

# Evaluating Collaborative Enterprises - A Workshop Report

Michelle P. Steves

National Institute of Standards and Technology  
msteves@nist.gov

Robert H. Allen

University of Maryland, College Park  
rhallen@nist.gov

## Abstract

*As part of the the 10th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaboration conference (WETICE 2001), the Workshop on Evaluating Collaborative Enterprises continued from its debut last year exploring the issues surrounding the evaluation of collaborative enterprises. This exploration included methods and tools for evaluating collaborative software as well as application-specific evaluation experiences. This paper describes the workshop's mission, summarizes the nine workshop presentations and one demonstration, and presents the main themes emanating from this workshop. These include the need for integrated theories, the need for advances in evaluation techniques and tools, and the need to confront non-controlled variables in case studies.*

## 1. Participants

In addition to the co-chairs, the following people participated in this workshop and contributed to the contents of this report: Kristina Buckley, The MITRE Corporation; Jeffrey Campbell, University of Maryland, Baltimore County; Jill Drury, The MITRE Corporation; Mark Klein, Massachusetts Institute of Technology; Julian Newman, Glasgow Caledonian University; David Pinelle, University of Saskatchewan; Elaine Raybourn, Sandia National Laboratories; Diane Sonnenwald, University of North Carolina, Chapel Hill; and Brent Stewart, University of Washington, School of Medicine.

In addition to the co-chairs, the following people served on the program committee. We gratefully acknowledge their time and energy: Janet Allen, Georgia Tech; Jeffrey Campbell, University of Maryland, Baltimore County; Jill Drury, The MITRE Corporation; Saul Greenberg, University of Calgary; Carl Gutwin, University of Saskatchewan; Doug Johnson, Hewlett Packard; Amy Knutilla, Knutilla Technologies; Emile Morse, NIST; Srinivas Nidamarthi, ABB Technologies; Catherine Plaisant, University of Maryland, College Park; William Regli, Drexel University; Eswaran Subrahmanian,

Carnegie Mellon University; and Marian Williams, University of Massachusetts, Lowell.

## 2. Introduction

Effective collaboration involves people sharing information. Internet technology has enabled basic communication infrastructures to facilitate people working together collaboratively over space and time. However, as Gutwin and Pinelle [1] point out, using these applications have proven much more difficult in enabling effective collaboration within and among enterprises than some Internet advocates would suggest. Researchers need tools to quantify the 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 will facilitate progress in Web-based collaboration.

The 2001 workshop on Evaluating Collaborative Enterprises builds upon the research presentations and the group discussions from last year's debut of this workshop [2]. Two common themes identified at WETICE 2000 were the need for research to better understand the characteristics of the collaborative enterprise of interest, and the need for identifying appropriate evaluation approaches, methods and techniques, and their associated metrics and tools. The workshop participants distinguished formative and summative evaluation techniques - the usefulness of various techniques at different stages of groupware development. We also highlighted many of the problems associated with groupware evaluation, which include but are not limited to: expense, participant willingness, cultural issues, group dynamics, and evaluator perspective.

This session continued in the vein set forth in the previous workshop and extended 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 is an increasingly attractive method for collaboration. Effective evaluation of

groupware holds one key to improving collaboration 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 groupware evaluation. Consequently, another outcome of this workshop is the plan for further information exchange among researchers, software developers, and users of collaborative systems to improve the development and use of methods for the evaluation of collaborative enterprises.

In this report, we present a summary of the nine papers presented (two of which were invited) and one demonstration. These papers and demonstration discussed evaluation methods and approaches (established and visionary), tools for evaluation (existing and prototype), and research needs. The papers and demonstration were presented and discussed in-depth across all three days of the workshop. This report summarizes the papers, demonstration, and discussions, and concludes with future research needs.

