Evaluating a Scientific Collaboratory: Results of a Controlled Experiment

DIANE H. SONNENWALD, MARY C. WHITTON, and KELLY L. MAGLAUGHLIN
University of North Carolina at Chapel Hill

The evaluation of scientific collaboratories has lagged behind their development. Do the capabilities afforded by collaboratories outweigh their disadvantages? To evaluate a scientific collaboratory system, we conducted a repeated-measures controlled experiment that compared the outcomes and process of scientific work completed by 20 pairs of participants (upper level undergraduate science students) working face-to-face and remotely. We collected scientific outcomes (graded lab reports) to investigate the quality of scientific work, post-questionnaire data to measure the adoptability of the system, and post-interviews to understand the participants' views of doing science under both conditions. We hypothesized that study participants would be less effective, report more difficulty, and be less favorably inclined to adopt the system when collaborating remotely. Contrary to expectations, the quantitative data showed no statistically significant differences with respect to effectiveness and adoption. The qualitative data helped explain this null result: participants reported advantages and disadvantages working under both conditions and developed work-arounds to cope with the perceived disadvantages of collaborating remotely. While the data analysis produced null results, considered as a whole, the analysis leads us to conclude there is positive potential for the development and adoption of scientific collaboratory systems.

Categories and Subject Descriptors: H.5.3 [Information Interfaces and Presentation]: Group and Organizational Interfaces—Computer-supported cooperative work; Evaluation/methodology; Synchronous interaction; Theory and models; J.2 [Physical sciences and engineering]: Physics; Chemistry

General Terms: Design, Human Factors, Experimentation

Additional Key Words and Phrases: Scientific collaboratory, geographically-distributed work, collaboration, controlled experiment, nanoscience

The development of the nM and this work have been funded by the NIH National Center for Research Resources, NCRR 5-P41-RR02170. The nanoManipulator project is part of the Computer Graphics for Molecular Studies and Microscopy Research Resource at the University of North Carolina at Chapel Hill.

Authors' addresses: Diane H. Sonnenwald, School of Information and Library Science, The University of North Carolina, Chapel Hill, NC 27599; Mary C. Whitton, Department of Computer Science, The University of North Carolina, Chapel Hill, NC 27599; Kelly L. Maglaughlin, School of Information and Library Science, The University of North Carolina, Chapel Hill, NC 27599; email: {dhs, maglk}@ils.unc.edu, whitton@cs.unc.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 1515 Broadway, New York, NY 10036 USA, fax: +1 (212) 869-0481, or permissions@acm.org.

© 2003 ACM 1073-0616/03/0600-0150 $5.00

1. INTRODUCTION

A collaboratory has been defined as “a center without walls, in which researchers can perform their research without regard to physical location—interacting with colleagues, accessing instrumentation, sharing data and computational resources, and accessing information in digital libraries” [Wulf 1989, P. 19]. Since the US National Research Council’s 1993 call for support to develop, refine, and evaluate the collaboratory concept in realistic settings [National Research Council 1993], a number of scientific collaboratories have been developed. They include the Worm Community System [Schatz 1991], Upper Atmospheric Research Collaboratory [Finholt and Olsen 1997], and the Environmental Molecular Sciences Laboratory [Kouzes et al., 1996].

The evaluation of scientific collaboratories has lagged behind their development. Of the 31 collaboratories we identified through print and electronic media, only eight have been formally evaluated or are currently undergoing evaluation. So few evaluations of scientific collaboratories exist that the fundamental questions have yet to be answered: Can distributed scientific research produce high quality results? Do the capabilities afforded by collaboratories outweigh their disadvantages from scientists’ perspectives? How does the scientific process change in the context of a collaboratory? Will scientists adopt the collaboratory software? How do scientific and organizational cultures impact adoption and nonadoption of the collaboratory system? Are there system features and performance characteristics that are common to successful collaboratory systems? Our goal is to help answer the fundamental questions by evaluating a specific scientific collaboratory.

Our context is a collaboratory system providing access to a specialized scientific instrument, a nanoManipulator [Taylor and Superfine 1999; Sonnenwald et al., 2001b]. The system is being evaluated in two phases: a controlled laboratory study and an ongoing longitudinal field study of the system in use by collaborating scientists. This paper is a detailed report of the controlled laboratory study.

2. EVALUATING SCIENTIFIC COLLABORATORIES

Evaluating scientific collaboratories has unique challenges, many of which can be attributed to the context in which science occurs and the substantial resources required to perform the studies. Evaluation purpose, the scientific context and available resources were all factors in the overall evaluation design for our study.

Collaboratory evaluation can have multiple purposes and goals. Examples include: increasing our understanding of individual behavior in geographically distributed collaboration, discovering new knowledge about collaborative scientific work processes as mediated by technology, informing the design of collaboratory technology, and providing insights regarding the efficacy of scientific collaboratories. These purposes are complex and multi-faceted, often requiring

---

1For comprehensive reviews of collaboratories see Finholt [2001] and Kouzes [2000].
multiple comprehensive studies that employ qualitative and quantitative research methods.

Scientific research often occurs in contexts where expertise, instrumentation and laboratory resources are scarce, costly, and in high demand, and where work processes are rapidly evolving. Using scarce resources to perform evaluation studies often competes with the progress of natural science research. Furthermore the population of potential study participants can be relatively small and in flux. A typical population includes undergraduate and graduate students, post-doctoral fellows, faculty and corporate scientists who have the required specialized scientific knowledge. Students and post-doctoral fellows, however, have naturally high turnover rates, and faculty and corporate scientists may face severe time and intellectual property ownership constraints. Another difficulty is that scientific research is dynamic; processes are evolving, sometimes very rapidly, and tasks may change over the course of an evaluation study. It is a challenge for evaluators to capture and understand the changing activities in order to determine methods and measures to evaluate them.

Other resources required to conduct an evaluation include personnel knowledgeable about evaluation methods and about the collaboratory system and its role in scientific investigations, as well as the equipment, supplies, travel funds, and time to collect and analyze evaluation data. The mere fact of a collaboratory’s geographic distribution may make a comprehensive on-site field study incorporating in-depth interviews and observations prohibitively expensive. Collaboratory technology may also severely limit the number of possible study settings and participants. For example, some collaboratories, such as the system we studied, require Internet 2 connections. Furthermore, programming resources are often required to develop tools to capture system-use data because most collaboratory systems and associated technologies do not have automatic logging capabilities or well documented programming interfaces.

The purposes of our evaluation study include providing insights regarding the efficacy of scientific collaboratories, increasing our understanding of collaborative scientific work processes mediated by technology, and informing the design of collaboratory technology. To address these multiple purposes and accommodate the scientific context of the evaluation using available resources, we selected a multi-faceted approach to evaluation.\textsuperscript{2} Our evaluation study design has two primary components: a longitudinal field study and a lab study. The first phase of the field study was completed before the controlled experiment. It yielded system design requirements as well as providing baseline data about scientists’ attitudes and collaboration behaviors. The second phase of the field study, investigating ongoing scientific collaborations using the system, is currently in progress.

