can lead to demotivation, and even result in newcomers dropping out. These barriers include not receiving an answer to a message and receiving impolite answers. Solutions to address reception issues were mapped to a section called "Communication." This section includes links to the mailing list, an embedded tool to connect to the project's IRC channel (if available), and a set of guidelines on how to interact with the community, including a suggested message template to guide the newcomers' first interactions.

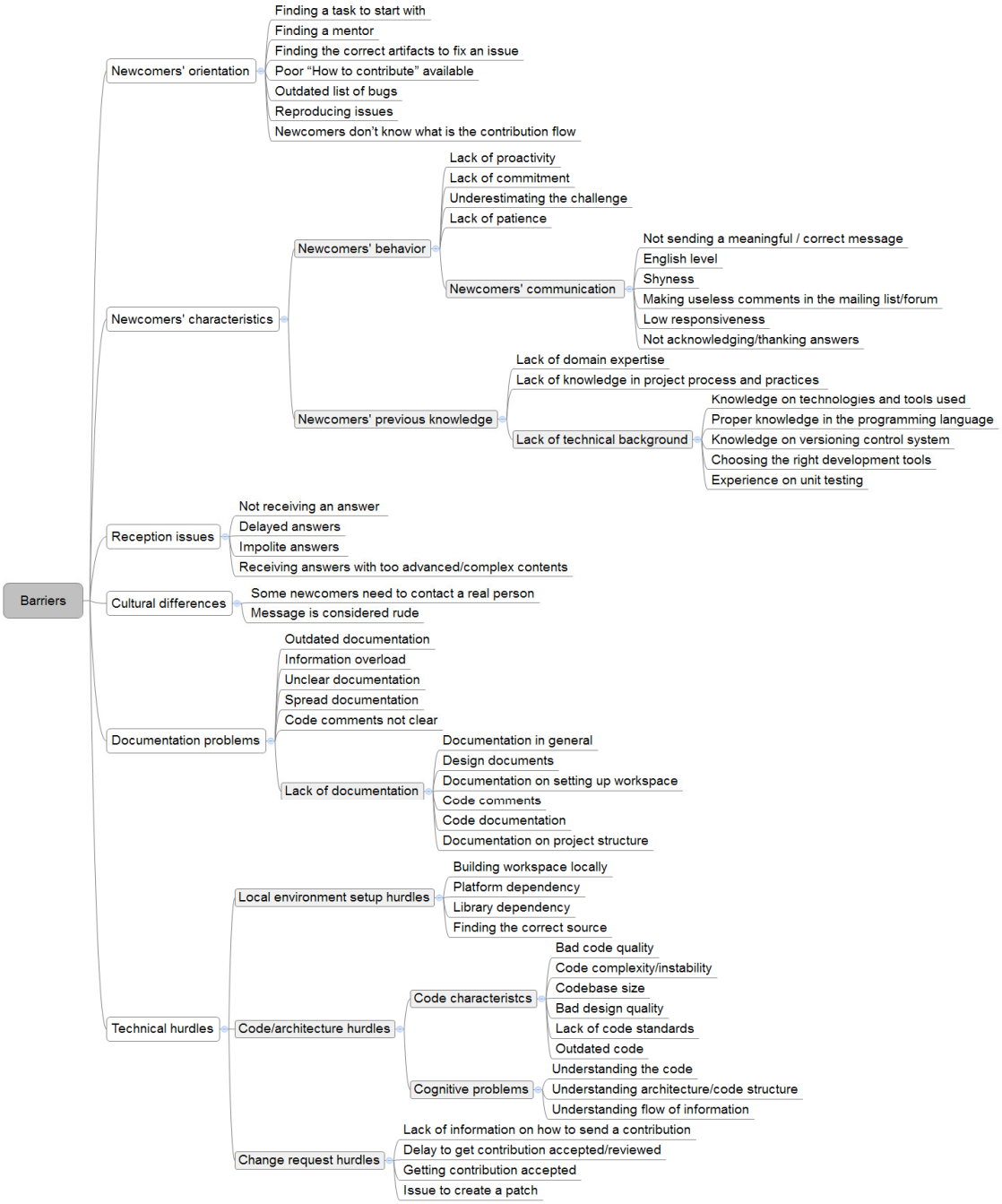

Figure 2. Barriers model proposed in [41, 42]**.**

- **Cultural differences**. Cultural differences can result in interaction problems. Two barriers were reported in the barriers model: the need to be in contact with a real person and messages considered to be rude. While building FLOSScoach, we found that the barriers under this category were strictly related to reception issues. Thus, strategies such as the suggested message template and links to regional mailing lists (when available), which aim to help address the cultural differences, were put under the section "Communication."
- **Documentation problems**. This category refers to the need to learn the project's technical and social aspects in order to be able to contribute. The barriers under this category define documentation problems that have been identified as barriers to newcomers, and include: outdated documentation; unclear code comments; information overload; and lack of documentation. We created a FLOSScoach section called "Documentation" and filled it with information about the documentation made available by the project, organizing it in subsections.
- **Technical hurdles.** This category consists of the project's technical barriers that arise when newcomers are handling the code. These barriers were placed under a single category and split into three subcategories: code/architectural hurdles, change request hurdles, and local environment setup hurdles. We mapped the information related to addressing the technical hurdles to a FLOSScoach section called "Technical issues." The section includes an embedded mailing list archive search and links to tutorials, code search engines, code standards, and code related documentation.

4 RESEARCH SETTING

In this section, we present details about the research setting, including the participant selection and the assignment that participants received. The study was conducted in two different iterations, and we present the details individually. Our participants were undergraduate students. Students are potential OSS project contributors, which is why there are currently several programs (e.g., Google Summer of Code, Facebook Open Academy) focusing on attracting them. The students chosen for this study had enough knowledge to fix small bugs in software projects and were motivated to contribute (since their grade depended on it). They joined a real project with real issues, and they interacted with the actual code and community.

For both iterations, we first profiled the participants to verify their experience level regarding software development, OSS, and programming languages. We also captured gender and age.

All the participants were asked to contribute code to a given project, by fixing bugs or implementing new features according to what was already reported in the project's issues tracker. They had one month to deliver the task. It was part of their assignment to write diaries logging their activities, issues, and everything that they did while working on the assignment in a shared document.

Iteration 1. Fourteen students attending a Software Engineering course (3rd year) at Federal University of Technology – Paraná (UTFPR) received a graded assignment. Most participants did not have any experience in software industry (outside university). Only five of them had worked in industry, but for less than one year. Only one participant reported previous experience in contributing to OSS projects. We directed the participants to two specific OSS projects: LibreOffice and JabRef. These two projects were part of our previous studies, with a large number of participants and communities receptive to our research. Moreover, this study's first author is a JabRef contributor.

