

# Workshop on Evaluating Collaborative Enterprises - Workshop Report

Amy J. Knutilla, Michelle P. Steves and Robert H. Allen

Knutilla Technologies, National Institute of Standards and Technology, and University of Maryland, College Park  
{knutilla, msteves, rhallen} @nist.gov

## Abstract

The workshop, Evaluating Collaborative Enterprises,<sup>1</sup> explored the issues surrounding the evaluation of collaborative systems including methods and tools for evaluating collaborative software as well as application-specific evaluation experiences. This paper describes the workshop's mission, summarizes the workshop papers and discussions, and concludes with future research needs.

**Keywords:** CSCW, groupware evaluation, collaborative enterprise evaluation

## Acknowledgments

The following people participated in this workshop and contributed to the contents of this report: Janet Allen, Evan Atkinson, Prasun Dewan, Carl Gutwin, David Hallenbeck, Steven MacGregor, Emile Morse, David Pinelle, Catherine Plaisant, Jean Scholtz, monica schraefel, Jonathan Sevy, Eswaran Subrahmanian and Angran Xiao.

In addition to the editors, the following people participated on the program committee. We gratefully acknowledge their time and energies: Jill Drury, The MITRE Corporation; Tom Finholt, University of Michigan; Saul Greenberg, University of Calgary; Carl Gutwin, University of Saskatchewan; Doug Johnson, Hewlett Packard; Srinivas Nidamarthi, Cambridge University; Leysia Palen, University of Colorado; Jennifer Preece, University of Maryland, Baltimore County; Jean Scholtz, U.S. Defense Advanced Research Projects Agency; Eswaran Subrahmanian, Carnegie Mellon University; Marian Williams, University of Massachusetts Lowell.

## Introduction

Effective collaboration involves people sharing information. Web technology has enabled basic communication infrastructures needed for collaborative applications to facilitate people working together over the Web. However, using the basic infrastructures has proven much more difficult in enabling effective collaboration within and among enterprises than the hype would suggest. Although higher technological layers are being developed and a general, deeper understanding of collaboration is slowly evolving, progress on efficient and effective Web-based collaboration lags behind these other gains. 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 the effectiveness of collaboration. We believe effective evaluation and appropriate standardization are mechanisms that will facilitate progress in Web-based<sup>2</sup> collaboration.

This workshop continues themes emanating from previous WETICE workshops held over the last several years. In general, these workshops addressed a number of issues, including:

- How can Web techniques be used to achieve or to improve collaboration within or among organizations?
- Can the Web serve as an infrastructure for both developing and implementing business applications in globally distributed, collaborative business environments?
- Can Web-based software answer the challenges that globally operating companies are facing?

This session continues in the vein set forth in previous workshops and extends it by exploring issues surrounding the evaluation of collaborative systems: When and how do we achieve collaboration successfully? When do we not? Why? With the increasing connectivity of the Internet and the increasingly distributed nature of organizations, groupware<sup>3</sup> is an increasingly attractive method for

---

1. For the purpose of this workshop, collaborative enterprises are systems of people and software technology involving group work, where more than one person is moving toward a common goal.

2. The term Web-based should not be interpreted as referring only to World Wide Web technology, i.e., technology requiring Web browsers for use. This workshop addressed the evaluation of distributed technologies for collaboration.

collaboration. Effective evaluation of collaborative systems holds one key to improving groupware effectiveness. This workshop explored methods and tools for evaluating collaborative systems as well as application-specific evaluation experiences.

The primary goal of this workshop was to facilitate the progress of distributed collaboration through improved development and use of methods for the evaluation of collaborative enterprises. The workshop also highlighted the value of bringing together people who are addressing the unique and challenging needs of collaborative enterprise evaluation. Plans for further information exchange among researchers, software developers and users of collaborative enterprises to improve the development and use of methods for the evaluation of collaborative enterprises were made.

