# LABORATORY SIMULATION METHODS
# FOR STUDYING COMPLEX COLLABORATIVE TASKS

William D. Humphries, Dennis Neale, and D. Scott McCrickard
Virginia Polytechnic Institute and State University
Blacksburg, VA

John M. Carroll
Penn State University
State College, PA

This paper describes a new laboratory model developed for studying complex CSCW phenomena. In our prior work a more ecological laboratory approach was developed to study activity awareness issues derived directly from earlier fieldwork. In the first study participants worked on a simulated long-term project with a confederate who introduced collaborative breakdown scenarios. Although the study produced many realistic behaviors, findings indicated that there was room for improving the model by making the simulation more ecological and engaging for participants. This paper examines the results of a follow-up study conducted with the goal of improving the laboratory model. The follow-up study varied two important elements of simulation from the first study. A real long-term project was used, and the confederate was replaced with another participant. Differences in results from the two studies indicate how the changes made to the model impacted its effectiveness.

## INTRODUCTION

To effectively understand the complex set of factors impacting computer-supported cooperative work (CSCW) systems, field methods used over extended periods of time have been necessary (Grudin, 1988). These types of studies can be contrasted with short-term laboratory research that uses a range of usability engineering methods developed to understand single-user systems. Until recently, there has been no inexpensive, short-term, lab-based option for assessing collaboration and multi-user systems.

Recently, Baker et. al. (2002) have suggested using heuristic evaluation methods adapted from single-user research on CSCW systems. This method does provide an efficient and inexpensive option for formative evaluation of groupware systems, but in considering the potential costs of this approach, ecological validity may be compromised (Cockton and Woolrych, 2002).

We recently conducted a study with the goal of validating a new laboratory model for observing complex CSCW phenomena. The goal of the research was to maintain ecological validity while capitalizing on the benefits of a laboratory study, namely short time span and minimal use of resources. Field studies with the goal of observing CSCW phenomena can involve months of frequent site visits. The laboratory model proposed here has a timeframe of less than a month. This laboratory model was used to observe activity awareness – a complex problem representative of difficult problems to research in CSCW systems (Dourish and Bellotti, 1992).

Activity awareness builds on earlier research into social (who is present) and action (what is he/she doing) awareness. It refers to the ability of the user to have and maintain “awareness of the overall situation, the social expectations and dependencies within their group, and their shared task goals and status” (Carroll et al, 2003). Activity awareness is knowing what has happened, what is happening, and what will likely happen in the future over extended periods of time. This means that in order to maintain awareness, the user must understand where they and their coworkers fit in to the long-term plans of the project. When activity awareness is not maintained, efforts at planning and completing the activity will lead to breakdowns as inconsistencies in different users’ interpretations of shared plans and goals become apparent. A well-designed system should keep the user in this state of awareness as some parts of the project plan undergo revision or as the project’s direction evolves and changes over time.

## A NEW LABORATORY MODEL

The model used in the first laboratory study combines the control and precision of existing laboratory experiments with simulated attributes designed to increase ecologic validity. Data from a field study conducted on activity awareness in middle-school students working on shared science projects (Carroll et al, 2003) was used to help develop a method that allowed authentic collaborative situations to be simulated and manipulated in a laboratory setting (Convertino et. al., 2004).

The new model has three aspects that set it apart from traditional laboratory studies. The first of these is the use of authentic tasks and collaborative situations, as observed and modeled directly from the fieldwork. The second key aspect involved the use of a confederate who manipulated the collaboration in order to introduce certain scenarios that were observed in the field study. Finally, experiments are conducted over multiple collaborative sessions. This is necessary due to the impact of changes that occur over time on activity awareness.

Six middle school students participated in four weekly one-hour laboratory sessions that required them to collaborate through a CSCW system with a confederate (acting as another student) to complete an environmental science project composed of several activities. A commercial CSCW package known as Groove was used for the study (Groove*, 2002)*. The Groove client consists of a planning tool, a chat area, a user list, and a tabbed workspace.