We split the participants and directed them to the two different projects. In order to randomly split the students into portal (received credentials to FLOSScoach) and control (did not receive credentials to FLOSScoach), we:

1. matched students and projects according to the expertise in the programming languages;
2. ranked the students according to their previous experience in OSS and industry (in case of ties, we ranked them randomly);
3. followed the ranked list (considering the project matching from step 1) to alternate between control and portal group.

Iteration 2. Iteration 2 was conducted during a Software Engineering course of a Computer Science major at University of São Paulo (USP). The initial number of participants was 51. Their profile was a little different from the participants of Iteration 1. The main difference is that the participants had slightly more industry experience.

As we were counting on a larger number of newcomers, we prepared FLOSScoach with information from four other OSS projects: Amarok, Empathy, Vim, and Audacity. We chose projects written in C/C++ to match the participants' language skills.

Following the same steps considered in Iteration 1, we split the participants by project, and into portal and control groups.

We would like to highlight that similar assignments had been applied to evaluate the students in previous editions of the courses in which the study was conducted. Thus, the task is part of the dynamics of the course and will continue to exist in following editions. Moreover, the grades were not related to the contribution itself, but based on the student diaries. We analyzed the process they followed, the frequency of entries, answers to feedback, persistence, and level of detail. We emphasized this to all the students before the assignment (as it had been done in the previous editions of the course). The assignment and writing diaries was mandatory, but completing the questionnaires was not. In the welcome page of the survey, there was a written notice stating that the survey was part of research, was not mandatory, and would be used in a study. The fourth author was in charge of the course in Iteration 2. No such conflict existed in Iteration 1.

5 PORTAL EVALUATION

In this section, we report the method used to answer Q1, Q2, and Q3 and the respective results.

Based on the research questions (see Introduction), we defined a research method in which we administered pre- and post-questionnaires and requested the participants to write diaries. For the sake of readability, we present a brief description of how each question was answered below before discussing the details for each question in the corresponding subsection.

To answer question Q1, we conducted a qualitative analysis using data obtained via diaries from newcomers trying to place their first OSS project contribution.

To answer question Q2, we administered pre- and post-study self-efficacy questionnaires. Self-efficacy is a measure of the confidence in the participants' perceived ability to perform a task, which can impact one's actual ability to complete a task [3]. It is correlated with the willingness to stick with a learning task, and has been studied in computer science education [7] and OSS [11] contexts.

Finally, to answer question Q3, we applied the Technology Acceptance Model (TAM) by conducting another questionnaire exclusively with the participants who used the FLOSScoach web portal. TAM is a model, which aims at assessing user perception about a technology's usefulness and ease of use, which are the determinants of a user's technology acceptance behavior [26].

As explained before, this study was conducted in two iterations. The first iteration was conducted with a small set of students and projects,

and the second iteration was conducted with a larger set of students and projects. To facilitate reporting our results, in this section, we identify the participants as follows:

- **C1-XX**: participant from Control group, in Iteration 1;
- **C2-XX**: participant from Control group, in Iteration 2;
- **P1-XX**: participant from Portal group, in Iteration 1;
- **P2-XX**: participant from Portal group, in Iteration 2.

Before detailing the findings, we would like to highlight that from the initial 65 participants we considered only 46 for analysis. We dismissed subjects if: fewer than three diary entries were written; the post-study questionnaire was not filled; or the contribution was not a code contribution. Thus, from the 34 participants in the control group (not using FLOSScoach), 19 were considered (56%) in the analysis. From the 31 participants assigned to use FLOSScoach, 24 were considered (77%). This may indicate that FLOSScoach fostered or facilitated the assignment completion.

It is important to note that three participants originally assigned to the portal group informed us that they did not use FLOSScoach. Therefore, we excluded them from the analysis.

5.1 How do newcomers use the web portal to overcome the contribution barriers?

In this subsection we present the details about the method followed to answer Q1 and the findings related to this question.

5.1.1 Method

To answer this research question, we employed a diary study methodology. A diary study allows for access to everyday behavior in a relatively unobtrusive manner, which affords access to the experience's immediacy and provides accounts of phenomena over time [47]. Diary studies have been used to collect longitudinal data in many disciplines, including psychology, health and medicine, education, anthropology, and architecture [11]. There have also been diary studies in various settings within the Computer Sciences [1, 4, 9, 11, 17, 21, 29]. For instance, Begel and Simon [4] collected data using 'daily events from novice Microsoft developers' video diaries.

We chose to conduct a diary study because, as reported by Davidson et al. [11], we could not observe our participants. They were able to work whenever they desired, conducting the work at a chosen time and place. Therefore, it was not possible to be with them every time they were working on the assignment. Since our goal was to explore how FLOSScoach supported newcomers throughout the contribution process, our diary study was unstructured, which enabled participants to write anything they wanted about their journey.

In diary studies, regular interaction between the investigator and participants is important because it helps the participants understand the diary entries' importance and researchers' desired level of detail [34]. Therefore, we used shared documents to keep the diaries. Each participant created his/her own document, which they shared with the researchers. We constantly followed the entries and provided prompt feedback to the participants via annotated comments. We guided them, asking for more details about what they posted, asking, for example, how they achieved something, why they made a given decision, or where they found a piece of information.

To analyze the data, we used open coding and axial coding procedures from Grounded Theory [46]. The analysis was conducted by one researcher, and reviewed by a second researcher. The review was followed by a discussion phase involving both researchers.

5.1.2 Results

We analyzed the diaries both from newcomers who used FLOSScoach and newcomers who did not use the web portal. In the following subsections, we report the findings related to question Q1 in detail, aiming to identify the web portal's influence on the newcomers' contribution process. We categorize the findings across three axes: process, social, and technical.

It is important to note that, since all the diaries and feedback sessions were conducted in the students' native language (Brazilian Portuguese), the quotes presented to ground our findings are free translations from the excerpts to English.

Process: the portal played an important role

The initial feeling of the participants who did not have access to FLOSScoach was uncertainty and doubt on how to start (evidenced in 12 out of 19 diaries). They received an assignment, and the big question came to mind: what to do now?

"... I will need to contribute to LibreOffice but I don't have any clue on how to do it" – C1-06 (first sentence of the diary)

"... I am a little lost, so I will try a bug that I think I can work with..." – C1-02 (in his first entry)

"When I accessed the project page for the first time, it confused me" – C2-07

"After lurking the mailing list archives I am still without clues on how to contribute…" – C2-05

This issue relates to at least two other barriers in our model: **spread documentation** (reported by 4 out of 19 participants from the control group) and **information overload** (reported by 3 participants). These problems can lead newcomers to feel lost when trying to better understand the project, as shown in this quote:

"I am feeling the necessity of finding something that will be my guide during this process, because until this moment I had to search information in different places" – C1-04

In one case, a student took a completely wrong path, analyzing a different codebase by cloning the wrong repository and trying to set it up. He spent around one third of his available time on this wrong path, but was put back on track when a colleague pointed out his mistake.