### 3. Summary of Papers

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

Kristina Buckley presented “A Practitioners Guide for Assessing Collaboration Systems”, which provided guidelines for assessing the effectiveness of collaboration systems in an operational environment. Specifically, she addressed three categories of metrics: usage statistics, business process and outcome metrics, and system usability metrics. After discussing when and how to collect metrics, she presented a hierarchy of understanding among metric sources, such as observation, anecdotal evidence, and number of user accounts and logons. The main conclusions are that effective evaluations should be characterized by three guiding principles. First, metrics should be focused, tying collaboration goals with process and outcome metrics. Secondly, data collection on usage statistics, outcome metrics and usability metrics, and usability metrics should be collected for a comprehensive evaluation. Thirdly, collection efforts should be streamlined to minimize required resources and imposition on participants.

“Validation, Evaluation and Virtual Enterprise Research”, by Julian Newman, used three case studies to highlight difficulties in Computer-Supported, Cooperative Work (CSCW) research, such as a multiplicity of paradigms and research cultures, and the weakness of the evaluators compared with others competing for project resources. He notes that, like Artificial Intelligence and software engineering researchers in the early days of their respective fields, CSCW researchers have difficulties because of the immaturity of their field. Additionally, because of the multi-disciplinary nature of CSCW, evaluators bring different perspectives on the proper scope and purposes of evaluation that are bound largely to the standards embodied within the research programs of their respective disciplines. Newman further notes that in the literature, collaborative system evaluation has many characteristics of a case study, and that evaluators sometimes refer to the deployment of the collaborative system as an “experiment”. While the evaluation of system deployment provides an important, empirical check on the workability of the system, it typically lacks most of the essential characteristics of the experimental method.

Michelle Steves presented “Looking at the Whole Picture: A Case Study of Analyzing a Virtual Workplace”, where she discussed her evaluation of a groupware system supporting automated, robotic welding. The initial analysis employed a user-centered evaluation method and found a majority of the usability problems. Additionally, the analysis revealed that 75% of the time the groupware system was used asynchronously. This led to a re-evaluation of the data from an artifact-centered perspective. The artifact-centric method (tracking email and persistent, shared objects) yielded additional usability issues, but more importantly, provided previously obscured process information about how the system was used. In combination, the two methods yielded superior usability results. And, the artifact-centric method added process information that can be used to help software developers prioritize addressing identified usability issues.

In “Group Task Analysis for Groupware Usability Evaluations”, David Pinelle introduced a model for analyzing and specifying group tasks. This task model incorporates the mechanics of collaboration, levels of coupling and variability in task performance. The model is designed to extract contextual information from practical work settings. By doing so, analysis of group tasks can be made more pertinent. The descriptions from this task analysis method can be used to contextualize a variety of inspection techniques including groupware walkthrough, a methodology based on cognitive walkthrough.

Jeffrey Campbell presented “Evaluation of Interference During Collaborative Document Development”, where he discussed an evaluation scenario using the development of a database design as a good scenario for evaluation of collaborative systems for the development of both text and diagram artifacts. Several metrics and measures are described for quantifying the occurrence of interference. In particular, five survey questions were added to a standard usability instrument to address interference and protection. Diagram interaction graphs provide a concise visualization for the large amount of log data recorded.

In “Using Innovation Diffusion Theory To Guide Collaboration Technology Evaluation: Work In Progress”, Diane Sonnenwald presented diffusion theory as a foundation for a survey to evaluate collaboration. Using five attributes from the theory - relative advantage, compatibility, complexity, trialability and observability - the authors developed a multi-scale survey to compare face-to-face interaction with distributed interaction.

Elaine Raybourn, “An Evaluation of an Intercultural Virtual Social-Process Simulation”, presented a virtual learning environment, DomeCity MOO (Multi-User Domain Object Oriented), which participants perceived as less threatening than face-to-face simulation. This was due in part to the anonymity of the virtual system and to the participant's perceptions about communication regarding identity and power.

In addition to the seven presentations corresponding to the reviewed papers that follow this report, the workshop hosted two invited presentations and one demonstration. These are briefly described below.

Mark Klein presented “What Complex Systems Research Can Teach Us About Collaborative Design”, where he shared his ideas about collaborative design being challenging because of strong, nonlinear interdependencies among designers. To help overcome competing, and sometimes, conflicting goals, designers must at times concede their optimum sub-system design for the optimum design of the system. The presentation concluded with some suggestions concerning a novel class of CSCW applications.

