

The CSCW Lab for Groupware Evaluation

Renata Mendes de Araujo^{1*}, Flavia M. Santoro^{2**}, and Marcos R.S. Borges³

¹Departamento de Ciências da Computação/UFRJ and UNIRIO
Caixa Postal 2324, Rio de Janeiro, 20001-970, RJ, Brasil
renata@nce.ufrj.br

²Departamento de Ciências da Computação/UFRJ
fmsantoro@yahoo.com.br

³Departamento de Ciências da Computação and NCE/UFRJ
mborges@nce.ufrj.br

Abstract. The aim of this paper is to report the initiative of a research project called the CSCW Lab. The CSCW Lab is an approach for applying evaluation methodologies in the context of a groupware research group. We identify the major dimensions of groupware evaluation and describe how the CSCW Lab addresses them. The first experiments using CSCW Lab are also described.

1 Introduction

Groupware evaluation is becoming a relevant research issue. The research area came to a point where there have been a lot of proposals and solutions. Now, the great challenge is to be aware of how effective are these solutions in real situations. The claim for groupware evaluation can be observed by the number of papers and research reports addressing this issue and by the recent workshops totally devoted to this theme [3][4]. We are very far from consensus about how groupware evaluation should be conducted although we see some initial results for building its “body of knowledge” [1][2][3][4][5][6][7].

Within the context of our research group – CHORD – many tools and prototypes have been specified and developed as the result of undergraduate, master and doctoral work [13]. The main difficulty faced by our research group is how to accomplish the evaluation of these tools and prototypes in order to verify if they answer the hypotheses outlined at the beginning of the research work. What is the scope of the evaluation, how do we define variables to measure the results, which instruments should we apply, how do we choose participants and settle down the groups that will take part in the experiment, how far should we go to validate our hypotheses?

In this work we present the CSCW Lab – an environment for the first-step evaluation of groupware prototypes. Besides an environment for evaluating our research

* sponsored by FAPERJ (process #E-26/152.176/2000).

** sponsored by FAPERJ (process #E-26/152.116/2001).

products, the CSCW Lab is also a research project, since we intend to study existing methodologies applied to groupware evaluation as well as the definition of new methods, instruments and/or tools. CSCW Lab defines a method for groupware evaluation based on the dimensions for evaluation and the steps to investigate each dimension.

This paper is structured as follows: in section 2 we present a summary of the major issues and approaches comprising groupware evaluation. In section 3 we identify and discuss the possible dimensions of groupware evaluation and how they influence each other. In section 4 we present the ideas concerning the CSCW Lab and how each dimension of evaluation could be addressed. In section 5 we illustrate the paper with some examples of evaluations conducted within the CSCW Lab. Finally, section 6 concludes the paper.

2 Groupware Evaluation: What Has Been Done?

Many authors have been reporting problems with the development and use of groupware applications. Evaluation failures of CSCW systems may be partially responsible for slow adoption of such systems. The system performance may depend on the varied behavior and personalities of the group members, the effect of social, motivational, economic and political dynamics, and the relevance of time as a factor in understanding interaction changes. All these issues interfere in the way people use a groupware, making it difficult, or almost impossible, to identify and to obtain control over all variables related to it. Thus, groupware evaluation is expensive and produces few general results.

Nevertheless, without the appropriate evaluation mechanisms, the developer's community does not accumulate enough experience to learn and build better systems [8]. It is claimed that four important, but still opened, topics should be considered in this case: What do we want to evaluate? What methodological approach should be taken? What criteria should be used to achieve the results? What instruments should be used?

There is no consensus on the methodology to be adopted in order to perform evaluations in groupware. Pinelle and Gutwin [2] present a survey on the most used evaluation methodologies, based on the works presented at the main conferences of the CSCW area. They classified the evaluations both in relation to the environment where they are accomplished (natural occurrence or simulation of the phenomenon), and the degree of the variables manipulation (rigorous or minimum control of variables) (Table 1). Their conclusion is that each work used different approaches, methodologies or techniques for conducting evaluations.