The guidance provided by FLOSScoach avoided this kind of situation. At the very beginning, the newcomers who used the web portal felt more comfortable and were aware of the expected steps, for example:

"The tool seems to be good, because it solves doubts that range from the skills needed to start to pointing out how to submit a contribution." – P1-07

"I could check what newcomers need to know regarding the development environment, accessing the links to documentation and relevant guidelines, understanding how to search for help and who to talk to in case of problems and, mainly, accessing the newcomer guide showing the flow and offering support to each step of the process." – P1-04

"…the tool helped me a lot, because it gave me an outstanding guidance about what I needed to do and, consequently, made me spend less time and made me more confident" – P1-05

"For me, the task was facilitated mainly by two factors: 1 - Presentation of necessary information only; 2 - organization of information" – P2-01

No participant who used the tool in the contribution process reported barriers related to **newcomers' orientation**, what are the next steps, or feeling lost.

Something that deserves highlight in the newcomer's first contact with the project is that FLOSScoach contains information on which **newcomer characteristics** the project requires, including **technical**

background and **newcomers' behavior**. Anticipating what specific technologies they need to know in order to contribute sets the newcomers' expectations, informing them what they should learn to achieve their goals. In our study, this was evidenced by four participants using FLOSScoach, for example:

"checking the skills that newcomers needed to know about the development, either technical – like language, versioning control system and preferable operating system – and behavior – respect, commitment and proactivity… I don't know the code review tool, Gerrit, so I will need to know more about it…" – P1-04

Making the newcomers aware of the technologies/languages used was well received, and created the understanding of what they need to learn to complete their contribution. It is important to note that in our study, we tried to remove the programming language barrier by directing newcomers to projects they could handle.

Participants also mentioned that FLOSScoach facilitated them **finding a task to start with** (as reported by 8 out of 24 participants of the portal group). The difference was mostly evident in analyzing JabRef participants because some JabRef bugs were tagged as easy, but there was no information about it on JabRef's website. This was stated by a participant attempting to contribute to JabRef:

"It was a little hard to find a task that I understood. I started from the tool [FLOSScoach], then I tried to find using Bugzilla by using the link provided at FLOSScoach. After some time, I went back to FLOSScoach and chose one from that list." – P1-07

In general, newcomers felt that FLOSScoach provided them with orientation. Participants mentioned benefits such as links that point to the correct information (mentioned by 4 out of 24 participants), suggestions for finding a bug to start with (mentioned by 8 out 24), and the suggested contribution flow (reported by 10 out of 24).

Social Interactions: the portal helped in some cases

Fifteen (out of 43) participants interacted (or attempted to interact) with the community during our study. It was common to find mentions to IRC chats, mailing list messages, and issue tracker entries in the diaries. Seven participants of the control group and 8 from the portal group interacted with the community in different ways.

Despite a high response rate from the community, in three cases the newcomers did not receive any answer from the community, which lead to frustration.

"I asked this question after searching for a certain period. I posed the question politely, as required by the community. I received no answer..." – P2-09

An important finding is that those newcomers who received responses reported no cases of receiving improper answers. All the participants who communicated with the community mentioned that they received welcoming messages and proper orientation.

Regarding the use of the web portal, one participant reported that the content of FLOSScoach was complete enough, and talking to the community was not necessary in his case:

"I did not need to talk to them. The tool was very clear. It is very easy, [the portal] is very good." – P1-02

Another participant mentioned a benefit of using FLOSScoach, namely that the message template provided was helpful:

"I liked the message template, showing how to introduce myself and to present the problems I am facing. Even having proficiency in English, I did not know the more polite way of asking for help. This example helped to be clear, concise to present the message objective, and also to reduce the shyness" – P2-01

Even with some indications that FLOSScoach supported social interactions, finding better ways to foster the communication with the community is a topic that deserves further investigation.

Technical: barriers are still there, in both cases

The diary analysis and debrief sessions revealed that web portal use did not overcome technical problems, such as workspace setup issues. Issues with workspace setup and difficulties finding their way in the code were widely reported by both the control and portal groups. This result aligns with the self-efficacy questionnaire results, which are detailed in Section 5.2. Furthermore, difficulty **understanding the architecture/code structure**, **understanding the code**, problems **finding the correct artifact to fix an issue**, and many barriers inside the **local environment setup hurdles** category were also evident and prominently represented in the diaries (in 32 out of 46). These barriers could not be diminished because FLOSScoach did not explicitly provide any new tool or mechanism to support newcomers in terms of technical issues.

Since the reports were written in parallel with the activity, and since most participants externalized their feelings about the problems, we could find evidence about the extent to which the barriers influence newcomers' motivation. We found some students relating these issues to frustration, irritation, or demotivation. We found that these feelings appeared in time-consuming activities with an unhappy end. In some cases, the bad feeling was reported when too much time was spent on fixing a single problem, mainly while also facing issues with missing dependencies:

"The issues with the dependency are still there, so I decided to clone the repository again. I am tired and frustrated" – C1-04

The other frequently reported barrier, which accompanied some bad feeling, was difficulty in finding the artifacts that should be amended (evidenced in 21 diaries). Participants used different strategies to approach this problem (facilities offered by FLOSScoach and by the project itself, or using previous knowledge), but the difficulty prompted them to mention irritation and demotivation.

"I checked that bug again... The complicated thing is to find where to find the code I need to work on, because there are too many files and lines of code" – C1-05

"I think I will have to take a look at all the documentation. I have no idea of where the code that I need to change is." – P1-07

Technical issues were the most influential barriers for the studied newcomers, and they are the main reason why most students were unable to deliver their contributions in the defined period. The lessons that can be learned from this analysis are: (i) communities should give special attention to workspace setup issues and provide proper indications to enable newcomers to find the artifacts that need to be changed; and (ii) novices to OSS development need to start with smaller and newer projects, which are easier to set up and provide code that is easier to understand.

To conclude, we so far could not find any strong evidence that FLOSScoach can offer technical problem support as this quote indicates:*"FLOSScoach is really interesting… It was a good starting point, that helped me learn the etiquette and the process, but it did not help me with the technical development problems" – P2-17*

5.2 Does the use of the web portal impact newcomers' self-efficacy?

Self-efficacy is a measure of the confidence in the participants' perceived ability to perform a task, which can impact one's actual ability to complete a task [3]. In this subsection, we present the details about the method followed to answer Q2, and the findings related to this question.

5.2.1 Method

Based on previous work that applied self-efficacy to OSS research [11, 30], we prepared a questionnaire with 10 items related to self-efficacy of contributing to OSS on a five-point Likert scale, as presented in Table 1. The items of the questionnaire were based on previous work [11, 30].

Table 1. Items on self-efficacy toward OSS activities.

