Abstract

Evaluation of group support is a complex task and its design must be carefully planned in order to produce useful results. Many factors may influence the evaluation design, some of them have been suggested in the literature. However, in a particular evaluation context, additional factors may arise, influencing and even compromising the evaluation design. We argue that preliminary evaluations must be performed before they are conducted in a real setting in order to verify if its design will provide the expected measurements and results. This paper presents the evaluation process of awareness information resources within software diagrams peer reviews. The paper describes the evaluation context, objectives, plan and design together with the enactment of a preliminary case study following this design. From this preliminary study, we outline some factors that arose during its enactment that influenced the refinement of the evaluation design.

1. Introduction

Software development processes are remarkably collaborative. Its collaborative nature has motivated the proposal and construction of a variety of tools and environments for supporting collaboration in this context [1][8][9][10].

Odyssey [13][6] is a research project aiming at defining an environment for component-based software development support. The collaborative aspect of software development is being considered within Odyssey Software Development Environment (Odyssey SDE) by the definition of tools for group interaction support. This collaborative version of the Odyssey SDE is now being called OdysseyShare SDE [7].

A set of awareness widgets was implemented in OdysseyShare for supporting remote and synchronous interactions. The widgets were built as software components in such a way that any tool provided in the environment could use them. In this work, we are concerned about how these widgets can be effectively used within OdysseyShare for supporting synchronous and on-line peer reviews. Particularly, we want to evaluate how these awareness widgets reduce the need of participants to explicitly indicate their location while interacting through a shared workspace during a review.

We are aware that promoting collaboration is a cultural challenge that does not only rely on the provision of supporting tools. To be effectively used, any tool must be tuned to the process participants’ objectives, needs and expectations. There is an increasing discussion nowadays about how collaborative tools must be evaluated in order to assure that they meet the group interaction needs and also to verify what are their impacts in the group culture [2][3][4].

Evaluation of group support is a complex and costly task, and its design must be carefully planned in order to produce useful results. We argue that preliminary executions of the evaluation design are important to verify if it will provide relevant results before they are conducted in a real setting. Besides, the design of groupware evaluations is an iterative process. After running an evaluation, results should be considered in order to refine its design.

The aim of this paper is to describe the evaluation plan and design of the awareness widgets use in OdysseyShare SDE within peer reviews. It also describes the enactment of a preliminary case study following this design. From this preliminary study, we outline some factors that arose during its enactment and how they should change the evaluation design. These results showed us how important the preliminary evaluation was to fine-tune our experiment before it was applied in a real setting.

The paper is structured as follows. Section 2 presents the objectives of the OdysseyShare SDE and the awareness support components it provides. In Section 3,
we present the method used for conducting the evaluation: its planning and design, the enactment of the preliminary evaluation and how its outcomes impact the evaluation design. Section 4 concludes the paper.

2. The OdysseyShare Environment

The OdysseyShare Project aims at constructing a collaborative software development environment to support Component Based Software Development (CBSD). CBSD is a particular approach in software engineering that aims at software reuse. CBSD can be defined as an approach to software development in which all software artifacts – from executable code to interface specifications, architectures, and business models – can be built by assembling, adapting, and wiring together existing components in a variety of configurations [11].

A CBSD method can rely on domain-specific information, i.e., concepts, processes, and tasks that are specific to an organization or business area, in order to build component specifications [12]. Thus, a domain-specific component may depend on information that is beyond the knowledge of a software engineer. On the other hand, domain experts usually have no knowledge about software development techniques but they have a deep insight of specific domains such as health care, telecommunications, finances, or law.

In this scenario, a collaborative software development environment aimed to support component-based software relying on domain models must focus on issues related to coordination, cooperation, and especially in the establishment of a community of domain experts and software engineers.

We expect that software engineers and domain experts spend most of their time in the collaborative authoring of documents while describing a domain, whether they be textual or graphical descriptions. Thus, collaborative authoring was identified as their primary activity within OdysseyShare. Therefore, our strategy is to at first place provide components that implement an adequate set of small possible actions needed to perform the editing in a shared workspace [14].

2.1. Awareness Components

Two classic widgets for real-time collaborative authoring were implemented as software components: a telepointer and a radar view component, both are commonly available in groupware frameworks and kits [17][16]. Odyssey SDE and OdysseyShare Components are implemented using The Java Platform [20].