The students in the laboratory study did not actually complete much of the work involved in the project. This aspect of the study was completely simulated. Instead they coordinated, collaborated, and planned how they would complete their work, and the work was simulated for them in between sessions. This allowed the lab study to remain shorter in duration than field studies. Individual work was simulated, while the collaborative components of the project, which the experimenter had an interest in, were maintained.

The participant and confederate were kept in separate rooms. They were monitored by a video camera, and spoke to the experimenter through a microphone. The experimenter was able to speak back to one or both of them. The experimenter also had monitors displaying the two participants' individual views of the workspace, so computer activity could be monitored in real time. The experimenter was also logged into the client workspace himself. This setting is appropriate, as it provides the experimenter with all necessary observatory capabilities while keeping the two participants remotely located. The experimenter was able to monitor or interact with either student without influencing the other (Convertino, 2004).

## DATA ANALYSIS AND RESULTS OF FIRST STUDY

Multiple data collection methods were used throughout the experiment. Video-cameras and screen-capture software recorded the participant's interactions with the computer. All interactions with the Groove system, such as chat and changes made to the workspace, were logged on a session-by-session basis. A Likert-type questionnaire was given to the participants upon completion of the project, and they were then asked to provide qualitative feedback on their answers to the questionnaire.

The data was analyzed using three different techniques. Analysis by scenario was used to assess the participant's activity awareness with respect to changes made in the collaborative scenarios that were introduced. Participants were evaluated fully aware if they spontaneously noticed inconsistencies, partially aware if they became aware of changes after being prompted by the confederate or experimenter, or unaware if they failed to notice changes. The second form of analysis used was evaluation of the Likert-type scales. The follow-up interview helped to contextualize rating scale responses.

Finally, the data was analyzed using breakdown analysis. A breakdown occurs when the expectations of one participant do not match with the action of another. Breakdowns were analyzed by the evaluation framework used in earlier field work (Carroll et al., 2003). Breakdowns were grouped into one of four categories based upon their cause.

1. Situational (environment)
2. Group/User (users and their roles)
3. Task (plans)
4. Tool (software and workspace).

Situational breakdowns refer to breakdowns that do not fit in any of the other three categories. These types of breakdowns can be influenced by the environment. For example, if a student believed that the laboratory setup was allowing him to interact with his partner through the microphone, a situational breakdown could occur.

Group/User breakdowns involve communication breakdowns between users. If a miscommunication leads one user to misinterpret his/her role for the current meeting, as explained by another user, a Group/User breakdown is occurring. These types of breakdowns can often highlight activity awareness difficulties.

Task breakdowns occur when one or more users do not share the same vision regarding one another's plans. This can occur due to a lack of communication, a poorly designed project plan, etc. These breakdowns are frequently associated with activity awareness breakdowns.

Tool breakdowns refer to difficulties experienced by one or more users when working with the software provided for them. A user who did not understand that the BRIDGE client (used in the second field study) automatically saves documents might have spent significant time attempting to locate the 'Save' button. These breakdowns can sometimes lead to breakdowns in activity awareness, but in many cases they do not.

The results from this experiment were compared to those gathered in the field study to determine their validity. Collaboration was clearly evident as users were interested in the activities and motivated to work. The users reacted to scenarios much the same way that users in the field study did when those scenarios evolved naturally. The questionnaire showed that most users felt that they were collaborating during the experiment, and the follow-up interviews revealed that several users felt informal communication with the confederate made collaboration seem more natural and informal.

The majority of the breakdowns observed (37%) were group breakdowns, where users failed to effectively communicate or understand the roles they and their partner played in the task. This is a well-established difficulty in computer-mediated communication. The loss of non-verbal communication and auditory cues makes it more difficult for users to maintain common ground.

Another third (32%) of the breakdowns found were determined by task factors. This category was most closely related to participants' activity awareness difficulties. Most of these breakdowns occurred due to the participants' inability to maintain a clear picture of the plan, progress made on the plan, and the time remaining to complete the project. The ability of the researchers to monitor and record activity awareness breakdowns reinforced the validity of the new laboratory model. Tool factors accounts for around one fourth (23%) of the breakdowns observed and situational factors caused only a small portion (7%).