The lab study was conducted early in the collaboratory’s life cycle. Its purpose was to generate data to allow a controlled comparison of collaborating face-to-face and remotely. Our goal was to gain insights into the potential adoption and use of a collaboratory, including participants’ attitudes towards the

\textsuperscript{2}For an overview of approaches to collaboratory evaluation, see the Digital Library Appendix.

collaboratory and how the collaboratory may impact scientific task processes and outcomes.

Previous research in computer supported cooperative work [Dourish et al., 1996; Olson and Olson, 2000] and theory of language [Clark 1996] would predict that working remotely would lack the richness of collocation and face-to-face interaction, for example, multiple and redundant communication channels, implicit cues, spatial co-references, that are difficult to support via computer-mediated communications. This lack of richness is thought to impair performance because it is more difficult to establish the common ground that enables individuals to understand the meaning of each other's utterances. Other research [Starr and Ruhleder 1996; Orlikowski 1993; Olson and Teasley 1996] would predict that working remotely may not be compatible with structural elements, such as reward systems and work practices, found in existing work environments, and thus not adopted by individuals. Thus, our hypotheses were:

H1: Study participants will be less effective collaborating remotely than collaborating face-to-face.
H2: Study participants will report more difficulty collaborating remotely than collaborating face-to-face.
H3: Study participants will report they are more likely to adopt the system after using it face-to-face than remotely.

In the following sections we report on the controlled lab study conducted to test these hypotheses, including discussions of the context of the evaluation, the lab study design, data collection and analysis, and the results of the controlled lab study and their implications.

3. EVALUATION CONTEXT: THE nanoMANIPULATOR COLLABORATORY

The collaboratory system we are evaluating provides distributed, collaborative access to a specialized scientific instrument called a nanoManipulator (nM). The single-user nM provides haptic and 3D visualization interfaces to a local (co-located) atomic force microscope (AFM), providing a natural scientist with the ability to interact directly with physical samples ranging in size from DNA to single cells. An nM can be used in live and replay modes. In live mode, an nM is used both to display and record data from an atomic force microscope and to control the microscope. The recorded data, including all data produced by the microscope, is saved in a “stream file” so that it can be replayed later for analysis. In replay mode, the nM is a display device where the stream file, instead of the live microscope, provides the data for the visual and haptic displays. Approximately 80% of nM use is in replay mode where scientists move forward and backward through the data, stopping at critical points to perform visualization and analysis. Details regarding the nM and its uses are described in Finch et al., 1995; Taylor II and Superfine (1999); Guthold et al. (2000); and Guthold et al. (1999).

The design of the collaboratory version of the nM was based on results of an ethnographic study from which we developed an understanding of the scientific research process, current collaborative work practices, the role of an nM...
as a scientific instrument, scientists’ motivation for interdisciplinary collaboration across distances, and scientists’ expectations regarding technology to support scientific collaborations across distances [Sonnenwald et al. 2001b]. The collaboratory system is shown in Figure 1 and demonstrated in [Hudson et al. 2000]. The collaboratory system is based on two PCs. One PC is equipped with a Sensable Devices Phantom™ force-feedback device. This PC and its associated software provide haptic and 3D visualization interfaces to a local or remote atomic force microscope (AFM) and support collaborative manipulation and exploration of scientific data in live and replay modes.

The collaboratory system allows scientists to dynamically switch between working together in shared mode and working independently in private mode (see Figure 2). In shared mode, remote—noncollocated—collaborators view and analyze the same (scientific) data. Mutual awareness is supported via multiple pointers, each showing the focus of attention and interaction state for one collaborator. We use optimistic concurrency techniques in shared mode [Hudson et al. 2003], eliminating explicit floor control and allowing collaborators to perform almost all operations synchronously. Because of the risk of damage to an AFM, control of the microscope tip is explicitly passed between collaborators. In private mode, each collaborator can independently analyze the same or different data from stream files previously generated. When switching back to private from shared mode, collaborators return to the exact data and setting they were using previously.

The second PC supports shared application functionality and video conferencing (via Microsoft NetMeeting™) and an electronic writing/drawing tablet. This PC allows users to collaborate using a variety of domain-specific and
off-the-shelf applications, including specialized data analysis, word processing and whiteboard applications. Video conferencing is supported by two cameras. One camera is mounted on a gooseneck stand so it can be pointed at the scientist's hands, sketches, or other physical artifacts scientists may use during experiments; the other is generally positioned to capture a head and shoulders view of the user. Collaborators have software control of which camera view is broadcast from their site. Previous research [Bellotti and Dourish 1997; Harrison et al. 1997] has illustrated the importance of providing the ability to switch between multiple camera views, as well as repositioning and refocusing cameras.

A wireless telephone connected to a commercial telephone network provides high quality audio communications for collaborators. A telephone headset and speakerphone options are also included to allow users mobility and provide the option of having others in the room participate in a conversation with a remote collaborator.

4. THE CONTROLLED LAB STUDY
The controlled lab study was a repeated measures controlled experiment comparing working face-to-face and working remotely with the order of conditions counterbalanced. This type of experiment is also referred to as a “mixed design” because it allows both within-group and between-group comparisons.
Twenty pairs\(^3\) of study participants conducted two realistic scientific research activities each requiring 2 to 3 hours to complete. Participants worked face-to-face on one occasion and, on a different day, collaborated remotely (in different locations). When face-to-face, the participants shared a single collaborative system; when collaborating remotely, each location was equipped with a complete collaborative system. We collected a variety of quantitative and qualitative evaluation data, including task performance measures to compare the quality of scientific work produced in the two collaboration conditions, post-interviews to gain, from participants’ perspectives, a more in-depth understanding of the scientific process in both conditions, and post-questionnaire data to evaluate the adoptability of the system.

4.1 Study Participants
The study participants were upper-level undergraduate natural science students from local Research I universities. We chose this population because it is relatively large and representative of individuals who perform scientific research, most often under the direction of faculty or postdoctoral fellows. The science and math skills of this pool are somewhat consistent, as they have taken a similar set of core science and math courses as first and second year students. Study participants were recruited through announcements in class, student newspaper advertisements, posters, and e-mail announcements. They received $100 for their participation (less than $10.00 per hour).

The majority of the 40 participants reported they were majoring in biology and had A/B grade point averages; no participant reported a GPA lower than a C (see the Digital Library Appendix for details). Thirty-six participants were Caucasian, 2 were African American and 2 were Asian/Indian. All were fluent in English and all but one appeared to be a native English speaker. Participants were randomly assigned to pairs without respect to their undergraduate major, self-reported GPA and ethnicity, and pair assignments did not change over the course of the experiment. We strove for a mix of gender composition in the pairs; 9 pairs were of mixed gender, 6 pairs were female only and 5 pairs were male only. To avoid bias or confounding results, we selected participants who had no experience collaborating across distances or using the nanoManipulator. In particular, no participant had any substantive knowledge of fibrin, the biological material under investigation in the collaborative activities.