Table 1. Evaluation Classifications [2]

		Manipulation	
		<i>Rigorous</i>	<i>Minimal/None</i>
Setting	<i>Naturalistic</i>	– Field Experiment	– Field Study – Case Study
	<i>Controlled</i>	– Laboratory Experiment	– Exploratory

Randall et al. [9] had also identified four orthogonal dimensions to classify the kinds of evaluation in groupware: Summative X Formative; Quantitative X Qualitative; Controlled Experiments X Ethnographic Observations; Formal and rigorous X Informal and opportunistic.

The authors state that the most used types of evaluations are the summative-controlled and experimental (considered a formal technique); and the formative-qualitative-opportunistic approaches (considered an informal technique).

Concerning the criteria for evaluating groupware, Baker et al. [8] proposed to analyze groupware through appropriate heuristics, which Nielsen, from the HCI (Human-Computer Interaction) area, defines as “general rules used to describe common properties of usable interfaces”. The heuristics are related to communication (verbal, gesture, body, shared artifacts), protection, activities and collaboration management and contact establishment.

Some other works are more concerned with quantitative matters, such as Baeza-Yates and Pino’s [10], who concentrate on the relationships among issues like the quality of a work outcome, the time spent on it and the total amount of work done.

Monk et al.[11] state that measures based on task performance – for instance, how well or quickly the work is completed - are only sensitive to gross changes in the facilities available for communication. They assume that looking directly at the effect of a manipulation on communication, rather than indirectly via the effect that communication has on the work, the effects observed will be clearer and easier to detect and the results will be easier to interpret, leading to a better understanding and hence more generalizable findings.

Steves and Allen [12] point out some other important issues related to groupware evaluation: a better evaluation will be supported by improved data collection, data formats, data categorization and data visualization tools; researchers cannot control all independent variables, and do not usually relate how they were handled in a case study documentation, that is, the analysis of data collection is typically labor intensive. Knutilla, Steves and Allen [3] asserts that researchers need tools to measure the incremental progress towards developing useful collaborative systems, as well as methods to evaluate the impact of specific technologies on collaboration.

Additionally, the ultimate goal of any evaluation is to validate some theory. The obstacles found in a groupware evaluation process makes it very hard to achieve, if we consider the straight meaning of validation. Therefore, Randall et al. [9] suggest that validation cannot be done in practice, but a provisional status should be accepted. For these authors, evaluations should focus on potential changes in human practice as well as machine functionality, and on the relationship between them.

Considering this variety of ideas, suggestions and approaches for groupware evaluation, the main objective of the CSCW Lab is to define a method for groupware evaluation comprising the steps for conducting the evaluation and guidelines for using any technique and instruments.

3 Dimensions of Groupware Evaluation

The first issue addressed in the CSCW Lab in order to define a method for groupware evaluation is to identify the concepts or dimensions that must be considered while evaluating groupware. We identify four dimensions for groupware evaluation (Fig. 1).

These dimensions are a first step towards building a conceptual framework for our studies in the CSCW Lab.

The conceptual framework considers that while evaluating groupware we may be addressing how to: evaluate and describe the context where the application under evaluation will be used; evaluate the application usability strengths and weakness; evaluate the level of collaboration achieved while using the application; and evaluate the technological and cultural impact achieved with its use along the time.

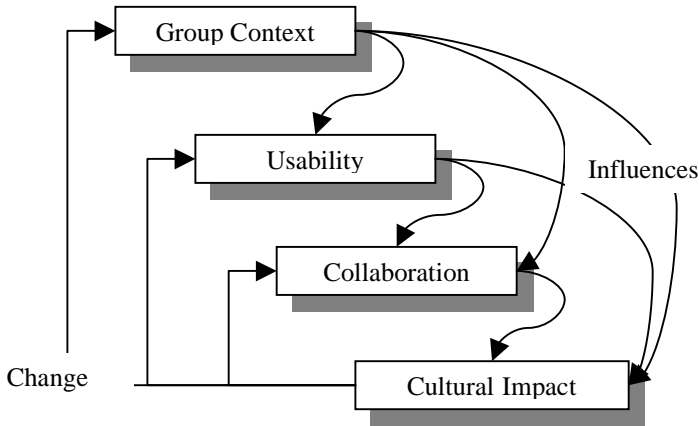


Fig. 1. Dimensions for groupware evaluation

Group context. It is a consensus in groupware evaluation research that groups are quite unique. Even if we try hard, it is almost impossible to find two groups with the same values to conform to our independent variables. Often we cannot find the “ideal” group to conduct our evaluations. To find or to build groups for evaluation is difficult and costly.