Sentence
Q1. I feel comfortable asking for help from the community using electronic communication means
Q2. I can write my questions and understand answers in English
Q3. I am good in understanding code written by other people
Q4. I have pretty good skills to write and change code
Q5. I feel comfortable with the process of contributing to an Open Source project
Q6. I think that contributing to an open source software project is an interesting activity
Q7. I feel I can set up and run an application if a set of instructions is properly given
Q8. I am pretty good at searching for solutions and understanding technical issues by myself
Q9. I can choose an adequate task to fix if a list of tasks is given
Q10. I can find the piece of code that needs to be fixed given a bug report presenting the issue

We asked participants to answer the questionnaire immediately before they started their assignment and immediately after concluding it. We aimed to discover whether participants had success performing the tasks (resulting in an increase in self-efficacy) or faced unexpected problems or failures (resulting in a decrease in self-efficacy) [39].

5.2.2 Results

To verify how the use of FLOSScoach influenced participants' self-efficacy, we analyzed the variation of the pre- and post-study answers. By applying a Wilcoxon signed-rank test, we found that the difference in the total self-efficacy score significantly decreased for the participants who did not use the tool ($p=0.005$), while there was no significant difference in the portal group. Moreover, we found that the total self-efficacy after the assignment was greater for the group that used FLOSScoach than for the control group ($p=0.013$), while before the assignment there was no significant difference.

In Iteration 1, the self-efficacy of five out of six participants who used FLOSScoach increased, i.e., most of the participants finished the study more confident than when they began. On the other hand, most of the participants who did not have access to FLOSScoach (5 out of 7) presented a decrease in self-efficacy.

A similar behavior was observed in Iteration 2, in which the self-efficacy of 11 out of 18 participants who used the web portal increased, two remained the same, and five decreased. For the participants who had no access to FLOSScoach, we observed that the self-efficacy score of 8 out of 12 participants decreased after the study, whereas three increased and 1 remained the same.

In accordance with the results presented by Davidson et al. [11] in their study of older adults in OSS, we attribute the decreases in self-efficacy to the idea that "you don't know what you don't know." If participants find unexpected barriers, their self-efficacy decreases.

Figure 3 shows the median of the answers per question, considering all participants from Iterations 1 and 2. We can observe trends that increased self-efficacy, and others that pulled it down. First, in consonance with the diary entries, the scores for the question related to the OSS contribution process (Q5) increased for the participants who used FLOSScoach, and decreased for those who did not. The decreasing trend observed for Q10 (finding the right piece of code) represents the newcomers' difficulty in finding the artifact they need to change in order to work on a selected task, reported by newcomers in the control group. We did not find any explanation for the variations in Q1 (asking for help) and Q2 (writing/understanding English), since there was no issue related to social interactions reported during the diary study.

It is worth noting that the self-efficacy of 17 out of 24 participants who used FLOSScoach increased, while 6 decreased. Most of the participants finished more confident than when they started the study. None of the participants with a decreasing behavior managed to proceed further than finding a bug to work with. The decreases were mainly related to decreases in social interactions (Q1 and Q2) and code issues (Q3 and Q10).

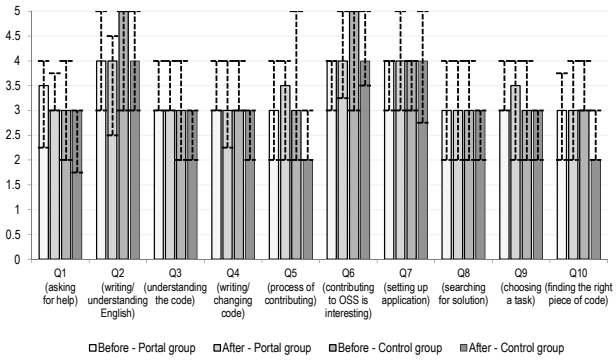

Figure 3. Self-efficacy results per question.

For the participants who had no access to FLOSScoach (control group), we observed the opposite behavior. The self-efficacy score of 12 participants decreased after the study, whereas only six increased. Among the six participants with an increased self-efficacy, two finished their assignment with their contribution accepted by the community and two fixed a bug but did not submit the code by the time the study was finished.

5.3 What is this portal's perceived usefulness, ease of use, and likely future use?

In this subsection we present the details about the method followed to answer Q3 and the findings related to this question.

5.3.1 Method

To answer this question, we applied the Technology Acceptance Model (TAM) [12], which is a model inspired by the theory of reasoned action (TRA) [13]. TRA asserted that attitude towards an action and subjective norms together impact behavioral intention, which in turn affects how people perform the action [22]. TAM aims at assessing user perception about a technology's usefulness and ease of use, i.e., the basic determinants of a user's technology acceptance behavior [26]. Perceived usefulness is defined as "the degree to which a person believes that using a particular technology would enhance his or her job performance," and perceived ease of use refers to "the degree to which a person believes that using a particular system would be free of effort" [12]. According to TRA, these factors positively correlate with one's intention to actually use a technology, i.e., self-reported future use, as modeled in Figure 4.

TAM has been widely applied in technology assessment, producing reliable results when users have worked with the technology for some time. King and He [22] report the results of a meta-analysis of 88 published TAM studies supporting the instrument's validity and robustness with a wide range of applications.

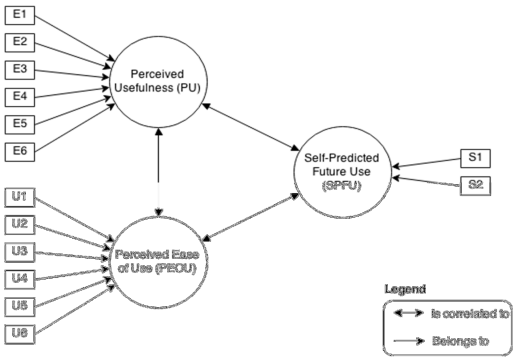

Figure 4. Model of usefulness, ease of use, and self-predicted future usage (TAM).

We found TAM to be a suitable starting point for developing an adapted measurement tool that could assess FLOSScoach's "ease of use" and "usefulness" in supporting the contribution process of newcomers. Table 2 shows our TAM-based assessment model adapted from [2, 12, 26, 48]. Sets of questions measure each of the three main constructs: "perceived usefulness" (Ui), "ease of use" (Ei), and "self-predicted future use" (Si).

Table 2. Scale items for measuring usefulness, ease of use and self-predicted future use.