All study participants had previous experience collaborating face-to-face with others while conducting scientific experiments and working on class projects. Twenty-five percent of the study participants (5 pairs out of 20) knew their partner before participating in the experiment, a situation that mirrors scientific and teaching practice. Scientists who collaborate may know each other, however, they frequently have their students or postdoctoral fellows, who do not know each other, work together to design and conduct the actual experiments and data analysis for their collaborative project. Occasionally, the

\(^3\)We were unable to perform a power analysis while planning the experiment because we did not have an estimate for the effect size. Thus, we reviewed the literature for similar studies and found that a sample size of 10 to 20 was common [Gale 1998; Gaver 1991; Gutwin and Roseman 1996].
students or postdoctoral fellows have previously met at a conference or in a
course, though more often they are meeting and working together for the first
time. Collaboratories, in particular, bring together scientists who are from dif-
ferent disciplines and locations, and who do not know each other. One scientist
may have knowledge of the scientific tool and methodology, and the other scien-
tist has knowledge of the sample to be investigated. Due to the small number
(and percentage) of previously acquainted pairs in our study, it was not possible
to determine if the previous acquaintance statistically affected the experimen-
tal outcome measures. However, the outcome measures from participants who
knew each other previously follow the same trends as the measures from the
participants who had not known each other previously.

4.2 Experiment Design
The controlled experiment consisted of three sessions as shown in Figure 3. The
sessions were iteratively refined in a pilot study that included 6 participants.

The introduction consisted of a presentation providing background informa-
tion on the controlled experiment, a thorough introduction to the natural sci-
cence used in the controlled experiment, and a brief hands-on demonstration of
the collaboratory system. During the presentation and demonstration particip-
ants were encouraged to ask questions. Study participants signed an informed
consent document and completed a demographic questionnaire, providing infor-
mation about their major, GPA, computer use and knowledge, computer game
use, age, ethnicity and preferred learning styles. This session typically lasted
45 minutes. In most cases paired participants attended the introductory session
together, however, in a few cases this was not possible.

Task sessions 1 and 2 were performed on different days and under different
conditions: face-to-face (FtF) and remote. The order of the conditions was coun-
terbalanced (see Table I). Pairs were randomly assigned to the two order con-
ditions. The task sessions took place within a one-week timeframe. Each task
session had three parts: a tutorial, scientific research lab, a post-questionnaire
and a post-interview.

The hands-on tutorial led participants through instructions on how to use
the features of the collaboratory system required for that day’s lab. The tutor-
ial before the remote collaboration session also included instructions on the
video conferencing system, shared applications, and the collaboration-specific features of the system. Each participant completed the tutorial in a separate location and was accompanied by a researcher/observer who was available to assist and answer questions. Participants were allowed to spend as much time as they wanted on the tutorial; typically they spent 45 minutes.

Both scientific research labs were designed in collaboration with natural scientists who regularly use the nanoManipulator to conduct their scientific research. The tasks were actual activities the scientists completed and documented during the course of their investigations. The labs were designed to be similar in difficulty as judged by the natural scientists and pilot study participants. To complete the tasks in the labs the participants had to engage in the following activities typical of scientific research: operate the scientific equipment properly; capture and record data in their (electronic) notebook; perform analysis using scientific data analysis software applications and include the results of that analysis in their notebooks; draw conclusions, create hypotheses and support those hypotheses based on their data and analysis; and prepare a formal report of their work. We did not require the study participants to design a natural science experiment or write a paper describing the experiment because the collaboratory system under evaluation was not designed to explicitly support these components of the scientific research cycle.

During each scientific research lab, study participants were asked to work together, using the collaboratory system in replay mode to manipulate and analyze data recorded previously during an experiment conducted by a physicist [Guthold et al. 2000]. As discussed above, the pre-recorded stream file contained an exact and complete record of all data collected from an AFM when the experiment was originally performed. All visualization options and controls on the system, except “live” microscope control, were available to the study participants in replay mode.

The subject of the scientific research labs was the structure of fibrin, a substance critical for blood clotting. In the first lab, participants were asked to measure distances between branch points of fibrin fibers and to discuss the possible relationship between these distances and the blood clotting process. In the second lab, participants were asked to measure additional structural properties of fibrin, and based on these measurements, discuss its possible interior structure. (See the Digital Library Appendix for more details on these tasks).

Study participants were asked to document their results, recording data they collected and their analysis of that data, in a lab report. The lab report mirrored lab notes created by the scientists when they originally conducted their fibrin investigation. Lab reports created by participants contain data images, tables of

<table>
<thead>
<tr>
<th>Table I. Conceptual Experiment Design: Repeated Measures with the Order of Conditions Counterbalanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order of Conditions</td>
</tr>
<tr>
<td>Condition: Group A: Collaborated FtF first</td>
</tr>
<tr>
<td>Type of Interaction: Face-to-Face (FtF)</td>
</tr>
<tr>
<td>Task Session 1</td>
</tr>
<tr>
<td>Type of Interaction: Remote (REM)</td>
</tr>
<tr>
<td>Task Session 2</td>
</tr>
<tr>
<td>Order of Conditions: Group B: Collaborated remotely first</td>
</tr>
<tr>
<td>Type of Interaction: Face-to-Face (FtF)</td>
</tr>
<tr>
<td>Task Session 2</td>
</tr>
<tr>
<td>Type of Interaction: Remote (REM)</td>
</tr>
<tr>
<td>Task Session 1</td>
</tr>
</tbody>
</table>

data values, explanatory text and annotated graphs illustrating their analysis of their data. A single report was requested from each pair of study participants for each task session. On average, participants spent 163 minutes working on the first lab, and 121 minutes working on the second lab (with standard deviations of 27.9 and 39.2 minutes, respectively). When working remotely, they typically spent 16.6 minutes longer on the first lab and 20.6 minutes longer on the second lab. However, the differences in completion times are not statistically significant. A multivariate analysis of variance (MANOVA) using a general linear model\(^4\) yielded no statistically significant differences in completion times due to condition or due to any interaction effect between condition and order, at the 0.05 level of significance.

During each task session a researcher observed the study participants and video- and audio-tape recordings were made of their interactions while completing the labs. After each lab, each study participant was asked to complete a post-questionnaire and to participate in a one-on-one interview with a researcher. The post-questionnaire took approximately 20 minutes to complete, and post-interviews lasted between 30 and 60 minutes. The questionnaires and interviews provided data regarding participants’ perceptions of the lab activities, of the technology in the collaboration system, and of the collaborative process as discussed below.