In groupware evaluation we must care about group diversity. While conducting evaluations, we should target on: comparing results from different groups, collecting empirical or statistical evidence on the effectiveness of a new technology or just observing a unique case on the field. In any case, we must look into the many different outcomes when supporting collaboration for different groups. To study the effects on diversity, a sharp characterization of the observed group is essential for the interpretation and discussion of any evaluation result.

Also, our framework considers that the singularities of the group context can influence all other evaluation dimensions. For instance, if a group is initially highly committed to perform a task, it is possible that they overcome any usability problems with the application in use. It is also possible that the level of collaboration will be high, since they are willing to work together and, finally, it is also possible that the tool will have a positive impact in the work or interaction environment.

System usability. The next dimension is evaluating usability. It is widely known that the level of usability of an application determines its proper use and also its acceptance by users. Evaluating groupware usability is a complex task since it

involves not only the evaluation of how a user can use the tool but also if the group can interact as expected using the tool [15][16].

Level of collaboration. Collaboration may occur at many levels and depends a lot on the nature and objectives of the group task. To evaluate collaboration it is first necessary to determine what are the measures or variables that determine how people collaborate. For instance, in a discussion forum, one possibility to measure collaboration is to count the number of contributions generated by the group. However, collaboration in a forum is only effective if contributions are not only inserted but also read by other participants [17].

Measuring collaboration also involves subjective metrics. Usually, people can feel if the members of the group they take part in collaborate with each other. By introducing instruments such as questionnaires or direct observation, evaluators can be aware of participants' satisfaction and have an indication about the collaboration that occurs among group members.

Cultural Impact. The technology may have completely different use in different contexts. The final dimension of group evaluation is how a groupware tool transforms: the way the members of a group work, the way the group work as an unit, what are the expected and unexpected ways the tool is used by the group and also how the organization or group incorporate or recognize the new tool introduced into their working culture.

Based on our previous evaluations, we observed that these four dimensions have a close relationship with each other in terms of their evaluation. For instance, depending on the group characteristics (its context), the reaction of using a specific tool can be quite different. Groups that are highly committed to an activity may try to overcome any usability problems that exist in the supporting tool. If the majority of people working in a group consist of optimistic persons, there is a tendency for the tool being used to have a positive cultural impact. If a tool has too many usability problems, collaboration may be completely compromised. If a high level of collaboration is achieved through the use of a groupware tool, the cultural impacts can be of greater dimension.

Additionally, evaluations should consider that the cultural impact (positive or negative) of the use of a groupware tool might change the group context (aims, expectative, attitudes), the way it is used and how people collaborate with each other.

4 The CSCW Lab

The CSCW Lab is an attempt to define a laboratory for groupware evaluation. By a laboratory we mean an environment where a methodology, including guidelines and instruments, are available for conducting groupware evaluation. Groupware evaluation involves a great amount of efforts. The planning, design, accomplishment and replication of an evaluation are costly activities. The design of an evaluation is an activity that should be carefully performed in order to guarantee that the results and measures are relevant for interpretation.

It is important that some evaluations should be conducted previously from real evaluations with the groups or organization where the tool will be used [1]. This first

evaluation stage – that we call “pilot-evaluation” – is important to tune the evaluation design and to outline the first general impressions, problems and benefits of the artifact been evaluated. The results are also relevant for identifying major usability problems in the developed tools before putting it to work in real settings.

The main objective of the CSCW Lab is to support the design and the conduction of pilot-evaluations. Each dimension described in the previous section can be considered as the subsequent steps of a method for conducting groupware pilot-evaluations:

Group context. The characterization of the groups to be observed is the first issue to be discussed. What are the essential group characteristics that must be collected before conducting the evaluation and that will be relevant to further understand the evaluation results? What are the guidelines for describing a group? How can we evaluate that a specific group is appropriate for our study? What are the group expectations with the use of the groupware tool? How these characteristics change from one domain to another?