In this report, we present a summary of the nine papers presented, which fell roughly into two categories: Collaborative System Evaluation Methods and Tools and Application-Specific Evaluation Experiences. These papers discussed evaluation methods and approaches (established and visionary), tools for evaluation (existing and prototype) and research needs. This report summarizes the papers and discussions and concludes with future research needs.

## Summary of Papers

This section briefly summarizes the presentations of papers and highlights some of the comments and issues generated by workshop participants as a result of the individual presentations.

### Collaboration System Evaluation Methods and Tools

*A Review of Groupware Evaluations*, by David Pinelle and Carl Gutwin, provides a review of all papers from the Association of Computing Machinery (ACM) Computer-Supported Cooperative Work (CSCW) conferences (1992-1998) that introduced or evaluated a groupware system. The main findings from the review are that almost one-third of the groupware systems were not evaluated in any formal way, that only about one-quarter of the evaluations involved a practical setting and that a wide variety of evaluation techniques are in use. The main conclusions from the review are that more attention must be paid to evaluating groupware systems, and that there is room for additional evaluation techniques that are simple and low in cost. This paper provides a good classification framework for collaborative evaluations that was found useful for further workshop discussions:

- Type of evaluation
- Placement of the evaluation in the software development cycle
- Evaluation techniques
- Focus of evaluation

*The Mechanics of Collaboration: Developing Low Cost Usability in Shared Workspaces*, by Carl Gutwin and Saul Greenberg, introduces a conceptual framework that addresses the mechanics of collaboration for shared-workspace groupware: the low level actions and interactions that must be carried out to complete a task in a shared manner. These include communication, coordination, planning, monitoring, assistance and protection. The framework also includes three general measures of these mechanics: effectiveness, efficiency and satisfaction. The underlying idea of the framework is that some usability problems in groupware systems are not inherently tied to the social context in which the system is used, but rather are a result of poor support for the basic activities of collaborative work in shared spaces. Gutwin believes that existing, low-cost, evaluation methods— heuristic evaluation, walkthroughs, user observations and questionnaires—can be modified to include this framework in a way that helps a groupware evaluator uncover these usability problems.

During post-presentation discussions for this paper, another group behavior was proposed: people in group situations quickly switch between activities, going back and forth quickly. Several general challenges in groupware evaluation were also discussed. For example, it is difficult to ascertain satisfaction levels for specific behaviors, it is difficult to identify precise measures and it can be difficult differentiating cause and effect. However, it was proposed that if all the correct questions were asked, many deficiencies would be uncovered.

*Techniques for Evaluating Collaboration Toolkits*, by Prasun Dewan, states that the most reliable approach to evaluating collaboration toolkits is to carry out field studies. This approach has the problem that it requires each toolkit project to create a complete, working system. More importantly, this approach does not allow a project to converge incrementally towards a complete solution. The paper identifies several other techniques for evaluating collaboration toolkits, including inspecting the design to see if requirements have been met, simulating other systems, implementing complete solutions to standard problems and performing self and lab studies. This paper discusses these techniques, points out their pros and cons, identifies the influential projects in which they have been used, and shows how they fit together. The presentation described a list of needs in collaboration toolkit evaluation, including the need to understand the value and cost of different evaluation approaches. It was pointed out that

---

3. Groupware is defined here as software and technology that enhances collaboration.

there are two aspects to toolkit evaluation: first, how to tell if the toolkit under evaluation can support a certain range of collaborative applications and secondly, how well these applications support collaborative tasks. The paper addressed the first aspect, leaving the second aspect for the workshop. The workshop participants also discussed the need for benchmark scenarios, applications and requirements to help developers and evaluators.

In *CollabLogger: A Tool for Visualizing Groups at Work* by Emile Morse and Micky Steves, a visual tool has been developed to support usability analyses of human-computer interaction of log data from a groupware system. The groupware was instrumented to log users' interactions with the system and each other. The CollabLogger addresses the problem of helping investigators analyze the volumes of log data that groupware tools can generate. Visual tools are powerful when large amounts of diverse data need to be explored.