4.3 Evaluation Measures

4.3.1 Task Performance (Outcome) Measure: Lab Reports. A primary goal of our overall evaluation study is to compare the quality of science produced in face-to-face collaborations to that produced in remote collaborations. Typically statistics such as number of publications, citation counts, number of grants and patents awarded, and peer reviews are used to measure science quality. These measures, however, require years of performance and data collection that are not possible in evaluation studies with a limited timeframe. Therefore, we chose to have study participants create laboratory reports that were modeled on scientists' lab notes that document scientists' data collection and analysis progress. We graded the reports and used the grades as a task performance measure—as a measure of the quality of science conducted face-to-face and remotely.

The instructions for the lab activities and for what should be included in the laboratory reports were designed in collaboration with natural scientists. As is typical in controlled experiments, the instructions were specific and guided participants’ actions. The information participants were asked to provide in the reports mirrored the information found in the scientists’ lab notes created when they conducted their original research on fibrin. Each pair of study participants collaboratively created a lab report under each condition, generating a total of 40 lab reports; 20 created working remotely and 20 created working face-to-face.

The lab reports were graded blindly; the graders had no knowledge of the lab report authors or under which condition the report was created. Initially

\(^4\)The multivariate analysis of variance statistical procedure used throughout this paper is part of the General Linear Model provided by SAS, version 8.2.
one instructor, a Ph.D. student in physics, created a grading template in collaboration with a physics professor. The graduate student and two instructors used the template to grade a subset of lab reports (6). The instructors discussed differences among their grades and modified the grading template, adding further details, to reflect their discussion. An additional subset of reports (6) was graded using the updated template. Intercoder reliability was calculated for these assigned grades using Cohen’s Kappa [Robson 1993]. Values of .75 and .79 were calculated for graded lab reports from the first and second task sessions respectively. Values above .70 are considered excellent [Robson 1993]. One instructor used the final template to grade all remaining lab reports.

4.3.2 Participants’ Perceptions: Post-interviews. To further our understanding of participants’ perceptions of the system, we conducted semi-structured interviews with each participant after each task session. Study participants were asked what they thought about their experience, including the most satisfying and dissatisfying aspects of their experience [Flanagan 1954]. In addition, we inquired about specific incidents that were noted by the observer, work patterns that emerged during the experience, and the impact technology may have had on their interactions with their research collaborator. After Task Session 2, participants were also asked to compare working face-to-face and working remotely. To better learn each participant’s perspective, participants were interviewed individually, for a total of 80 interviews, each lasting from 30 to 60 minutes. Each interview was audio-recorded and transcribed.

The interviews were analyzed using both open coding and axial coding [Berg 1989]. During open coding a subset of the interviews were read thoroughly and carefully by two researchers, and the researchers identified coding categories, or coding frames. For example, a category that emerged was negative references to aspects of the technology. After the initial set of categories was discussed among the research team, three team members analyzed another subset of interviews. Additional coding categories emerged from this analysis. All three researchers analyzed an additional subset of interviews. No new coding categories emerged, and researchers were in agreement regarding the application of the codes. Intercoder reliability, calculated using Cohen’s Kappa, yielded values of .86 and .81. Values above .70 are considered excellent [Robson 1993]. In the final step, that is, during axial coding, all interviews were re-read and analyzed using the coding categories. For the purposes of this paper, we analyzed the following codes: references to working face-to-face; references to working remotely; comparison between working face-to-face and remotely; positive aspects of the technology; and negative aspects of the technology. (See the Digital Library Appendix for coding examples.)

4.3.3 Innovation Adoption Measure: Questionnaire. Innovation adoption and diffusion theory provided us a foundation for investigating the potential of the collaborative system for adoption by scientists. Synthesizing over five decades of innovation adoption and diffusion research, Rogers [1995] identifies
five attributes of innovations that are correlated with the adoption of innovations. The five innovation attributes are: relative advantage, compatibility, complexity, trialability and observability (Table II).

Relative advantage is the degree to which potential adopters perceive that an innovation surpasses current practices. Compatibility is the degree to which an innovation is perceived to be consistent with adopters’ existing values, past experiences and needs. It includes individual, group and organizational goals, needs, and culture, and is concerned with the level of congruence between a group’s traditional work patterns and the work patterns required by the innovation. Complexity refers to the perceived difficulty of learning to use and understand a new system or technology. When a system is perceived as complex, it is less likely to be adopted. Trialability refers to the ease of experimenting with an innovation. It includes the level of effort needed and the risk involved in observing and participating in small scale demonstrations of the system, including the ease with which you can recover from (or “undo”) an action taken using the system and the cost of reversing the decision to adopt. Observability is the degree to which the results of the innovation are easily seen and understood.

Numerous researchers have validated these attributes in a variety of domains including medicine, engineering, and airline reservation information systems [Rogers 1995; Tornatsky and Fleischer, 1990]. Researchers [Grudin 1994; Shniederman 1997; Olson and Teasley 1996; and Orlikowski 1993], have also identified the importance of the attributes in computer supported cooperative work (CSCW) contexts. Rogers’ theory and the five attributes guided the construction of our post-questionnaire.

We used the same questionnaire under both collaboration conditions to enable a comparison of results. Note, the terms “technology” and “system” in questions, such as “I learned new ways of using the technology from my partner” and “Interacting with the system is frustrating,” included a more extensive set of functionality when the collaboratory system was used remotely than when used face-to-face. That is, when participants collaborated face-to-face, the terms “technology” and “system” referred to data analysis, word processing and visualization software. When participants collaborated remotely, “technology” and “system” also included video conferencing, shared application software and collaborative features of the visualization software. In each tutorial, we introduced these various system features to the participants as needed for the task and condition. Thus, the terms “technology” and “system” in the questionnaire always referred to the system features the study participants used in the task session immediately proceeding administration of the questionnaire. Details regarding the construction and validation of the questionnaire...
Table III. Participants’ Perspective on Experiment Format and Content

<table>
<thead>
<tr>
<th>Question</th>
<th>FtF (n = 40)</th>
<th>Remote (n = 40)</th>
<th>FtF Session 1 (n = 20)</th>
<th>Remote Session 1 (n = 20)</th>
<th>FtF Session 2 (n = 20)</th>
<th>Remote Session 2 (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentrated fully on activity</td>
<td>4.27</td>
<td>4.35</td>
<td>4.40</td>
<td>4.45</td>
<td>4.15</td>
<td>4.25</td>
</tr>
<tr>
<td></td>
<td>(0.75)</td>
<td>(0.89)</td>
<td>(0.75)</td>
<td>(0.83)</td>
<td>(0.75)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>Time given to perform tasks</td>
<td>4.52</td>
<td>4.33</td>
<td>4.25</td>
<td>4.10</td>
<td>4.80</td>
<td>4.55</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
<td>(1.14)</td>
<td>(1.12)</td>
<td>(1.33)</td>
<td>(0.70)</td>
<td>(0.89)</td>
</tr>
<tr>
<td>I was provided ample training</td>
<td>4.15</td>
<td>4.25</td>
<td>4.05</td>
<td>4.20</td>
<td>4.25</td>
<td>4.30</td>
</tr>
<tr>
<td></td>
<td>(0.62)</td>
<td>(0.87)</td>
<td>(0.61)</td>
<td>(0.83)</td>
<td>(0.64)</td>
<td>(0.92)</td>
</tr>
<tr>
<td>I believe this lab is similar to work scientists do</td>
<td>3.78</td>
<td>3.58</td>
<td>3.75</td>
<td>3.55</td>
<td>3.80</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(1.11)</td>
<td>(0.85)</td>
<td>(1.05)</td>
<td>(0.95)</td>
<td>(1.19)</td>
</tr>
</tbody>
</table>

instrument can be found in the Digital Library Appendix and Sonnenwald et al. [2001].