Group context characterization is a dimension that relies heavily on psychology, sociology and ethnographic studies. It is an objective of CSCW Lab to work collaboratively with researchers from these domains in order to define techniques and instruments for characterizing group contexts.

Usability. Usability is the next dimension to be evaluated. Groupware usability is being widely discussed and some approaches and techniques for its evaluation have been proposed [15][16]. Usability evaluation techniques for single-user applications have been adapted to the evaluation of collaborative applications. In CSCW Lab, we aim at collecting, using and evaluating these suggested techniques as instruments for assessing the usability of the applications.

Usability has a strong relation to interface design and research. Some initial collaboration with other research groups [14] in this area has been settled to combine investigations on this theme.

Level of Collaboration. When we design and construct a groupware tool, our aim is to support and to promote collaboration. If we cannot determine if this collaboration really occurred, the tool loses its relevance and cannot be validated. Although we may not disregard the other dimensions of groupware evaluation, since they influence each other, collaboration evaluation is the main focus of groupware research.

To address this issue in the CSCW Lab, we define what we call Collaboration Maturity Models. Similarly to what has been defined in software engineering domain [18], we believe that it is possible to describe a set of collaboration levels a group can achieve and what are the characteristics of each level. Two models for evaluating the level of collaboration have already been developed in the context of the research group [19][20] providing resources for establishing other models for other domains.

Cultural Impact. After characterizing the group, evaluating the tool usability and the level of collaboration achieved by group members, we are able to evaluate its cultural impact. In CSCW Lab, we assume that cultural impact evaluation must consider the following levels: the individual level, the group level and the organizational level. In all these levels, it is possible to measure impressions, perceptions, satisfaction, commitment, learning and change of attitudes among other impact indicators.

Another objective of the CSCW Lab is to collect, organize and store data about the design and results of the evaluations conducted in the Lab. This data can be used as a source for analyzing the evaluation results as also to reuse the evaluation design specifications.

5 Experiences in the CSCW Lab

In this section we described the first set for experiments conducted in the CSCW Lab. These experiments were conducted as an attempt to validate the products of two doctoral research works of CHORD group in the domains of software development and collaborative learning. Our objective in this section is to show how the dimensions of CSCW Lab can be followed as evaluation steps according to each specific evaluation objective.

PIEnvironment evaluations. One group of evaluations was accomplished in the context of the use of workflow systems and awareness mechanisms for software process learning and improvement. An environment, named PIEnvironment, was built as an extension of a commercial workflow system in order to provide information about the collaboration that occurred within a defined process being executed through the workflow system. A description of the environment and its complete experiment design and results can be found in [21].

The environment was conceived in order to help software process participants to: follow and perform their tasks, be aware of the process they execute; learn about the process, and participate suggesting process improvement opportunities. The basic hypothesis of PIEnvironment was that process visualization and learning could change the culture of software development, enforcing the collaboration that exists within it.

The attempt of the pilot-evaluations, in this case, was to observe the potential of PIEnvironment in providing awareness, knowledge and consciousness to process participants about their work, and whether they felt satisfied and in favor of the idea of using a defined process.

In order to accomplish this objective of PIEnvironment evaluations, each of the dimensions of the CSCW Lab were defined:

- Group context: to evaluate the levels of awareness, consciousness and knowledge about the process they execute, it should be important to consider, for instance: the previous experiences of the participants with software development, their knowledge and practice about using defined processes, their practice and satisfaction with process definition and improvement initiatives and so on. Subjective measures about group context should also consider, for instance, the group commitment to the task and to the evaluation process.
- Usability: usability was not the main focus of PIEnvironment evaluations at this time. We addressed this dimension by allowing participants to report any problems and causes of dissatisfaction they encountered during process enactment, reflecting interface and/or usability troubles.
- Level of collaboration: PIEnvironment was conceived to make collaboration more visible and, consequently, to enforce it. To evaluate if the environment was

able to turn participants aware of the collaboration they took part was our first attempt at this time. Measuring how collaboration was enforced within the group was subjective, using the questionnaires to let participants report their feelings about how collaboration occurred and if they felt satisfied with it.