The radar view is an awareness widget commonly used in graphical and textual editors to help in determining current editing position in a large document. For example, Acrobat Reader, Aladdin Ghost View, and Microsoft PowerPoint provide radar views. A radar view component, in the context of groupware, displays information about the local user and remote users’ positioning, as well. The user can scroll editing window by clicking within the radar view window.

The telepointer presents the position of the remote user’s mouse, or another pointing device, cursor in the document. It requires no additional user interaction; mouse pointer information is captured automatically. Telepointers are popular in teleconferencing, but are seldom seen in non-groupware applications.

Figure 1 – Radar view (small window at left hand side) and telepointers (cursor hands at center) in the OdysseyShare SDE Collaborative Editor.

On a real-time collaborative editing session, these awareness components transfer information about the local workspace to other participants, increasing their perception of the other users’ actions and location. This kind of workspace awareness information would have a positive effect on the collaborative editing and on the shared workspace in general [15].

The development of software components in OdysseyShare follows an iterative and incremental process. Before adding more functionality or greater reliability to these two components there shall be a measure of their applicability within the environment. Therefore, conducting evaluations on the use of the proposed components plays an important role on their own design and also on their use within the OdysseyShare development activities.

3. Evaluation Method

It is a consensus in groupware research that groupware evaluations are costly [14]. Therefore, it is essential to have a method for conducting evaluations that encompasses its planning, design, enactment and analysis,
in order to obtain significant results. In the next sections we present our decisions and outcomes in each of these steps.

3.1. Planning

The evaluation planning was guided by the definition of the evaluation hypotheses, the target population and the task to be performed during the evaluation.

Hypothesis. The specific communication challenge we address in this work is the difficulty that a participant is faced when trying to maintain its partners informed about his current location and editing position in a WYSIWIS-relaxed session in the shared space. Participants may overcome this challenge by explicitly verbalizing their position to his remote partner through audio or text chat. This explicit verbalization may burden the communication and, when inadequate, may provide a negative influence on interaction productivity and participant satisfaction.

Our hypothesis is that the awareness widgets developed within OdysseyShare SDE help participants to locate each other in the shared workspace, thus, allowing them to maintain the focus on the task being performed. As a consequence, this feature will lead to interactions that are: (a) more productive: decreasing the need for verbal orientation; (b) less error prone: by preventing misunderstandings; and (c) more satisfactory: making participants feel that the task was performed in a productive and correct way.

Population. OdysseyShare users compose the target population of our evaluation. These users are computer science students because Odyssey SDE is an academic tool. Should this evaluation be applied to a commercial tool, the population should be changed accordingly. Evaluation results, therefore, are used to gain some understanding about a localized group of environment users. We do not expect that results generalize to a broader population, such as component-based software engineers in general.

Task. A peer review task allows us to test our evaluation hypothesis focusing on how the components can effectively provide information about participants’ position in the shared space.

The purpose of a peer review [5] is to look for defects and improvement opportunities in a work artifact under construction. It is argued that by conducting peer reviews as formal activities throughout the development process the chances of finding errors earlier is increased and thus the costs of fixing them is decreased.

Review sessions do not aim at editing an artifact to remove its defects. While conducting reviews through online interactions, the challenge is focused on window sliding and view synchronization.

We are not concerned about providing full support for a defined review methodology. In the Odyssey SDE, the development must follow a defined process where peer reviews activities could be formally placed. However, we focus on the conduction of informal review interactions. Within OdysseyShare, collaboration is motivated by allowing users to conduct peer reviews in an ad hoc manner, i.e., whenever process participants want to, using any of the provided development tools as the scenario for conducting the reviews.

We could plan their evaluation in the context of a collaborative editing session. However, collaborative editing is a complex task that comprises other factors than besides supporting participants’ localization. In order to control the number of other variables in the evaluation, we decided to use a task where the positioning in the shared space plays a fundamental role.

3.2. Design

The evaluation design comprises the definition of the set of variables to be measured during the evaluation session, the instruments to collect data and the particular scenario for task accomplishment.