Brent Stewart presented “Tumor Conferencing Tools for Regional Collaborative Cancer Care Using the Next Generation Internet” where he discussed the design, construction, and evaluation of specific collaborative tools using the Next Generation Internet for three key steps in the diagnostic work-up, management and treatment of a cancer patient. He discussed patient assessment, diagnosis and selection of treatment modality, and treatment. He focused on tumor board conferencing, where all clinical information is contemporaneously brought to a group of cancer specialists. Three multi-disciplinary teams were

assembled to develop these collaborative tools: the technical infrastructure, contextual inquiry and design, and telepresence and collaborative work teams.

David Pinelle gave a short presentation and demonstration of work done at the University of Saskatchewan entitled “Log Data Exploration with the Flexible Log Visualizer”. The Flexible Log Visualizer extends work presented at last year's workshop on the CollabLogger [3]. The Flexible Log Visualizer accepts a flexible (tagged) log format, has an improved query interface, and improved timeline visualizations. Several participants voiced interest in using the tool in their evaluation work.

#### 4. Summary of Discussions

Many of the papers presented had common themes among them. The principal ones are as follows:

- Challenges still exist in how we evaluate collaborative enterprises; after 15 years of CSCW research, there is no integrated theory of collaboration. Integrated theories will benefit evaluation and vice versa.
- Researchers can not control all independent variables, and worse, do not usually relate how they were handled in case study documentation.
- Advances are needed in evaluation techniques and tools. Better evaluation will be supported by improved data collection, data formats, data categorization, and data visualization tools.
- Groupware will better support the way we work, especially as switches between synchronous and asynchronous modes become seamless.

Additionally, participants recognized that there are many challenges that are faced in evaluating collaborative enterprises. They include:

- Some data collection schemes are automatic, such as logging; however, others are labor intensive, such as interviews. Analysis is typically labor intensive.
- Collaborative system evaluation can involve many participants, requiring more extensive techniques than traditional (singleware) evaluation [4].
- Cultural and organizational issues are often problematic.
- Evaluators often have inherent biases.

Another theme that emerged was the lack of control and measurement of many variables. For example, collaborative technology is being put in place for the Intelligence Community to reduce the risk of negative events, such as terrorist attacks. However, if an event occurs, but not for a long time, how is the effectiveness of the system measured? In addition, those metrics difficult

to control and measure are often underreported in the literature.

As identified last year, there are many methods and techniques for evaluating collaborative enterprises [1]. However, many of the methods, such as data logging and visualization tools, are in their developmental stages and need to become more robust. Intelligent data logging is on the horizon and should contribute to better, more streamlined data analysis.

The theme that groupware would prevail in our working culture was echoed by many of the case studies reported in the papers. Even those papers in which it was reported that groupware was underutilized, participants felt a high sense of reward by learning how to collaborate in different ways. This message was the most consistent throughout the workshop.

## **5. Conclusions and Future Work**

While there is an opportunity to exploit the work done in other disciplines and adopt it for groupware evaluation,

as reported in several of the workshop papers, most discussions held during this workshop came around to the need for future research to contribute towards an integrated theory of collaboration. It was concluded that next year's workshop would address that issue.

## **6. References**

- [1] D. Pinelle and C. Gutwin, "A Review of Groupware Evaluations", Proceedings of WETICE 2000, (Gaithersburg, MD, Jun. 2000), IEEE Computer Society, pp. 86-91.
- [2] A. Knutilla, M. Steves, R. Allen, "Workshop on Evaluating Collaborative Enterprises - Workshop Report", Proceedings of WETICE 2000, (Gaithersburg, MD, Jun. 2000), IEEE Computer Society, pp. 79-85.
- [3] E. Morse and M. Steves, "CollabLogger: A Tool for Visualizing Groups at Work", Proceedings of WETICE 2000, (Gaithersburg, MD, Jun. 2000), IEEE Computer Society, pp. 104-109.
- [4] Nielson J, Usability Engineering, Morgan Kaufmann, San Francisco, CA 1994.