5. RESULTS AND DISCUSSION

During the controlled experiment we collected both quantitative and qualitative data including lab report scores, post-interview data and post-questionnaire responses. The quantitative data analysis did not support the hypotheses. No statistically significant negative differences in the measures of scientific outcome and perceived adoptability of the system attributable to collaboration condition emerged. The analysis of the qualitative interview data helped explain this null result. Participants reported advantages and disadvantages working under both conditions and developed work-arounds to cope with the perceived disadvantages of collaborating remotely.

We present the detailed results in several parts. First, we examine the face validity of the scientific research labs from the participants’ perspective. Next, we examine data from each measure, examining similarities and differences that arise when working face-to-face and remotely, with respect to our hypotheses regarding scientific outcomes, participants’ perceptions of the scientific work process and technology, and collaboratory adoption.

5.1 Face Validity of Research Laboratory Tasks: Participants’ Perspective

To investigate the face validity of the tasks and task sessions from the participants’ perspective we included questions in the post-questionnaire and in the post-interview. These data (Table III) show participants reported they concentrated on the activity, had sufficient time and training to perform the tasks and believed the lab was similar to work scientists do. A multivariate analysis of variance (MANOVA) using a general linear model yielded no statistically significant differences due to condition or due to any interaction effect between condition and order, at the 0.05 level of significance. The analysis treats the scores of each participant independently of his or her partner’s score. This approach is similar to other types of evaluations, for example, end of semester class evaluations. Each individual may experience the same situation in unique
Table IV. Graded Lab Report Statistics

<table>
<thead>
<tr>
<th>Condition</th>
<th>Group A: Collaborated FtF first (n = 20)</th>
<th>Group B: Collaborated remotely first (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>FtF</td>
<td>70.0</td>
<td>16.75</td>
</tr>
<tr>
<td>Remote</td>
<td>75.1</td>
<td>10.49</td>
</tr>
</tbody>
</table>

Table V. Multivariate Analysis of Variance (MANOVA) of Differences among Lab Report Scores

<table>
<thead>
<tr>
<th>Type of Comparison</th>
<th>MANOVA Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
</tr>
<tr>
<td>Between Group</td>
<td></td>
</tr>
<tr>
<td>Condition: FtF vs. Remote</td>
<td>1</td>
</tr>
<tr>
<td>Condition &amp; Order:</td>
<td>1</td>
</tr>
<tr>
<td>FtF first &amp; Remote second vs.</td>
<td></td>
</tr>
<tr>
<td>Remote first &amp; FtF second</td>
<td></td>
</tr>
<tr>
<td>Within Group</td>
<td></td>
</tr>
<tr>
<td>Group A: FtF first vs. Remote second</td>
<td>1</td>
</tr>
<tr>
<td>Group B: Remote first vs. FtF second</td>
<td>1</td>
</tr>
</tbody>
</table>

ways, and when given the opportunity to independently share her or his perspective, as done in the study, may do so. To test this assumption, we examined the correlation of score residuals within pairs. The correlations were low: -0.20, -0.51, -0.10, and 0.06, with an average correlation of -0.19. Thus, treating the responses as independent is warranted.

Face validity was further supported by interview data. Participants reported:

*Everything was like for real.*

*I thought it was not too long, not too short...[it] didn’t make me feel unsure of what I was doing.*

The post-questionnaire and post-interview data appear to verify the face validity of the tasks from the participants' perspective.

5.2 Task Performance (Scientific Outcomes): Analysis of Graded Lab Reports

Hypothesis H1 suggests that collaborating remotely would have a negative impact on scientific task performance outcome measures. Only minimal support was found for this hypothesis. The average lab report scores for the first task session were identical (70/100) for both the face-to-face and remote condition (Table IV). Furthermore, using a multivariate analysis of variance (MANOVA) test (row 1, Table V), the differences in scores for the face-to-face and remote conditions are not statistically significant.5

5The average lab report scores were greater in the second task session for both conditions, indicating a possible learning effect. This difference is accounted for in the analysis of variance computation.
However, the data suggest that collaborating remotely first may have a positive effect on scientific outcomes in this context. When order is taken into account using a multivariate analysis of variance (MANOVA) test (row 2, Table V), participants who collaborated remotely first (Group B) scored significantly higher on the second task than did those who collaborated face-to-face first (Group A) \((p < 0.01)\). Furthermore, there is no statistically significant difference between FtF and remote lab scores for participants (Group A) who collaborated face-to-face first (row 3, Table V). However, there is a statistically significant difference \((p < 0.01)\) between the FtF and remote lab scores for participants (Group B) who collaborated remotely first (row 4, Table V).

The only statistically significant correlation (at the .05 level) between scores across conditions and order occurs among scores within Group B. Using a Pearson correlation test\(^6\) the value of the correlation between scores in Group B is \(0.698, p = 0.025\). That is, if participants received a high grade on their first lab report created when collaborating remotely, then they were likely to receive a high grade when collaborating face-to-face. The converse is not supported—the score participants received when collaborating face-to-face did not predict their score when collaborating remotely.\(^7\)

Previous research [Olson and Olson 2000] would predict that scores from a remote first session would be lower because the remote session would lack the richness of collocation and face-to-face interaction, including multiple and redundant communication channels, implicit cues, and spatial co-references, that are difficult to support via computer-mediated communications. This lack of richness is often thought to impair performance. Perhaps technical features such as seeing your partner’s pointer and functions, optimistic shared control of scientific instrumentation and applications, improved video that provides multiple views, and high quality audio communications may be “good enough” for scientific tasks focusing on collecting, analyzing and interpreting data.

Further, the literature would predict that participants would learn more working together face-to-face and thus have higher scores after working face-to-face, whereas our data indicate participants performed better in a second, face-to-face collaboration after first collaborating remotely. There are several alternate explanations for the difference in scores. One explanation is that the activities in the second task were inherently more difficult to perform remotely than face-to-face. Another explanation is that collaborating remotely first provided more time for participants to independently learn to operate the system. Therefore, when subsequently working face-to-face, they could better interpret their partner’s actions vis-à-vis the system, and allocate task responsibilities more efficiently. These possible explanations cannot be eliminated at this point. Further analysis of both the lab reports and the recordings of the sessions, and/or replication of the study using a Solomon four-group design to

\(^6\)A Pearson correlation test was used because it tests the systematic correlation between two continuously scaled variables [Cohen and Cohen 1983].