The workshop participants discussed and concluded that there is real value in determining appropriate sets of events that groupware systems should log so that tools such as the CollabLogger can be employed to help evaluate groupware. Sets of events would be linked to characteristics of the groupware systems so that groupware system developers could easily determine which events to log. Further, a mapping of these events to the mechanics of collaboration identified by Gutwin and Greenberg in the previous paper was determined to be of value.

The paper, *An Evaluation and Comparison of the Industrial and Educational Usage of CSCW Within the Design Process*, by A. I. Thomson, S. P. MacGregor and W.J. Ion, describes the findings of three recently completed research projects conducted at the University of Strathclyde, Scotland. Each of the projects involved the implementation and usage of CSCW technology within the design process. Two of the projects described were education-based while the other was conducted in collaboration with industry and based on live industrial projects. The paper describes the methods adopted to evaluate the impact of the technology together with the findings of each of the projects in the form of implementation guidelines to promote effective and efficient technology adoption. Opportunities for cross-sector lessons were also presented.

In *Towards an Evaluation Methodology for the Development of Research-Oriented Virtual Communities* by Mark Chignell, Janet Ho and m. c. schraefel, a virtual community is defined as a group of people whose shared needs and/or interests are largely communicated within, and mediated by, Web-enabled interactions, e.g., within a website. In order to design effective virtual communities, methods are needed to evaluate the usability and effectiveness of collaborative environments and the degree

to which they function as virtual communities (VCs). In this paper the authors explore, through the discussion of a pilot study of a research-purposed VC, how to evaluate a VC infrastructure. Workshop participants posed the question, In what ways can we increase the effectiveness of collaboration system evaluation while reducing the cost?

## **Application-Specific Evaluation Experiences**

The paper, *Deploying and Evaluating Collaborative Studios* by Jonathan Sevy, Vera Zaychik, Thomas T. Hewett and William C. Regli, describes a vision of a collaborative engineering environment, the Collaborative Design Studio. It considers issues and questions relating to the effective design and evaluation of such systems, and describes proposed studies for evaluating and evolving the current environment. While preliminary development of the studio has proceeded using relatively informal input from professional design engineers, it is essential to incorporate direct evaluation by intended users of the system to ensure that it meets their needs accurately and fits into their work environment. Such information will be obtained through surveys and observation of participants as they complete specified tasks. The information will then be analyzed and used to guide further development of the system. Researchers also used archived communication and search (pattern) activity analysis during evaluation for tool development. One important issue raised as a result of this presentation is that long-term use of systems for field studies is very difficult to sell to industry.

In *A Method to Design Process Architecture in the Distributed Product Realization Environment* by Angran Xiao, Janet Allen, David Rosen and Farrokh Mistree, a distributed process structuring method (DPSM) was presented to design the process architectures in a distributed environment. With DPSM, a process is represented as a hierarchy by partitioning the process and reorganizing the activities. Four typical architecture templates are introduced: sequential, parallel, integrated and distributed. The process architectures are constructed by mapping the process representation into the templates. This method is a state-of-the-science method to structure a process in the distributed environment, while ensuring modularity and robustness of the architecture. Finally, an engineering case is investigated with this method and the distributed architecture is implemented using agent technology. The evaluation involved testing the validity of the dependencies within the context of a task and ranking the templates based on effectiveness relative to the context. Workshop participants discussed the issue that it is often difficult to determine what is really happening when things go wrong because of participant group dynamics.