Perceived Usefulness (PU)
U1. Using a web portal like FLOSScoach when joining a new OSS project, I would be able to contribute more *quickly*
U2. Using a web portal like FLOSScoach would improve the *performance* of newcomers contributing to OSS projects
U3. Using a web portal like FLOSScoach would enable newcomers to increase their *productivity*
U4. Using a web portal like FLOSScoach would enhance newcomers' *effectiveness* in contributing
U5. Using a web portal like FLOSScoach would *make it easier* for newcomers to contribute to OSS projects
U6. I would find a web portal like FLOSScoach *useful* to contribute to an OSS project
Perceived Ease of Use (PEOU)
E1. *Learning* to operate FLOSScoach would be easy for me
E2. I would find it easy to get FLOSScoach to *do what I want it to do*, to support the first steps of newcomers who want to make their first contribution to an OSS project
E3. My interaction with FLOSScoach would be *clear* and *understandable*
E4. It would be *easy to become skillful* in using FLOSScoach
E5. It is *easy to remember* how to perform tasks using FLOSScoach
E6. I would find FLOSScoach *easy to use*
Self-prediction of Future Use (SPFU)
S1. Assuming FLOSScoach would be available for any project, I *predict that I will use it in the future*
S2. I would *prefer* using FLOSScoach to the project pages for guiding me to contribute to an OSS project

We used the items presented by Laitenberger et al. [26] based on the item set originally proposed by Babar et al. [2]. To adapt the assessment instrument we: (i) used "FLOSScoach" as the object of the questionnaire and (ii) named the process: "newcomer contribution to OSS projects." We used a Likert scale to measure participants' perceptions, using a six-point semantic scale that asked for their degree of agreement with a statement , from "extremely disagree" (1) to "extremely agree" (6). In accordance with the study by Babar et al. [2], we chose a 6-point scale with no neutral value available so subjects would more clearly express their tendency towards a positive or negative evaluation.

In the following subsection, we present the results for FLOSScoach's perceived usefulness, ease of use, and prediction of future use. For this analysis, we considered all the 24 participants who used FLOSScoach from both Iteration 1 and 2.

5.3.2 Results

To assess usefulness, ease of use, and future use, we followed the Technology Acceptance Model. Before presenting the results, we first checked the questionnaire items' reliability and factorial validity. This was done to verify whether in this context the instrument and the observations provided reliable and valid results.

The reliability analysis was conducted to ensure the internal validity and consistency of the items used for each factor. A Cronbach's Alpha reliability level that exceeds a threshold of 0.8 indicates a reliable measure [6]. In our study, the Alpha values were 0.930 and 0.922 for Usefulness and Ease of Use items, respectively. Thus, the results demonstrate that the questionnaire is a reliable instrument.

Factorial validity is concerned with whether the usefulness and ease of use items form distinct constructs. We checked it using factor analysis. We present the adapted TAM questions' factor loading in Table 3. The literature reports the threshold level for sufficient loading [26] as 0.7; since the results for *ease of use* questions E1 to E6 load well with the first factor, we interpret this factor as "Ease of Use."

The second factor loaded well for the results of questions U1 to U6, as expected. We interpret this factor as "Usefulness." We can observe that U6 loads below the threshold of 0.7. However, since it loaded (slightly) higher on usefulness than on ease of use, we attributed it to the former, as in some other reports [2, 22].

Table 3. Factor validity for TAM's constructs.

	Ease of Use	Usefulness
Easy to learn (E1)	**0.92**	-0.23
Easy to perform (E2)	**0.70**	0.40
Clear and understandable (E3)	**0.71**	0.21
Easy to become skillful (E4)	**0.85**	0.16
Easy to remember (E5)	**0.87**	0.23
Easy to use (E6)	**0.91**	0.23
Work more quickly (U1)	0.53	**0.75**
Improve performance (U2)	-0.02	**0.91**
Increase productivity (U3)	0.06	**0.94**
Effectiveness (U4)	0.14	**0.95**
Makes job easier (U5)	0.49	**0.73**
Useful (U6)	0.50	0.58

Usefulness of FLOSScoach

The aggregated results for usefulness range between 18 and 34 (median 27.5). Considering that the maximum possible rating is 36, we conclude that most participants found FLOSScoach useful.

To better explore the results, we present each item's detailed results in Figure 5. The figure displays the large number of participants who agreed with the tool's usefulness. Furthermore, the number of participants who did not agree with U2, U3, U4, and U5 was at most four (for U4). Furthermore, U1 (quickness), U2 (job performance), U3 (productivity), and U6 (usefulness) had more than 50% of agreement or strong agreement.

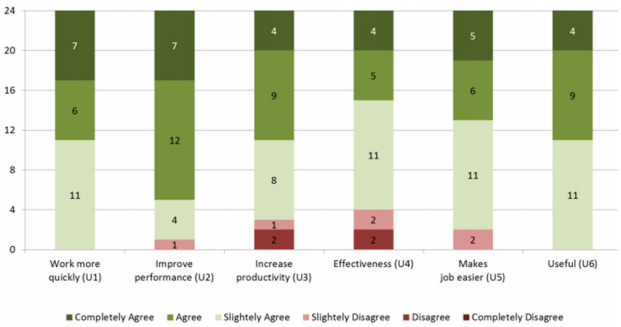

Figure 5. Perceived Usefulness: summary of the answers.

Ease of Use of FLOSScoach

The questionnaire items' aggregated score results for ease of use range between 16 and 36 (median 28.5). Regarding the maximum rating of 36, the ease of use rating can be considered high overall. The results are very similar to perceived usefulness. We can conclude that most participants found FLOSScoach easy to use.

In Figure 6, we can observe the answers' distribution per item. More than 50% of the participants agreed or strongly agreed with the items E1, E3, E4, E5, and E6, confirming how easy it was to use FLOSScoach.

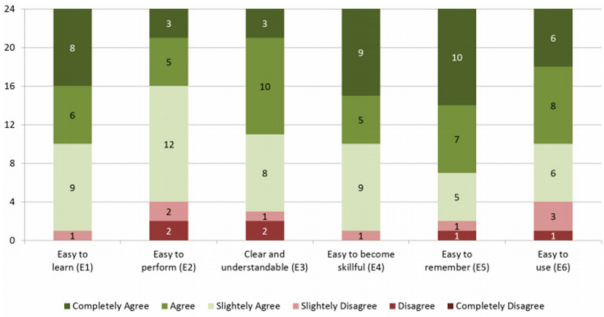

Figure 6. Perceived Ease of Use: summary of the answers.

Self-predicted Future Use

Figure 7 reports self-predicted future FLOSScoach use. We can observe that 15 participants (62%) agreed or strongly agreed that if FLOSScoach were available for any project in the future, they would use it (S1). Compared to the traditional way of finding information (using the project page), a large number of participants (21) agreed to some extent with a preference for FLOSScoach, and three disagreed at some level. Only one participant strongly disagreed, i.e., he preferred the traditional form to FLOSScoach.