\(^7\)Note, a Pearson correlation test yielded no statistically significant correlation between lab completion times and grades, between times, grades and condition, between times, grades and order, or between times, grades, condition and order.
Table VI. Interview Analysis: Participants' Comments on Remote Collaboration Compared to Face-to-Face Collaboration

<table>
<thead>
<tr>
<th>Disadvantage</th>
<th>Significance, Coping strategy, or Relative advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction less personal</td>
<td>Doesn’t matter for this work</td>
</tr>
<tr>
<td>Fewer cues from partner</td>
<td>Need to talk more frequently and descriptively</td>
</tr>
<tr>
<td>Some tasks are more difficult</td>
<td>Easier to explore system &amp; ideas independently; Having identical views of data visualization is better; Working simultaneously on the data visualization increases productivity</td>
</tr>
</tbody>
</table>

obtain data from two consecutive face-to-face and remote sessions is needed to provide additional insights regarding any possible task or condition order effect. We also looked to the post-interview data for insights.

5.3 Participants’ Perceptions of the Scientific Process: Post-Interview Analysis

Hypothesis H2 proposes that participants would find working remotely more difficult than working face-to-face. Analysis of the interviews provided only partial support for this hypothesis. As expected, participants reported disadvantages to collaborating remotely. However, participants also reported that some of these disadvantages are not significant in scientific work contexts, and that coping strategies, or work-arounds, can reduce the impact of other disadvantages. Furthermore, participants reported that remote collaboration provided several relative advantages compared with face-to-face collaboration (see Table VI).

Similar to previous studies [Olson and Olson 2000; Olson and Teasley 1996], study participants reported remote collaboration was less personal than face-to-face collaboration. When comparing working face-to-face and remotely, participants reported collaborating face-to-face was:

- more personal
- easier to express yourself
- [we] did more chatting [face to face]

However, participants also reported that lacking this type of interaction when working remotely did not seem to have a negative impact on their work. The impersonal nature of remote collaboration increased their productivity and facilitated collaborative intellectual contributions. As participants explained:

If we were...working side by side, we might tell more stories or something like that...[However] if you’re trying to get something done, sometimes the stories and stuff can get in your way.

It does make for a less interpersonal experience if you’re not working right beside someone...but [when working remotely] I had time to figure things out for myself instead of [my partner] just doing it and me just accepting what he was doing, or me doing it and him accepting what I did. This time [working remotely], we both got to figure it out and say ‘hey, look at this’ in collaboration.
I think that being in separate rooms helps a little bit because it's more impersonal... [You] just throw stuff back and forth more easily.

Participants also reported that when working remotely they received fewer implicit cues about what their partner was doing and thinking. Similar to previous research [Clark, 1996], the study participants explained that without these cues, it may be difficult to follow social interaction norms and assist your collaborator:

[when collaborating face to face] it was a lot easier to ask questions of each other... since you have a feeling [about] when to interrupt them... if you're in the same room... you'll wait [to ask a question] until the other person is not doing as much or not doing something very specific

It is hard to get the context of any question that's asked because you're not paying attention to what the other person is doing because they're in a little [video-conferencing] screen.

To compensate for this lack of cues, several participants reported they needed to talk more frequently and descriptively when collaborating remotely. Participants reported:

Even though we were in separate rooms, it kind of seemed like there was more interaction compared to being face-to-face, which seems kind of strange... It just seemed more interaction was expected... Maybe needed.

We had a really good interaction [when collaborating remotely]... You're conscious that you're not together and you can't see [some things, and] so you think more about [interacting. For example, you think] 'I need to let this person know that I'm about to do this' or 'this is what I'm seeing and I'm trying to let you know so, and you're like doing the same to me'. Yeah, so [our interaction] was probably more. Interaction was really easier. It made [working together] better.

You have to be more descriptive with your words.

Thus to compensate for the absence of implicit cues in the remote condition participants provided explicit cues for their partner. When working remotely, it appears that individuals recognize they do not have a common shared physical reality and subsequently may not have a shared cognitive reality. However, humans are intrinsically motivated to develop a shared reality [Schutz and Luckman 1973, 1989]. Subsequently, study participants developed and adopted a strategy of providing explicit cues to their partner, to develop a shared reality. These explicit cues appear to be joint actions [Clark 1996] that help coordinate activities between participants. The cues may contribute to a faster and more accurate formation of common ground and mutual understanding.

It is interesting to note that even with the disadvantages of remote collaboration and the need for coping strategies, many participants reported they
could work and assume the roles similar to those they typically assume when collaborating face-to-face. Participants commented:

[Collaborating remotely] was just like if we had to sit down and do a group project and we were sitting right next to each other.

I tend to naturally take on the role of coordinator. So if anything seems like it’s not getting done fast enough, I’ll go and say, ‘Well, you need to do this’ or ‘I need to do that.’ So I think I did this [collaborating remotely] because I do that with everything I do.

Schutz and Luckman [1973, 1989] suggest that when developing a shared reality or acting within the context of different realities, individuals assume that differences will not keep them from achieving their goals. In Schutz and Luckman’s terms, individuals assume there is a congruence of relevance systems. This assumption may explain why participants assumed similar roles as if working face-to-face and could be successful working remotely.

In addition to receiving fewer cues from a partner when collaborating remotely, participants also reported that some physical tasks are more difficult. These tasks include creating and sharing sketches of scientific structures, manipulating mathematical equations, and jointly using shared applications in NetMeeting. Some of these problems may be remedied by including more tools in the systems, such as MATLAB®. Others may be remedied by advances in technology, such as shared applications that support multiple pointers and use optimistic concurrency for floor control. Participants explained:

[When collaborating face to face] you could draw more easily, communicate diagrams more easily, and you could look at the other person and see their level of understanding more easily.

The thing that frustrated me the most [collaborating remotely] was the shared applications [NetMeeting]... you could see the other person doing things but you couldn’t do anything [simultaneously.]

I caught myself pointing at my screen sometimes but [my partner] couldn’t see my finger pointing at the screen.

Although technology made some tasks more difficult, study participants also reported that the collaboratory system provides advantages over collaborating face-to-face. These advantages include the ability to work independently as well as collaboratively, having identical and unconstrained views of the data visualization, and working simultaneously with the data visualization.

I liked that we were separate. I think it gave a whole new twist on the interactions, and if one of us got snagged up with something the other could independently work and get it done rather than both of us being bogged down by having to work on it simultaneously.