## FOLLOW-UP LABORATORY STUDY

A second study was conducted to address issues raised in the first study, validating the new laboratory model in a more realistic setting. In an effort to increase ecological validity, the participants had to complete a real project for course credit. This ensured that the project would be as engaging as any the students worked on for their classes. Adding this level of realism to a lab study was expected to keep the students fully engaged for the duration of the study. Work was no longer simulated, and the confederate was replaced by a second participant. Scenarios were not introduced; the users were expected to run into potentially problematic situations naturally. The informal communication likely to occur between two classmates was expected to enhance collaboration as well.

Undergraduate computer science students from Virginia Tech participated in this study, and the same laboratory setting was used, with four groups of two students meeting three times over the course of three weeks.

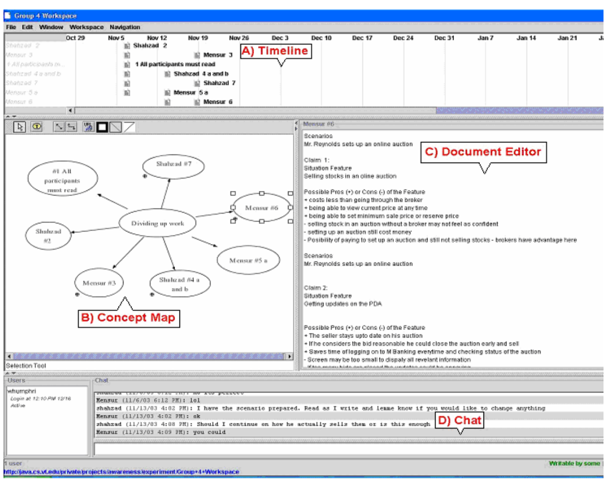


Fig. 1, BRIDGE Client Interface

The BRIDGE client, used previously in the large-scale field study, was used in the follow-up study (Ganoe, et. al., 2003). This was done to allow some comparison between the two based on results of the two studies. The BRIDGE client has five elements within its interface (Fig. 1). Along the top is the Timeline tool, which provides document histories for all documents in a workspace. On the left is the concept map tool, where concept maps are used to show relationships between the documents in the workspace. On the right is the document editor. Finally, there is a user list on the bottom left, showing all users logged in, and a chat tool on the bottom right.

## FINDINGS FROM FOLLOW-UP STUDY

Initial results of the second study seemed contradictory (Table 1). The participants' uniform interest in the assignment encouraged collaboration that was observed through the questionnaire. Comparison of the responses revealed that the users did not feel they were communicating and collaborating with their partners as well as the users in the first study. This can be seen on items four, five, six, eight, and ten of the questionnaire, all of which average .7 to 1.2 lower on the second study than on the first. Initially it was believed that this simply reflected the more realistic experimental setting, where two students interacted with each other instead of with a generally agreeable confederate. These findings indicated that collaboration had not gone as well in the second study, and so more breakdowns were expected than in the first study.

Table 1. Likert Survey Results. 1=Strongly Disagree, 4=Neither Agree Nor Disagree, 7=Strongly Agree

| Likert-type Questions (7-point scale agree/disagree) | Study 1 | Study 2 | **Ave. Diff. from 1 to 2** |
|---|---|---|---|
| | Ave. *S.D.* | Ave. *S.D.* | |
| 2. I could tell what my partner was doing while we were collaborating online. | 6.2 *0.8* | 6.1 *0.9* | **-0.1** |
| 3. My partner and I planned adequately. | 5.8 *0.8* | 5.8 *1.4* | **0** |
| 4. I always knew what my partner was going to work on over the week. | 6.3 *0.8* | 5.6 *2.4* | **-0.7** |
| 5. My partner and I communicated well with each other. | 6.5 *0.5* | 5.8 *1.6* | **-0.7** |
| 6. I became more aware of my partner's plans over time. | 6.2 *1.2* | 5 *1.0* | **-1.2** |
| 7. It was always clear what my partner was going to do. | 5.7 *0.5* | 6.3 *0.5* | **+0.6** |
| 8. I enjoyed collaborating with a partner online. | 6.5 *0.5* | 5.6 *2.3* | **-0.9** |
| 9. It was easy to find what my partner had worked on within the collaborative space. | 5.0 *1.5* | 4.8 *2.4* | **-0.2** |
| 10. My partner contributed equally to this project. | 7.0 *0.0* | 6.0 *1.2* | **-1.0** |
| 11. My partner collaborated with me to complete the project. | 6.5 *0.5* | 6.5 *0.8* | **0** |