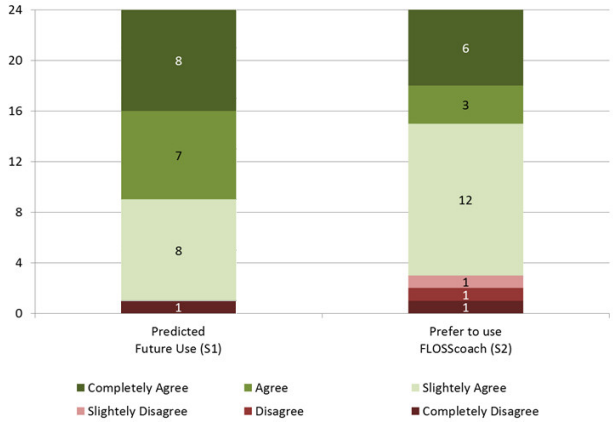

Figure 7. Self-Predicted Future Use: summary of the answers.

Correlating Usefulness, Ease of Use, and Self-predicted Future Use

Individually analyzing usefulness, ease of use, and self-predicted future use, we did not analyze how participants' opinion influenced user acceptance. We conducted a regression analysis to investigate the correlation of usefulness, ease of use, and self-predicted future use. According to the TAM model, usefulness and ease of use strongly correlate with self-predicted future use, i.e., the intention to use the technological innovation. Table 4 presents the correlations among the three factors. We observed a correlation between both variables and self-predicted future use.

These results show that usefulness and ease of use were important determinants that correlate with self-predicted future use. Most of our participants consider FLOSScoach useful and easy to use. Moreover, they indicated the possibility of future use, which firmly aligns with what we discovered in the diaries analyzed beforehand. We can conclude that participants perceived FLOSScoach as a useful and easy to use web portal, and would potentially use this tool to support future project contributions. Our results were positive in both iterations, with few negative scores.

Table 4. Correlation among Usefulness, Ease of Use, and Self-predicted Future Use.

	PU	PEOU	SPFU
Perceived Usefulness (PU)	1.00	0.65	0.59
Perceived Ease of Use (PEOU)	0.65	1.00	0.80
Self-predicted Future Use (SPFU)	0.59	0.80	1.00

6 DISCUSSION

In this section, we present changes we made to the barriers model based on the development and evaluation of FLOSScoach, we describe the impact of the web portal, and, based on our experience, we discuss features that have the potential of helping OSS project newcomers overcome entry barriers.

6.1 Updating the barriers model

After developing and evaluating FLOSScoach, we updated the barriers model using the feedback from this study as input. We made the following changes:

- A category called **communication** was created to accommodate problems related to **newcomers' communication** and community **reception issues**.
- **Cultural differences** barriers are now included in the **communication** and **newcomers' behavior** categories. This change was made in response to the barriers' proximity to the mentioned categories, which we perceived during the conception of the web portal.
- The **technical hurdles** category was dismissed, making **code/architecture hurdles**, **change request hurdles** and **local environment setup hurdles** first-level categories. This rearrangement was suggested after Iteration 1, because the two-level structure for accessing such important information was confusing.
- **Documentation problems** that crosscut other categories were accommodated into other categories, which retained "documentation" in their name. This change was based on suggestions from Iteration 1 participants, who reported existing redundancy.

The resulting model and more details can be found in [40].

6.2 Impact of FLOSScoach

We could not identify any new barrier that emerged due to the use of the web portal. The few issues reported by the participants were classified as enhancement opportunities, since they were mostly related to the way FLOSScoach presented information. Thus, we observed that categorizing the portal according to the barriers model benefitted the newcomers.

By conducting this study we could notice that offering the correct guidance or, as stated by Dagenais et al. [10], providing maps and signs brings benefit to newcomers, as they felt more confident and oriented starting from the very first contact with the project.

We understand that the web portal's main contribution is better orienting newcomers by organizing already provided strategies and documents. This brought direct results in terms of reducing **newcomers' orientation** barriers as well as **documentation** barriers.

We also observed that some social barriers, mainly those related to **communication**, previously categorized as **reception issues** and **newcomers' communication,** were softened by reducing the need

for community interaction, as reported by one participant: *"I did not need to talk to them. The tool was very clear. It is very easy, very good."* – P1-02

Using the portal did not eliminate all barriers. However, we observed that these barriers were reduced, since they were reported by the control group and not by the group using the web portal.

We also highlight the strong influence of technical issues. Barriers related to the **technical hurdles** category were recurrently reported by participants from both control and portal groups. We found that the recurrence and difficulty related to this category of barriers made the newcomers feel irritated and frustrated, which can lead to demotivation.

6.3 Features for newcomers

Based on our work, we can identify various features that have the potential of helping OSS project newcomers overcome entry barriers. For example, in our experience, providing a clear contribution flow helps newcomers gain confidence about what to do and in what order. Aligned with this, we suggest to create a newcomer specific page containing only the resources they need and not flood them with every possible resource, since too much information can confuse them.

The identification of tasks that are considered easy or simple for new contributors helps them finding their way. Some informative tags for the issue tracker that may guide newcomers include: difficulty level, module affected, language/technology skills needed, and members who can help. Many OSS projects already triage their issues in this way, but more important than triaging is making newcomers aware.

Regarding communication, providing a message template facilitates and encourages newcomers to interact with the community, and helps them sending a meaningful and complete message. Moreover, in our experience, it seems important to explicitly tell newcomers that it is important to find partial answers to their questions in the mailing list archives before dropping a question to the list. We addressed this by embedding a custom search box (which enables newcomers to search the mailing list archives) in the initial page of the "Communication" web portal section.

We believe that communities and researchers can put more effort into providing ways to facilitate newcomers' onboarding. According to our results, the points that deserve more attention are facilitating local workspace setup and providing ways to find the correct set of artifacts to work on once a task is selected.

7 THREATS TO VALIDITY

Although we used data from a variety of projects, the findings are not generalizable to all projects. We are aware that each project has its singularities and that the level of support and the barriers can differ according to the project. Our strategy of considering different projects aimed to explore different ways to use the web portal and overcome barriers.

Diary studies can introduce some threats to validity. First, it is impossible to ensure that participants write their diary entries in an unfiltered way. Second, diary studies follow a case-study approach. In the diary study, we aimed for in-depth understanding, rather than statistical validity. We read students' entries consistently, and frequently requested additional information or explanation, which should have counteracted any withholding.

Also regarding diaries, we acknowledge that there was no way to ensure that participants from both groups received equivalent feedback on their diaries. To reduce this possible threat, the researchers visited the diaries on a daily basis, posting comments to on all diaries that changed since the prior visit.

Regarding the possible benefit received by students that had been assigned to use FLOSScoach, we report that even students that were in the control group and were not able to contribute could receive an excellent grade, since the process was more important than the final result. We highlight that the median grade obtained by the control group was 7.0 (on a scale 0.0 - 10.0), while the median of those from the portal group was 7.25, and we did not find a statistical difference between both means applying the Mann-Whitney test ($p=0.3418$).