- Cultural impact: this could be considered the main objective of PIEnvironment evaluations – to evaluate how the awareness of the process and the collaboration provided by the environment helped participants to be more receptive to idea of following defined processes. To measure this level of individual cultural impact, we defined subjective questions in the questionnaire where participants could report if they felt satisfied with the overall process enactment and about how they suppose to use defined processes in the future.

COPLE evaluations. Another group of evaluations aimed at validating a conceptual model for collaborative project-based learning [20]. For this purpose, COPLE (Cooperative Project-Based Learning Environment) was implemented. This educational groupware presents an approach where teachers and students should define their work process, it means, the stages, the tasks and relationship among them, identifying the best ways to interact and to integrate individual solutions in order to accomplish a collaborative project. Our hypothesis is that collaboration enhances learning, thus it is necessary to enhance collaboration. Therefore, the goal was to observe if the process design helped to stimulate collaboration within the group.

A methodological set, including a series of criteria was defined to evaluate collaboration. In the sense of the four dimensions stated in the CSCW Lab Project, we can discuss the following subjects:

- Group context: It was very important to understand the personal characteristics of the public who used the environment, since we were in real case situation, and the cultural issues could interfere in the results obtained. We had two groups, working for a pos-graduate course, thus the background of each participant was also relevant in terms of the contributions to the final product made by the group. We observed that the personal commitment and availability of each member are the most determinant factors to the success of the project.
- Usability: Although we recognize this dimension to be a fundamental basis for any groupware evaluation, we did not attained very much on it this time. In the questionnaires we used to inquire students about their work, we just tried to raise the HCI problems that could possibly had influence on the way students developed their projects.
- Level of Collaboration: Evaluating the level of collaboration was the focus of this research. Therefore, we defined a set of criteria concerned to the following issues: Communication, Contributions for collective knowledge building, Coordination and Awareness. These criteria combined quantitative and qualitative measure units due to the nature of the collaborative process.
- Cultural Impact: In this case, to measure cultural impacts means to verify if the kind of work proposed brings a new perspective on the way apprentices and teachers view learning-teaching process. It would be necessary to go on using it in real situations with diverse teachers and students' groups, while reflecting the observations made by each study in new functionalities for the COPLE environment. That is what we intend to do within CSCW Lab project due to our goal to establish a methodological approach for groupware evaluations.

6 Conclusions

This paper presented our ideas concerning the CSCW Lab. The CSCW Lab is an initiative to organize and address the main issue of evaluating the groupware products developed within a research group. The paper identifies from the literature what are the main dimensions of groupware evaluation and how they are addressed in the Lab as steps for an evaluation method. Additionally, some experiments already conducted following CSCW Lab method are described.

From the evaluations conducted within CSCW Lab, we can conclude that the first step for designing groupware evaluations is to properly identify which of the dimensions are to be evaluated during the experiments depending on the research hypothesis. Next, each dimension must be defined in terms of the necessary variables to be set and how to measure them using specific techniques. To accomplish this definition, the CSCW Lab aim at providing guidelines, variables, models and instruments specific for groupware evaluation

As we can observe from the examples of evaluations conducted in CSCW Lab, they had very different targets and consequently, different aims and dimensions definitions. Another goal of the CSCW Lab is to define guidelines for helping researchers to map evaluation objectives to evaluation variables and techniques.

Other set of evaluations are now being planned in the context of our research group following the CSCW Lab method. Each of them will help us to continuously detail the method steps and techniques.

Finally, there are special issues that we would like to deeply address in CSCW Lab. Among them is the dimension of the collaboration level achieved using groupware by the definition of the collaboration maturity models. Additionally, we intuitively observed that there are strong relations between group commitment, tool usability and the level of collaboration achieved. Thus, we aim to continue to study this special relation in our experiments within CSCW Lab.