*Evaluation Challenges for a Federation of Heterogeneous Information Providers: The Case of NASA's Earth Science Information Partnerships* by Catherine Plaisant and Anita Komlodi, presented an experiment to assess the ability of a group of widely heterogeneous, earth-science data or service providers to self organize and provide improved and cheaper access to an expanding, earth-science, user community. As it is organizing itself, the federation is mandated to set in place an evaluation methodology and collect metrics reflecting the health of the Federation and benefits to the public. The authors described the challenges of organizing self-evaluation of such a federated partnership and discussed the issues encountered during the metrics definition and early data collection phases. Some metrics used were changes over time of identified measures, for example: number of delivered data products, number of rendered services, number of Web page hits, number of promotional products, volume of available data and number of available services. One lesson learned was that metrics must be chosen carefully to reflect the goals of the organization, because resources were diverted to enhance the products and services with associated metrics leaving other offerings to languish. Another lesson learned was that while good metrics were difficult to identify, collect and assess, anecdotal information had significant impact in reports. Challenges identified were 1) identifying useful and obtainable metrics, 2) the relative unavailability of baseline data and 3) trying to use metrics gathered from very diverse entities within an enterprise to compare how those entities are performing and collaborating among themselves and with their respective constituencies.

## Summary of Discussions

Three main questions were addressed during the discussion portion of the workshop:

- What are the approaches to evaluating collaborative systems?
- What are the current problems with collaboration system evaluation?
- What are future research needs?

The remainder of this report summarizes the discussions and conclusions related to these issues.

### What are the approaches to evaluating collaborative systems?

One useful goal for this workshop was to identify the methods and techniques for evaluating collaborative systems and to discuss how those methods are best applied. The participants noted that it will also be important to ascertain when in the development cycle each method is

best applied, although it was acknowledged more research needs to be done in this area. In an attempt to frame the scope of the discussion, we first outlined a statement on context for evaluation of collaborative enterprises.

*The evaluation of a collaborative enterprise should be set in some context, e.g., a particular application or problem space. Effective evaluations do not take place out of context.*<sup>1</sup>

We then turned to what we know of where single-user system and collaborative system evaluation diverge. A major difference between single user system evaluation and collaborative system evaluation is that we are trying to do more than traditional usability testing when evaluating groupware. Testing must also consider the usefulness of the software in addressing the needs of the collaborative enterprise. Sufficient guidelines for understanding how software meets collaborative enterprise *needs* do not exist. (Note, that the workshop participants were being intentionally vague in the use of the term *needs*, it should be defined in future research.)

There are essentially two basic reasons people use groupware: they need to collaborate or they want to improve an existing collaboration. Some form of separation, either geographical or temporal, often drives the incorporation of groupware to facilitate collaboration. Evaluation provides an important method for understanding how to improve the groupware-supported collaborative experience. Given the need to evaluate a collaborative enterprise, there are several approaches that have been used, including exploratory methods, field experiments, laboratory experiments, field and case studies. Building on McGrath's work [4], Pinelle organizes these evaluation types based on the level of manipulation and the setting, see Figure 1. Further research needs to be

		Manipulation	
		Rigorous	Minimal/ None
Setting	Naturalistic	Field Experiment	Field Study, Case Study
	Controlled	Laboratory Experiment	Exploratory

Figure 1: An organization of evaluation types

1. It was also noted that the context needs to be reasonably well understood by the evaluator(s) to identify and examine pertinent variables during analysis.

performed to determine if some approaches are more applicable to the evaluation of collaborative systems than others.

In gathering metrics, evaluators are interested in assessing the effectiveness, efficiency and satisfaction with the product and the process. Gutwin describes these measures in his workshop paper with respect to usability testing,

*“Effectiveness considers whether the activity was successfully completed, and the number and severity of errors made during that activity. A usable groupware system will not prevent the mechanics of collaboration from taking place, and will not cause group members to make undue errors in those activities.*

*Efficiency considers the resources (such as time or effort) required to carry out the activity. A good groupware system will allow the activities of collaboration to proceed with less time and effort than will a system with usability problems. Note that any measures of efficiency must be carefully focused on task activities, since groups often engage in off-task activities that are not detrimental to the overall shared work. Satisfaction considers whether the group members are reasonably happy with the processes and outcomes of each of the activities of collaboration. Satisfaction will sometimes overlap with efficiency and effectiveness (that is, problems in the other areas are likely to reduce satisfaction).”*

There are many methods and techniques for evaluation. The workshop participants generated the list below, acknowledging that these methods and techniques are to be used in combination, at different stages of the trial, and have varying levels of effectiveness depending on the situation being evaluated. There is clearly a need for research to understand and to document these (and potentially others) with respect to groupware evaluation:

- Archival analysis
- Informal discussion
- Heuristic evaluation
- Real-time
  - Walkthroughs
  - Participant observations
- User Questionnaires
  - Numerically scaled
  - Free response
- Inspection
- Simulation
- Standard problems (scenarios)
- Automatic data logging/visualization tools
- Think-out-loud studies
- Interaction protocol analysis
- Interviews

Developing useful metrics is key to successful evaluation. Ideally, we would like to be able to match the

type of collaboration and the type of tools being evaluated with approaches, methods and tools to generate useful metrics. There is room for considerable research to enable the measurement of product and process efficiency, effectiveness, and satisfaction in a collaborative enterprise. There are issues related to gathering baseline data. This certainly must be done, but often evaluations are done in an environment where metrics have never been identified or gathered. After an attempt to match qualitative and quantitative measures to groupware and collaboration characteristics, the workshop participants decided that this is an area ripe for additional research.

### **What are the current problems with collaboration system evaluation?**

Workshop participants rapidly developed a list of problems and challenges associated with collaborative system evaluation. Then we tried to identify existing and potential solutions for these problems.

**Long-term evaluations are difficult and expensive.** This problem occurs in collaborative enterprise evaluations where tasks are spread out over many months. The field study evaluation method is particularly prone to this problem. One solution is to incorporate automated data collection with only periodic evaluator involvement during the data gathering phase of the field study.

**Evaluating many participants is difficult.** As opposed to single-user evaluation, with its documented techniques [5], collaboration system evaluation can involve many participants. It is necessary to develop evaluation techniques that can be performed requiring few or no participants while the software is in early development phases to minimize the number of trials involving lots of people. It was noted that some scaling issues may not always be detectable with only a few participants and that some tests with large numbers of participants will still have a role.

**Evaluation costs are high.** These costs include researcher costs, access time and participant costs. These costs may be keeping effective evaluations from being completed, resulting in more failures when deploying ineffective collaborative systems. It is important to develop alternative, low-cost methods and tools for evaluation. Furthermore, guidelines for collaborative-software graphical user interfaces (GUIs) like those existing for single-user GUIs, would enable less expensive and evaluation and development of collaborative systems. Along these lines, we also need to develop groupware benchmarks, scenarios and requirements against which to complete future evaluations. There is also a need to

develop data-collection tools with data-viewing capabilities to reduce the time required to understand the vast amounts of data gathered. Further, we need to educate users on how their costs can be reduced with implementation of application-appropriate groupware systems. An interesting area for research is in the development of groupware design patterns [1, 2, *et al.*], like those used for single-user software engineering.

**Getting participants for field studies is difficult.** Field studies of collaborative systems offer the opportunity to truly understand a system's effectiveness, efficiency and satisfaction levels. However, potential participants are wary of inviting evaluators into their shop for fear of productivity losses due to these evaluations. Some useful techniques to minimize this would be automated data collection and embedded, infrequent, random, short, on-line questionnaires. There is also a need to collect success stories in order to "sell" the value of the implementation of a collaborative system as well as the knowledge gained in identifying enterprise processes to potential participants.

**Field study evaluation is difficult.** While there are many challenges with all types of collaborative system evaluations, field studies are particularly difficult. Obviously, it is necessary to simplify methods for field studies and to develop cheaper field methods. The value in the use of design ethnography [3] was noted. Two potential solutions include: more instrumented beta testing and more robust instrumented feedback.

**Cultural and organizational issues affect ability to evaluate cause and effect.** When something does not work, is it because of the software or is it due to cultural or organizational effects? Often, studies fail to ascertain the source of success or failure or, worse yet, attribute it incorrectly. We need research to categorize group characteristics and then be able to match appropriate groupware characteristics. We also need to know what to look for so that cultural and organizational issues can be properly attributed during evaluation. Furthermore, evaluations and publication of field studies should include cultural and organizational issues, even if the discussion of such things is merely anecdotal.