The same method of breakdown analysis used in the first study was used here. However, significantly fewer breakdowns occurred in the second study than in the first. Each experiment involved 24 hours of lab time. Within that time, 108 breakdowns were observed during the first study. Less than 20 breakdowns were observed throughout the second study.

Most of the breakdowns in the first study were group breakdowns dealing with communication and task factors (Table 2). In the second study, the overwhelming majority of breakdowns (57%) were tool breakdowns dealing with the BRIDGE client's timeline or document editor. Many of these breakdowns occurred in the first week, when users were still familiarizing themselves with the BRIDGE client interface. Task-related difficulties accounted for 21% of the breakdowns in the second study. Again, this type of breakdown generally indicated activity awareness problems.

Table 2. Breakdown Analysis Results.

| Breakdown Type | Study 1 Percentage | Study 2 Percentage |
|---|---|---|
| Task | 32% | 21% |
| Tool | 23% | 57% |
| User | 37% | 14% |
| Situational | 7% | 7% |

It seemed contradictory that groups who felt they were not collaborating as well with their partner would run into fewer breakdowns. Even Group 3, whose two members stated in separate interviews that they had not planned adequately, completed the assignment with minimal confusion. One conclusion that can be drawn from all of these factors was that the assignment was not of sufficient complexity to force common collaborative scenarios, such as plan revision and difficulties with interdependencies in the plan, to occur.

The assignment plan given to students for the first lab study instructed them to carry out two related science experiments with their partner, the confederate. Both experiments involved multiple steps that built upon each other in such a way that the task could not be continued if a particular step was not carried out properly. The second lab study's assignment was similar to assignments the participants had been completing for a class they were enrolled in. Although the assignment was developed with the idea of building in task dependencies, the participants' experience may have allowed them to continue through the experiment using knowledge gained from prior assignments rather than from full understanding of all portions of the assignment given to them during the study. A user who realized that the portion of the assignment s/he was working on was dependant upon information gained from a part of the assignment that was not yet completed could use knowledge gained from outside of the experimental setting to complete that portion of the assignment satisfactorily.

## DESIGN RECOMMENDATIONS FOR SIMULATED LABORATORY STUDIES

The results of the two studies have provided us with insight into the best way to design laboratory studies for the observation of complex CSCW phenomena. Recommendations based upon this knowledge are provided below.

1. A confederate is helpful in producing effective simulated collaborative breakdowns. This experiment demonstrated that removing the confederate from the experiment seriously hinders the ability of the experimenter to observe breakdowns because fewer are likely to occur. This issue outweighs the realism gained from using two real participants.
2. Encouraging informal communication between the confederate and participant is important for the participant to develop a relationship with the confederate that underlies the collaborative process. The participant will have to communicate more information than she would in a non-CSCW environment due to the loss of common ground. Because of this, it is important that s/he be comfortable communicating with her partner.
3. The experimental sessions should extend over an adequate amount of time. Multiple experimental sessions are required to raise task coordination issues. Many task breakdowns occur as the result of a user's inability to keep track of a plan as it is changed over time. It is imperative that multiple sessions be used, so that if a user's activity awareness is not adequate there will be time for a situation to arise where this can be observed.
4. The introduction of collaborative breakdown scenarios must extend over time as well. Breakdowns introduced by the experimenter should involve components that extend over multiple sessions. This makes them more realistic and allows the researcher more opportunities for discovering their effects on users.
5. The participant must be actively engaged in the completion of tasks for the simulation to be realistic. This simulated component of the method must be carefully balanced with actual task completion.
6. The tasks need to have enough complexity to insure that interdependencies exist between subtasks. This will require more extensive planning and plan revisions that bring to light collaboration issues relevant to groupware.
7. The experimenter and confederate must be able to communicate independently of the participant during the simulation to insure adequate experimental control. There may be situations where the confederate is required to alter their script due to the behavior of the participant. The experimenter has greater observatory capabilities with regards to the participant, and so s/he will be in a better position to determine when a change on the confederate's part is required. It is essential that the experimenter be able to communicate instructions of this nature to the confederate without upsetting the realism of the experiment for the participant.