Variables. In order to validate the hypothesis, we defined a set of dependent variables, addressing each outlined issue:

- **V1.** Amount of verbal orientation between participants during their interaction.
- **V2.** Task productivity. Measured by:
  - **V2.1.** Time spent doing the task;
  - **V2.2.** Task correction.
- **V3.** Participants’ satisfaction performing the task using the shared workspace.

Independent variables address the following issues:

- **V4.** Participants’ knowledge and experience. In the following topics:
  - **V4.1.** Peer review practice;
  - **V4.2.** Shared workspaces operation;
  - **V4.3.** Odyssey SDE operation.

Instruments. The evaluation involved five instruments to record participants’ commitment, actions and contributions. Namely they are: a term of agreement, before and after review session questionnaires, an after-session informal group talk, and in-session participants voice and screen recording.

A term of agreement is a document to state clearly to participant the evaluation objectives, his responsibilities within the evaluation, expected session duration in time, and any other information that is required for the participant to decide if he is going to accept to continue
participating. This document is brief and requires the participant’s signature. The term also informs that evaluation data is not subject to be used to rank individual performance nor will influence his evaluation as student or as employee, whenever this applies to the participant.

The term is sent to the participant in advance of the evaluation session and asks for an appropriate date for the session to take place or other data relevant to session planning.

Questionnaires are adopted to make the session more objective and to make participants concentrate on a particular aspect of the evaluation. The phrasing of questions and adoption of standard scales are important aspects of the questionnaire design. We choose to adopt standard answer scales, whenever they apply.

There are two questionnaires: a before session questionnaire which aims at obtaining information about participant background, skills, and other descriptive data (independent variables); and an after session questionnaire which aims at obtaining data about the session from the participant point-of-view. The after session questionnaire is based on previous usability and system evaluation questionnaires [19].

In an after session informal talk the evaluation team explore participant’s opinion in a less constrained format than the questionnaire. This talk is performed with the presence of both participants. In this talk, participants are able to comment different aspects of the evaluation. This talk is also used to confront participant’s answers, to try and explore differences between them and obtain their rationale for action.

Session recording helps the evaluation team during analysis. To record the session we use screen and voice recording software. We decided to use screen recording instead of video recording because it captures the participant point-of-view. We adopted screen recording using TechSmith Camtasia recording software. Lotus ScreenCam failed to record Java-based applications.

Scenario. A set of three documents provides the task description. A document stating the task objectives, a requirement description document and an UML class diagram.

The requirement description document is divided in two parts, to induce participants’ communication. In this way, both participants need to share and obtain information from his partner in order to review the whole diagram.

During the evaluation, the UML diagram is available on-line on the screen; the other two documents are distributed as hard copies.

The whole evaluation process involves the following activities:
- The participants fill out the agreement term;
- the participants fill out the first questionnaire;
- the participants use the Odyssey environment and its widgets to collaborate in reviewing the class diagram based on the requirements description they received;
- the participants fill out the second questionnaire.

3.3. Preliminary Evaluation

The evaluation planning and design were tested in a preliminary enactment to detect problems and to determine their viability. This preliminary evaluation was performed before the actual enactment step.

Two voluntary graduate software engineering students participated in our preliminary evaluation. Both students are Odyssey SDE users familiarized with UML diagrams. They had already participated in informal source code peer reviews, but have not reviewed UML diagrams before.

The preliminary evaluation occurred in a room having the participants close enough for easy conversation but disposed in such a way that no direct view of the other participant display was possible.

The evaluation team performed the necessary software installation, OdysseyShare SDE login and session setup. Hear-phones and microphones were installed for each participant. A one-minute voice and screen recording test was performed before the scenario enactment.

These preliminary evaluation participants have a deep knowledge of the Odyssey SDE. They have participated in the programming of the original Odyssey SDE and are regular users of the system. This evaluation was their first session using the OdysseyShare SDE.

Participants declared insufficient previous experience in peer reviews. However, both participants recall previous training in reviews in their software engineering classes. Both participants reserved the term ‘software review’ to state that a formal software review is being performed.

Both participants acknowledge no previous experience in shared workspaces, although they admit intensive use of Internet, instant messaging and chat programs. Participants do not recognize on-line computer games as shared spaces.