The use of students could have affected the results' generalizability. Moreover, most of our participants were novices to software development in real settings (with no previous industry experience), and thus it is possible that some barriers they faced are not specific to OSS development. However, we highlight that students are potential contributors to OSS projects. Students may have felt that they needed to provide positive feedback in the surveys. To avoid this, we emphasized to both groups that the feedback had no bearing on their grade. Moreover, according to previous studies [18, 36], students may provide an adequate model of the professional population. Runeson [35] identified a similar trend when comparing freshman, graduate, and professional developers.

Another potential threat to our results is the data subjectivity of the qualitative diary analysis. We grounded our findings in the data collected and presented excerpts to mitigate this threat.

8 CONCLUSION

In this paper, we presented and evaluated FLOSScoach, a web portal to support newcomers. We found that the web portal improved newcomers' experiences of the contribution process, serving as a compass during the contribution journey. This finding was evident both in the qualitative analysis of contribution diaries and in the self-efficacy results.

Newcomers who used the web portal felt more confident deciding which steps they needed to take to achieve their goals. The participants highlighted the contribution (step-by-step) flow and the helpful organization of information as FLOSScoach's main beneficial features. However, we could not identify any significant improvement in supporting newcomers to overcome technical barriers. The study's practical application of the barriers model led to observations that resulted in a new version of the model, which included some rearrangements.

The results of our work point to a set of features that can help OSS project newcomers overcoming entry barriers, including triaged lists of tasks, a clear contribution flow linked to useful documentation, and a message template to support newcomers' first interaction in mailing lists.

Currently, we are working on a new, more collaborative, version of FLOSScoach, which will enable creating and maintaining entries for other projects. We are also working on automating the identification, extraction, summarization, generation, and presentation of documentation that is relevant to newcomers in OSS projects. Using a variety of existing and novel natural language processing techniques, we propose to automatically parse documentation that is available for an OSS project from different sources, such as the official documentation on the projects' website, issue trackers, forums, Stack Overflow, or blogs.

9 ACKNOWLEDGEMENTS

The authors thank the students who participated in this study. We also thank Pernille Bjørn for suggesting the use of diary study and David Redmiles for the guidance provided during Igor's PhD. We thank CNPq (process 477831/2013), CAPES, Fundação Araucária, FAPESP (processes 2015/07399-1 and 2014/21899-4), NAP-SoL, NAWEB and FAPEAM for their financial support. Marco A. Gerosa receives individual grant from CNPq.

10 REFERENCES

[1] Adler, A., Gujar, A., Harrison, B.L., O'Hara, K. and Sellen, A. 1998. A Diary Study of Work-related Reading: Design Implications for Digital Reading Devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Los Angeles, California, USA, 1998), 241–248.

[2] Babar, M.A., Winkler, D. and Biffl, S. 2007. Evaluating the Usefulness and Ease of Use of a Groupware Tool for the Software Architecture Evaluation Process. *Proceedings of the First International Symposium on Empirical Software Engineering and Measurement* (Sep. 2007), 430–439.

[3] Bandura, A. 1986. *Social foundations of thought and action: a social cognitive theory*. Prentice-Hall.

[4] Begel, A. and Simon, B. 2008. Novice Software Developers, All over Again. *Proceedings of the Fourth international Workshop on Computing Education Research* (2008), 3–14.

[5] Canfora, G., Penta, M. di, Oliveto, R. and Panichella, S. 2012. Who is Going to Mentor Newcomers in Open Source Projects? *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering* (Cary, North Carolina, 2012), 44:1–44:11.

[6] Carmines, E.G. and Zeller, R.A. 1979. *Reliability and Validity Assessment*. SAGE Publications.

[7] Cassidy, S. and Eachus, P. 2002. Developing the computer user self-efficacy (CUSE) scale: investigating the relationship between computer self-efficacy, gender and experience with computers. *Journal of Educational Computing Research*. 26, 2 (Jan. 2002), 133–153.

[8] Cubranic, D., Murphy, G.C., Singer, J. and Booth, K.S. 2005. Hipikat: a project memory for software development. *IEEE Transactions on Software Engineering*. 31, 6 (Jun. 2005), 446–465.

[9] Czerwinski, M., Horvitz, E. and Wilhite, S. 2004. A Diary Study of Task Switching and Interruptions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria, 2004), 175–182.

[10] Dagenais, B., Ossher, H., Bellamy, R.K.E., Robillard, M.P. and Vries, J.P. de 2010. Moving into a new software project landscape. *Proceedings of the 2010 ACM/IEEE 32nd International Conference on Software Engineering* (New York, NY, USA, 2010), 275–284.

[11] Davidson, J.L., Mannan, U.A., Naik, R., Dua, I. and Jensen, C. 2014. Older Adults and Free/Open Source Software: A Diary Study of First-Time Contributors. *Proceedings of The International Symposium on Open Collaboration* (2014), A5.

[12] Davis, F.D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*. 13, 3 (Sep. 1989), 319–340.

[13] Fishbein, M. and Ajzen, I. 1975. *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley Pub.

[14] Fogel, K. 2013. *Producing Open Source Software: How to Run a Successful Free Software Project*. O'Reilly Media.

[15] Halfaker, A., Kittur, A. and Riedl, J. 2011. Don't Bite the Newbies: How Reverts Affect the Quantity and Quality of Wikipedia Work. *Proceedings of the 7th International Symposium on Wikis and Open Collaboration* (2011), 163–172.

[16] Hars, A. and Ou, S. 2001. Working for free? Motivations of participating in open source projects. *Proceedings of the 34th Annual Hawaii International Conference on System Sciences* (2001), 1–9.

[17] Hess, J. and Wulf, V. 2009. Explore Social Behavior Around Rich-media: A Structured Diary Study. *Proceedings of the Seventh European Conference on European Interactive Television Conference* (Leuven, Belgium, 2009), 215–218.

[18] Höst, M., Regnell, B. and Wohlin, C. 2000. Using Students as Subjects – A Comparative Study of Students and Professionals in Lead-Time Impact Assessment. *Empirical Software Engineering*. 5, 3 (Nov. 2000), 201–214.

[19] Jensen, C., King, S. and Kuechler, V. 2011. Joining Free/Open Source Software Communities: An Analysis of Newbies' First Interactions on Project Mailing Lists. *Proceedings of the 44th Hawaii International Conference on System Sciences* (Jan. 2011), 1–10.

[20] Jergensen, C., Sarma, A. and Wagstrom, P. 2011. The Onion Patch: Migration in Open Source Ecosystems. *Proceedings of the 19th ACM SIGSOFT Symposium and the 13th European Conf. on Foundations of Software Engineering* (2011), 70–80.

[21] Kersten, M. and Murphy, G.C. 2005. Mylar: A Degree-of-interest Model for IDEs. *Proceedings of the 4th International Conference on Aspect-oriented Software Development* (Chicago, Illinois, 2005), 159–168.

[22] King, W.R. and He, J. 2006. A meta-analysis of the technology acceptance model. *Information and Management*. 43, 6 (Sep. 2006), 740–755.