I think the technology helped the interaction... because...one person could do a task and then the other... has the chance to say, ‘OK, well maybe we can do it this way.’
Table VII. Mean Questionnaire Responses for Collaboratory Adoption Attributes

<table>
<thead>
<tr>
<th>Adoption Attribute</th>
<th>FtF (n = 40)</th>
<th>Remote (n = 40)</th>
<th>FtF Session 1 (n = 20)</th>
<th>Remote Session 1 (n = 20)</th>
<th>FtF Session 2 (n = 20)</th>
<th>Remote Session 2 (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative advantage</td>
<td>4.13</td>
<td>4.05</td>
<td>3.94</td>
<td>3.83</td>
<td>4.31</td>
<td>4.27</td>
</tr>
<tr>
<td></td>
<td>(0.60)</td>
<td>(0.72)</td>
<td>(0.54)</td>
<td>(0.87)</td>
<td>(0.61)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>Compatibility</td>
<td>4.15</td>
<td>4.20</td>
<td>3.97</td>
<td>4.20</td>
<td>4.33</td>
<td>4.19</td>
</tr>
<tr>
<td></td>
<td>(0.64)</td>
<td>(0.60)</td>
<td>(0.60)</td>
<td>(0.66)</td>
<td>(0.64)</td>
<td>(0.55)</td>
</tr>
<tr>
<td>Complexity</td>
<td>1.26</td>
<td>1.30</td>
<td>1.41</td>
<td>1.25</td>
<td>1.10</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td>(0.62)</td>
<td>(0.75)</td>
<td>(0.61)</td>
<td>(0.78)</td>
<td>(0.62)</td>
<td>(0.73)</td>
</tr>
<tr>
<td>Trialability</td>
<td>4.10</td>
<td>3.89</td>
<td>4.30</td>
<td>3.78</td>
<td>3.90</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>(0.80)</td>
<td>(0.82)</td>
<td>(0.49)</td>
<td>(0.96)</td>
<td>(1.00)</td>
<td>(0.65)</td>
</tr>
<tr>
<td>Observability</td>
<td>3.42</td>
<td>3.50</td>
<td>3.38</td>
<td>3.45</td>
<td>3.47</td>
<td>3.55</td>
</tr>
<tr>
<td></td>
<td>(0.85)</td>
<td>(0.72)</td>
<td>(0.83)</td>
<td>(0.77)</td>
<td>(0.89)</td>
<td>(0.68)</td>
</tr>
</tbody>
</table>

Sometimes when you’re working side by side with somebody, you have to deal with ‘Well, you’re looking at [the data] from a different angle than I am, and so you’re seeing a different perspective there.’ Now [working remotely] we could both of us be straight on, having the exact same perspective from where we’re sitting. It made it easier.

[My partner] could be changing the light focusing somewhere, while I could be zooming or moving [the plane] around. And that was really helpful because you’re thinking, ‘OK, as soon as I’m done moving the light I want to go ahead and shift [the plane]…[to be able to] say to [my partner], ‘Why don’t you [shift the plane] while I’m shining the light,’ was really cool. It was really helpful.

The participants in this study reported experiencing disadvantages of remote collaboration similar to those that have been previously reported in the literature. However, the study participants also reported that some disadvantages had minimal impact on their scientific work, and that they developed and used coping strategies to compensate for disadvantages. In addition, they perceived remote collaboration to provide some advantages relative to face-to-face collaboration. They also reported that collaborating remotely was compatible with their previous ways of collaborating face-to-face. These findings elucidate our null result regarding scientific outcomes. Next we look at our data on innovation adoption.

5.4 Collaboratory Adoption: Post-Questionnaire Data Analysis

Analysis of the collaborative adoption post-questionnaire data (Table VII) yielded no support for hypothesis H3. We performed a multivariate analysis of variance (MANOVA) using a general linear model to investigate whether differences in the adoption questionnaire responses can be attributed to condition, that is, working face-to-face or working remotely, or to any interaction effect between condition and order—working face-to-face first or remotely first.
Table VIII. Multivariate Analysis of Variance (MANOVA) of Collaboratory Adoption Attributes Responses Testing for Differences due to Condition and Order

<table>
<thead>
<tr>
<th>Adoption Attribute</th>
<th>Differences due to Condition (FtF vs. Remote)</th>
<th>Differences due to any Interaction Effect between Condition and Order (FtF first or Remote first)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>F</td>
</tr>
<tr>
<td>Relative advantage</td>
<td>2.38</td>
<td>0.47</td>
</tr>
<tr>
<td>Compatibility</td>
<td>2.38</td>
<td>0.30</td>
</tr>
<tr>
<td>Complexity</td>
<td>2.38</td>
<td>0.03</td>
</tr>
<tr>
<td>Trialability</td>
<td>2.38</td>
<td>3.23</td>
</tr>
<tr>
<td>Observability</td>
<td>2.38</td>
<td>0.57</td>
</tr>
</tbody>
</table>

The results (Table VIII) indicate null result. The differences in questionnaire responses due to condition are not statistically significant (at the \( p < .05 \) level). That is, participants’ perceptions of the system’s relative advantage, compatibility, complexity, trialability and observability were not significantly different from their perceptions after using the system face-to-face.

The data analysis indicates there is only one statistically significant difference in questionnaire responses due to the interaction between condition and order. This difference is for relative advantage (\( p < .01 \)). Participants’ mean score for relative advantage was always greater after their second lab session, irrespective of the order of conditions.

Similar to the analysis of the face validity questions, this analysis treats the scores of each participant independently of his or her partner’s score. To examine this approach, we tested the correlations of the score residuals within pairs. The correlations were low: 0.09, 0.18, 0.15, 0.52, 0.13 for relative advantage, compatibility, complexity, trialability and observability responses, respectively. These low correlations, especially in conjunction the low correlations found in the residuals of the validity responses, justify this approach.

The null results are surprising because intuition would suggest that participants would perceive that the system provides fewer relative advantages when working remotely and that using the system face-to-face would be more compatible with participants’ existing work patterns, norms and values primarily developed from face-to-face experiences. Furthermore, we expected the system would be perceived as being less complex when working face-to-face because a partner who could provide assistance was collocated, and that participants would not be able to observe their partner as well remotely as face-to-face. While surprising, the results are consistent with the analysis of the interview data which did not reveal strong negative attitudes towards using the system remotely, and, in fact, revealed some positive attitudes towards using it remotely.

5.5 Limitations

This study has several limitations. One limitation is the repeated measure design. A Solomon four-group design would have allowed additional comparisons among data from two consecutive face-to-face sessions and two consecutive remote sessions. These comparisons could potentially increase our understanding.
of the differences between working face-to-face and remotely, including differences caused by varying the order of working face-to-face and remotely, and the impact of any differences between the first and second task. However, a Solomon four-group design would have required substantial additional resources.