**Evaluators may have unclear understanding of the technical applications in use in a collaborative system.** In other words, those who may be familiar with the groupware being used or with group collaboration evaluation techniques may not understand the end-users' disciplines, resulting in additional evaluation challenges. It would be useful to develop a typology that addresses how categories of log data correlates with specific metrics, e.g.,

tag log data so that it relates to user activities. This would facilitate intelligent logging of data.

**There are insufficient heuristics for collaborative software evaluation.** There are no basic groupware design principles, and therefore, there are no basic heuristics that can be used for collaborative software evaluation, as opposed to basic design principles and heuristics for single-user, software systems. There is a need to develop and document heuristics for use in groupware evaluation.

## Conclusions and Future Research Needs

Most discussions held during this workshop eventually recognized the need for future research to better understand the complexities and nuances of collaborative enterprise evaluation. A common theme for future work is that there needs to be a better understanding of the characteristics of the collaborative enterprise of interest and then what evaluation approaches, methods and techniques and the associated metrics and tools are most appropriate for use. Additionally, there is an opportunity to exploit some of the work done with single-user systems, for example the design patterns<sup>1</sup> work recently applied to HCI (Human Computer Interaction) was cited.

Ideally, there needs to be some sort of framework or taxonomy that can answer:

- Which evaluation approaches are best for the evaluation of different types of collaborative systems?
- When in the design cycle of collaborative software development are particular approaches most effective?
- What combination of methods and techniques for gathering metrics are most effective for the situation under evaluation?
- Which approaches, methods and techniques address product and process effectiveness, efficiency and satisfaction?
- Which metrics address product and process effectiveness, efficiency and satisfaction of the collaborative enterprise under study?
- What evaluation tools and mechanisms are best for generating specific metrics?
- What categories of log data generated within collaborative software correlate with specific metrics?

Additional notes:

- There is a need to develop collaboration evaluation methods and tools that are low cost and can be done with few or no participants and that use design ethnography.

---

1. <http://www.gespro.com/lafrenid/PSA.pdf>

- Research is needed in the area of groupware design heuristics.
- As noted earlier, specifications of groupware event sets need to be generated to facilitate automated data collection/logging. Similarly, there is much work to be done in event tagging to enable intelligent data logging and visualizations of logged data.
- Benchmark collaboration scenarios and their associated evaluation measures need to be developed to help developers assess their groupware systems during design and development phases.

An issue in collaborative enterprise evaluation is that it is a multi-disciplinary activity and it is very often difficult to effectively accomplish a thorough evaluation. A significant outcome of this workshop is that NIST will be establishing a forum for information exchange and an on-line repository for useful information about collaborative system evaluation. An email distribution list (eval-ce@nist.gov) will further enable information exchange among interested participants. This workshop formalized

initial discussion in an area of great importance to the future success of collaborative enterprises.

## References

- [1] Erickson, T., Putting It All Together: Pattern Languages for Interaction Design Workshop, Proceedings of the ACM's CHI 97 Conference, Atlanta, GA, 1997.
- [2] Foote, B., Harrison, N., et. al., *Pattern Languages of Program Design 4*, Addison Wesley, 1999.
- [3] Ford, J. and Wood, L., An Overview of Ethnography and System Design. In *Field Methods Casebook for Software Design*, Wixon, D. and Ramey, J. (eds.), John Wiley & Sons, Inc., New York, 1996, pp. 269-282.
- [4] McGrath, J.E., Methodology Matters: Doing Research in the Behavioral and Social Sciences. In *Readings in Human-Computer Interaction: Toward the Year 2000*, 2<sup>nd</sup> Ed., Baecker, R.M., Grudin, J., Buxton, W.A.S., Greenberg, S. (eds.), Morgan Kaufman Publishers, San Francisco, 1995, pp. 152-169.
- [5] Neilsen, J., *Usability Engineering*, Morgan Kaufmann, San Francisco, 1994.