3.4. Analysis

Our analysis is made based on the data obtained in the preliminary study. This data is insufficient to assess correlations among evaluation variables and to test our hypothesis. Instead we were focused on analyzing the
evaluation viability. Overall scenario and infrastructure setup were demonstrate feasible in the preliminary evaluation. However some problems have been detected:

**Measurement instruments.** Verbal orientation recording was successful, but the measurement and analysis procedure must be enhanced. The process of scanning voice recording is tedious and imposes some restrictions to the evaluation of a large number of observations. Further, separated voice recording requires a separated scanning for each participant. So, a 40-minute session requires an 80-minute scanning.

Verbal orientation occurrences vary in importance and complexity. A classification of verbal orientation information is a refinement to consider in voice recording analysis [21]. Thus, in future analysis evaluations, occurrences of verbal orientation should be associated to a given record time, to its refereeing objects (e.g. class diagram) and also to a given verbal class.

Confronting data from informal talks and questionnaires reveal that participant interpretation of questions differs from the evaluators’ intention. The participants were more formal in the questionnaire than in the talk, as expected. In general, participants have taken the phrasing in a more formal way than was initially intended by evaluators.

Question phrasing was problematic in some points. Questions asking for “the best computer program”, or citing “a book on peer review” created a feeling of uneasiness to the participants. They consider that a simple rephrasing such as “one of the best programs” or “some book” would lead to a better question statement.

**Task scenario and material.** Time spent to perform the task and task correction were both within expected values. Some “defects” where undetected due to the high-level nature of the software diagram. Evaluators and participants were not able to reach a consensus about some diagram details such as relationship cardinality. In some interpretations, a detail is seen as a defect and in others it is not, but both seems to be plausible interpretations. Future evaluations should include questions to determine participant interpretation.

Screen recording analysis reveals the prevalence of the use of the voice channel over the use of awareness mechanisms. Radar view and telepointers were active during the evaluation session. In the informal talk, participants state that they do not feel the need to interact with radar view and believe that a more complex task would be more useful to evaluate awareness support.

For instance, screen recording reveals that scrollbars were used to move windows instead of the click-scroll control available in the radar view component. One participant screen recording reveals that he is aware of this control, but he preferred to use scrollbars. Nevertheless, the participants state that they feel more comfortable in using traditional scrollbars.

This awareness support sub-utilization could be a consequence of the task scenario. The participants considered the textual document as correct and complete. Most of verbal localization events referred to this document, which is hard-copied. Besides that, this requirement document was written in natural language and its structure is linear. We have evidence that most of the scrolling activity was performed off-line over this document.

As a curiosity, the same participant states that he would prefer a drag-scroll control instead of the available click control. This participant asked for a new functionality: to be able to move another participant window during a session. He expects to obtain the attention of another user by scrolling or by ‘shaking’ a remote window.

As a conclusion, the support and recording infrastructures worked as expected. Question phrasing problems were detected and possibly corrected.

The introduction of an easy to manipulate hard-copied, linear textual document was the most problematic point revealed in the preliminary evaluation. It seems that attention was taken away from the editing support by the introduction of this document. Currently, we are concerned on designing a new scenario of evaluation that can be performed without a textual specification.

Despite of radar view control apparent sub-utilization, according to the second questionnaire, the participants considered the shared workspace useful to realize the review task.

### 4. Conclusion

The aim of this paper was to present and discuss the steps performed to design the evaluation of awareness components developed within the OdysseyShare - a component-based software development environment. The awareness components were built as a first step towards a strategy for providing OdysseyShare with adequate support for the interaction among the environment users – software engineers and domain experts – while editing software engineering documents. This work is an example of the complexity of group support evaluation. To obtain relevant results, an evaluation must be carefully planned in advance. Even if carefully planned, factors occur during the evaluation that may indicate the need of refinements. For instance, our preliminary evaluation revealed problems with the questionnaires, the task scenario and the technological infrastructure used to record the review session.

These findings reinforce our recommendation that preliminary evaluations should be conducted in order to validate the evaluation design and to check out any misconceptions.
As future work, we intend to refine the evaluation based on our major findings and to perform a sequence of evaluations using different audiences/population and tasks. These future evaluations will bring us concrete data about the components use and how they affect the collaborative work within OdysseyShare.

Acknowledgements
The authors would like to thank CAPES, CNPq, and FAPERJ (process #E-26/152.176/2000) for financial support.

References