References

- [1] Pinelle, D. 2000 *A Survey of Groupware Evaluations in CSCW Proceedings*. Research Report 00-1, Computer Science Department, University of Saskatchewan.
- [2] Pinelle, D., and Gutwin, C., 2000, "A Review of Groupware Evaluations", *Proceedings of Ninth IEEE WETICE 2000 Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, Gaithersburg, Maryland, June.
- [3] Knutilla, A.J., Steves, M.P., Allen, R.H., 2000, *Workshop on Evaluating Collaborative Enterprises 2000 – Workshop Report*, as part of the 9th Int. Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, June, National Institute of Standards and Technology, USA. (report available at: www.mel.nist.gov/msidlibrary/doc/steves00a.pdf)
- [4] Steves, M.P., Allen, R.H., 2001, *Evaluating Collaborative Enterprises – A Workshop Report*, as part of the 10th Int. Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, June, MIT, Cambridge, MA, USA. (available at: www.mel.nist.gov/msidlibrary/doc/steves01b.pdf)

- [5] McGrath, J.E., "Methodology Matters: Doing Research in the Behavioral and Social Sciences". In: *Readings in Human-Computer Interaction: Toward the Year 2000*, 2nd Ed. Baeker, R.M., Grudin, J., Buxton, W.A.S., Greenberg, S. (eds.), Morgan Kaufman Publishers, San Francisco, 1995, 152-169.
- [6] Ramage, M., Twidale, M., Sommerville, I., *Evaluation of Cooperative Systems Project*, (<http://www.comp.lancs.ac.uk/computing/research/cseg/projects/evaluation/>)
- [7] Dennis, A., Valacich, J.S., 2001, "Conducting Research in Information Systems", *Communications of the Association for Information Systems*, Vol. 7, Article 5, July.
- [8] Baker, K., Greenberg, S., Gutwin, S., 2001, "Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration". In: *Proceedings of the 8th IFIP Working Conference on Engineering for Human-Computer Interaction (EHCI'01)*, Toronto, Canada.
- [9] Randall, D., Twidale, M., Bentley, R., 1996, "Dealing with Uncertainty- Perspectives on the Evaluation Process". In: Thomas, P.J. (ed.), *CSCW Requirements and Evaluation*, Springer.
- [10] Baeza-Yates, R., Pino, J.A., 1997, "A First Step to Formally Evaluate Collaborative Work". In *Proceedings of Group 97*, Phoenix, USA.
- [11] Monk, A., McCarthy, J., Watts, L., Daly-Jones, O., 1996, "Measures of Process". In: Thomas, P.J. (ed.), *CSCW Requirements and Evaluation*, Berlin, Springer.
- [12] Steves, M.P., Allen, R.H. (2001) *Evaluating Collaborative Enterprises – A Workshop Report*. WETICE'2001.
- [13] CHORD Group – www.chord.nce.ufrj.br
- [14] Semiotic Engineering Research Group (SERG) – www.serg.inf.puc-rio.br
- [15] Baker, K., Greenberg, S., Gutwin, C., 2001, "Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration", *Proc. of the 8th IFIP Working Conference on Engineering for Human-Computer Interaction*, ECHI'01, May 11-13, Toronto, Canada.
- [16] Prates, R.O, Souza, C.S., Assis, P.S., "Categorizing communicability evaluation breakdowns in groupware applications", *CHI-SA 2001*, Pretoria, South Africa.
- [17] Borges, M.R.S., Pino, J.A., 1999, "Awareness Mechanisms for Coordination in Asynchronous CSCW". In: *Proceedings of the 9th Workshop on Information Technology and Systems (WITS'99)*, Charlotte, North Carolina, EUA, Dec.
- [18] Paulk, M.C. et al, 1993, *Capability Maturity Model for Software, version 1.1*. CMU/SEI-93-TR-24, Carnegie Mellon University, Software Engineering Institute, Pittsburgh, Pennsylvania, Feb.
- [19] Santoro, F.M. , 2001, *A Cooperation Model for Project-Based Learning*. Doctoral Thesis. COPPE Sistemas, UFRJ, Rio de Janeiro, Brazil. (in Portuguese).
- [20] Borges, M.R.S., Pino, J.A., 2000. "Information Technology in Latin America: Two Decades of Collaboration", *Information Technology for Development*, V 9 N 3-4, p. 189-204
- [21] Araujo, R.M., Borges, M.R.S., 2001. Extending the Software Process Culture – An Approach Based on Groupware and Workflow. In: *Proceedings of the Third International Conference on Product Focused Software Process Improvement PROFES'2001*, Kaiserslautern, Germany, September, p. 297-310.