## CONCLUSION

The follow-up study highlighted the strengths of the laboratory model developed in the first study. The studies revealed that controlled situations in the laboratory under the right conditions do effectively reflect the complex dynamics found in actual collaborative work contexts, especially if we extend the experimental sessions over time. A lack of ample awareness breakdowns in the follow-up study, especially with the group that failed to plan sufficiently, indicates that the assignment given to the participants needs to be sufficiently complex to force the natural development of realistic breakdown events.

These conclusions and the design recommendations that have resulted from the two laboratory experiments should allow future CSCW evaluation efforts to be performed in the laboratory. This will provide experimenters with reduced cost and time factors, while the new techniques provide realism and ecological validity that has been previously unavailable in laboratory studies involving complex phenomena.

Future work should examine the relationship between assignment complexity and experimental realism. The role of the confederate also needs to be better understood for this method to be effective in a variety of contexts. Different elements of future assignments should be simulated to determine which elements are not essential to keeping participants engaged and which elements do not reveal the quality of collaboration occurring. This methodology should be tested out on more varying user populations as well.

## ACKNOWLEDGEMENTS

This research is supported by the US National Science Foundation (NSF) Information Technology Research program (IIS 0113264).

We thank Gregorio Convertino for his prior work on the lab methodology discussed herein.

## REFERENCES

Baker, K. (2002). Heuristic Evaluation of shared Workspace Groupware based on the Mechanics of Collaboration. *MSc Thesis, Department of Computer Science*, University of Calgary, Calgary, Alberta, Canada.

Carroll, J. M., Neale, D. C., Isenhour, P. L., Rosson, M. B., and McCrickard, D. S. (2003). Notification and awareness: synchronizing task-oriented collaborative activity. *International Journal of Human-Computer Studies*, 58, 605-632.

Clark, H. H (1996). *Using Language.* New York: Cambridge University Press.

Cockton, G., & Woolrych, A. (2002). Sale must end: Should discount methods be cleared off HCI's shelves? *Interactions, 9,* 5, 13-18.

Convertino, G., Neale, D. C., Carroll, J. M., Rosson, M. B., L. Hobby (2004). A laboratory model for studying activity awareness. (In Press). *In Proceedings of NordiCHI 2004*. Tampere (Finland): Association for Computing Machinery.

Dourish, P., and Bellotti, V. (1992). Awareness and coordination in shared workspaces. *In Proceedings of the ACM CSCW '92 Conference on Computer Supported Cooperative Work* (pp. 107-113). New York: Association for Computing Machinery.

Ganoe, C. H., Somervell, J. P., Neale, D. C., Isenhour, P. L., Carroll, J. M., Rosson, M. B., and McCrickard, D. S. (2003). Classroom BRIDGE: Using Collaborative Public and Desktop Timelines to Support Activity Awareness. *In Proceedings of the ACM '03 Symposium on User Interface Software and Technology* (pp. 21-30). New York: Association for Computing Machinery

Groove Networks (2003). *Groove Networks Inc.* Retrieved November 2003, from http://www.Groove.net.

Grudin, J. (1988). Why groupware applications fail: Problems in design and evaluation. *Office: Technology and People, 4(3),* 245-264.