[23] Kraut, R.E., Burke, M., Riedl, J. and Resnick, P. 2012. The Challenges of Dealing with Newcomers. *Building Successful Online Communities: Evidence-Based Social Design*. R.E. Kraut and P. Resnick, eds. MIT Press. 179–230.

[24] Krogh, G. von and Hippel, E. von 2003. Editorial: Special issue on open source software development. *Research Policy*. 32, 7 (Jul. 2003), 1149–1157.

[25] Krogh, G. von, Spaeth, S. and Lakhani, K.R. 2003. Community, joining, and specialization in open source software innovation: A case study. *Research Policy*. 32, 7 (2003), 1217–1241.

[26] Laitenberger, O. and Dreyer, H.M. 1998. Evaluating the usefulness and the ease of use of a Web-based inspection data collection tool. *Proceedings of the Fifth International Software Metrics Symposium* (Nov. 1998), 122–132.

[27] Malheiros, Y., Moraes, A., Trindade, C. and Meira, S. 2012. A Source Code Recommender System to Support Newcomers. *Proceedings of the IEEE 36th Annual Computer Software and Applications Conference* (2012), 19–24.

[28] Nakakoji, K., Yamamoto, Y., Nishinaka, Y., Kishida, K. and Ye, Y. 2002. Evolution Patterns of Open-source Software Systems and Communities. *Proceedings of the International Workshop on Principles of Software Evolution* (Orlando, Florida, 2002), 76–85.

[29] Naur, P. 1983. Psychology of Computer Use. *Psychology of Computer Use*. T.R.G. Green, S.J. Payne, and G.C. van der Veer, eds. Academic Press. 159–170.

[30] Park, Y. 2008. *Supporting the learning process of open source novices : an evaluation of code and project history visualization tools*. School of Electrical Engineering and Computer Science - Oregon State University.

[31] Park, Y. and Jensen, C. 2009. Beyond pretty pictures: Examining the benefits of code visualization for open

source newcomers. *Proceedings of the 5th IEEE International Workshop on Visualizing Software for Understanding and Analysis* (Sep. 2009), 3–10.

[32] Pinto, G., Steinmacher, I. and Gerosa, M. 2016. More Common Than You Think: An In-Depth Study of Casual Contributors. *23rd IEEE International Conference on Software Analysis, Evolution, and Reengineering, SANER 2016, Osaka, Japan* (2016).

[33] Qureshi, I. and Fang, Y 2011. Socialization in Open Source Software Projects: A Growth Mixture Modeling Approach. *Organizational Research Methods*. 14, 1 (Jan. 2011), 208–238.

[34] Rieman, J. 1993. The Diary Study: A Workplace-oriented Research Tool to Guide Laboratory Efforts. *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands, 1993), 321–326.

[35] Runeson, P. 2003. Using students as experiment subjects–an analysis on graduate and freshmen student data. *Proceedings of the 7th International Conference on Empirical Assessment in Software Engineering* (2003), 95–102.

[36] Salman, I., Misirli, A.T. and Juristo, N. 2015. Are Students Representatives of Professionals in Software Engineering Experiments? *37th IEEE International Conference on Software Engineering (ICSE 2015)* (May. 2015), 666–676.

[37] Scacchi, W. 2002. Understanding the requirements for developing open source software systems. *IEE Proceedings Software*. 149, 1 (Feb. 2002), 24–39.

[38] Schilling, A., Laumer, S. and Weitzel, T. 2012. Who Will Remain? An Evaluation of Actual Person-Job and Person-Team Fit to Predict Developer Retention in FLOSS Projects. *Proceedings of the 2012 45th Hawaii International Conference on System Sciences* (Washington, DC, USA, 2012), 3446–3455.

[39] Smith, S.A., Kass, S.J., Rotunda, R.J. and Schneider, S.K. 2006. If at first you don't succeed: Effects of failure on general and task-specific self-efficacy and performance. *North American Journal of Psychology*. 8, 1 (Apr. 2006), 171–182.

[40] Steinmacher, I. 2015. *Supporting newcomers to overcome the barriers to contribute to open source software projects*. University of São Paulo. Available at http://www.teses.usp.br/teses/disponiveis/45/45134/tde-30112015-131552

[41] Steinmacher, I., Chaves, A.P., Conte, T. and Gerosa, M.A. 2014. Preliminary empirical identification of barriers faced by newcomers to Open Source Software projects. *Proceedings of the 28th Brazilian Symposium on Software Engineering* (2014), 1–10.

[42] Steinmacher, I., Conte, T., Gerosa, M.A. and Redmiles, D. 2015. Social Barriers Faced by Newcomers Placing Their First Contribution in Open Source Software Projects. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada, 2015), 1379–1392.

[43] Steinmacher, I., Conte, T., Gerosa, M.A. and Redmiles, D.F. 2015. Social Barriers Faced by Newcomers Placing Their First Contribution in Open Source Software Projects. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada, Feb. 2015), 1–13.

[44] Steinmacher, I., Silva, M.A.G., Gerosa, M.A. and Redmiles, D.F. 2015. A systematic literature review on the barriers faced by newcomers to open source software projects. *Information and Software Technology*. 59, (Mar. 2015), 67–85.

[45] Steinmacher, I., Wiese, I.S., Chaves, A.P. and Gerosa, M.A. 2013. Why do newcomers abandon open source software projects? *Proceedings of the 2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering* (2013), 25–32.

[46] Strauss, A. and Corbin, J.M. 2007. *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory*. SAGE Publications.

[47] Symon, G. 2004. Qualitative research diaries. *Essential Guide to Qualitative Methods in Organizational Research*. C. Cassell and G. Symon, eds. SAGE publications. 98–113.

[48] Vaz, V., Conte, T. and Travassos, G.H. 2013. Empirical Assessments of a tool to support Web usability inspection. *CLEI Electronic Journal*. 16, 3 (Dec. 2013), 16 pp.

[49] Vora, P. and Komura, N. 2010. The n00b Wikipedia Editing Experience. *Proceedings of the 6th International Symposium on Wikis and Open Collaboration* (2010), Article 36.

[50] Wang, J. and Sarma, A. 2011. Which bug should I fix: helping new developers onboard a new project. *Proceedings of the 4th International Workshop on Cooperative and Human Aspects of Software Engineering* (Waikiki, Honolulu, HI, USA, 2011), 76–79.

[51] Ye, Y. and Kishida, K. 2003. Toward an Understanding of the Motivation Open Source Software Developers. *Proceedings of the 25th International Conference on Software Engineering* (Portland, Oregon, 2003), 419–429.

[52] Zhou, M. and Mockus, A. 2012. What make long term contributors: Willingness and opportunity in OSS community. *Proceedings of the 34th International Conference on Software Engineering* (Jun. 2012), 518–528.