A second limitation can be found in our population sample. We used upper level undergraduate science students, one segment of the overall population who conduct scientific research and are potential collaboratory users. Graduate and undergraduate research assistants, postdoctoral fellows and faculty also comprise this population. However, due to the small number of individuals in these groups locally, the variance in their scientific knowledge, and the demands on their time, we did not include them in our population sample. The entire participant sample for the ongoing ethnographic study of the collaboratory system is taken from this working scientist population. In that study we will conduct interviews and make observations, gathering data similar to that collected during the controlled experiments. The presence or lack of correlation between these data will help confirm or refute the validity and reliability of the current study.

A third limitation is sample size. We had 20 task sessions, a minimum adequate for statistical analysis purposes; however, data from additional task sessions may yield differences not present in the current data. This sample size limitation is linked to our design. If we had designed a simpler task that required less time, more task sessions and study participants could have been included. However, natural science takes time; scientists who use the nanoManipulator report they use it from 2 to 24 hours at a time. We decided that asking participants to complete a sequence of tasks based on scientists’ actual work and work patterns should take precedence over using a larger sample size.

To explore this sample size limitation, we conducted a post-study power analysis. The analysis suggests that a sample size of 30 pairs of participants generating lab report scores with a distribution identical to that in this study would be needed to show that differences in scores due to condition are statistically significant. Similarly, an average difference of 7.4 points (on a hundred point scale) in scores between conditions would be needed for a sample size of 20 pairs to yield statistically significant results. In our study the difference between the mean scores due to condition is 5.65 points. The practical significance of a decrease in performance in this range (under 10%) for remote collaboration would need to be evaluated on a case-by-case basis; such a performance penalty may or may not be acceptable to reduce travel costs and/or time delays.

We also performed a post-study power analysis of the collaboratory adoption post-questionnaire data. The sample sizes needed to show statistical significance in responses due to condition are 322, 752, 5262, 49 and 266 for relative advantage, compatibility, complexity, trailability and observability, respectively. Sample sizes of 53, 203, 280 and 178 would be needed to show statistical significance due to condition and order for compatibility, complexity, trailability and observability. These numbers are based on the assumption that the distribution of responses from the larger samples would be identical to the
distributions in this study. An average point difference of 1.8 to 2.4 points in post-questionnaire responses (on a scale of 1 to 5) would be required with a sample size of 40 participants to yield statistically significant results. The largest point difference in this study was 0.52. Given these results and inherent limitations of controlled studies, the issue of adoption is best investigated in longitudinal field studies, such as the one we currently have underway, that investigate the use and non-use of collaborative systems in scientists' labs.

A fourth limitation focuses on the tasks. Although the tasks are representative of natural science data collection, analysis and interpretation, they do not encompass the entire life-cycle of the scientific process. For example, problem formulation, research design and research dissemination were not included in the tasks. Furthermore, the tasks in session 1 and 2 differed. Although designed to be similar in complexity, additional investigation may uncover aspects of the tasks that are inherently impacted by an interaction condition.

6. CONCLUSIONS AND FUTURE RESEARCH

This study produced a null result; conclusions from this study need to be interpreted with caution. The data from the scientific task outcome measures, post-interviews and collaboratory adoption post-questionnaire do not support the hypotheses that working remotely would be less effective and more difficult than working face-to-face, or that working remotely would have a negative impact on participants' perceptions regarding innovation adoption.

To summarize, task performance when collaborating remotely appears equivalent to task performance when collaborating face-to-face. Mean scores for the first task session were the same for the face-to-face and remote conditions, and there was no statistically significant difference between lab scores for the two conditions (without considering order). Additionally, collaborating face-to-face before collaborating remotely did not produce the anticipated positive impact on the scientific outcomes or process. Pairs who collaborated remotely first had statistically significant higher lab report grades for Task Session 2 than those who collaborated face-to-face first. Interview data revealed that participants could identify both unique advantages and disadvantages of the technology and that many of them were able to develop coping strategies to reduce the impact of the disadvantages that emerged when collaborating remotely. Lastly, the only statistically significant difference in questionnaire responses between the face-to-face and remote conditions with respect to relative advantage, compatibility, complexity, trialability and observability was that relative advantage rated higher after the second task session.

Thus, the multiple types of evaluation data combine to illustrate how, in this particular study, the system mediates collaborative scientific work processes without a statistically significant negative impact on scientific data collection and analysis task outcomes or participants' attitudes toward adoption. While the data analysis produced null results, considered as a whole, the analysis leads us to conclude there is positive potential for the development and adoption of scientific collaboratory systems. Participants were able to adequately complete scientific work when collaborating remotely,
and they readily developed and used strategies to compensate for system deficiencies.

Schutz and Luckmann’s theory of the life world [1973, 1989] can be used to explain some of the behaviors and responses we saw and provide theoretical support for our conclusion. Working remotely can be considered an example of a problematic situation in which an individual cannot assume his or her physical world is the same as the physical world of her or his collaborator’s. However, humans have a desire to develop a shared reality. Although individuals may have different types and degrees of motivation in establishing a shared reality, we strive to assume a shared reality, an intersubjectivity, at least to the degree necessary for current purposes [Clark, 1996].

When developing a shared reality or acting within the context of different realities, Schutz and Luckmann propose that individuals assume that differences will not keep them from achieving their goals. That is, individuals assume there is a congruence of relevance systems. Schutz and Luckmann further propose that individuals assume that if you were with me, you would experience things the same way I do, that is, individuals assume there is an interchangeability of standpoints.

When working remotely, participants’ different physical locations and the system’s limitations in fully and accurately representing the remote location may provide strong evidence that causes participants to believe they do not have a shared reality. However, as humans, they are motivated to develop a shared reality. Subsequently, they seem willing to proactively work to develop it, and appear to assume that the physical location differences will not keep them from completing their tasks (congruence of relevance systems). For example, no study participant reported that they could not do science when working with their partner remotely. This is especially interesting considering that 75% of the study participants had not worked with their partner previously. The participants appear to further assume there is an interchangeability of standpoints. They take explicit joint actions to develop a shared reality, using language to share their experiences and standpoint. For example, participants said that when collaborating remotely they discussed what they were currently doing with their partner more frequently and in greater detail than when working face-to-face. These explicit joint actions may help to create a shared reality and assist in task performance. The joint actions compensate for a lack of physical collocation and for limitations in the system’s ability to represent the remote physical location fully and accurately.

In comparison, when working face-to-face, individuals may, perhaps erroneously, assume that a shared reality already exists, or that it is more comprehensive than it really is, because there is a shared physical location. The shared physical location helps individuals believe there is also a shared reality. Knowledge about each other gained through the interactions that commonly occur in face-to-face situations may also reinforce the perception of an existing shared reality. For example, the study participants reported they have more personal interactions when collaborating face-to-face. Personal knowledge about a collaborator and a shared physical location may influence or strengthen an individual’s assumptions about a shared reality, and subsequently reduce
the type and number of joint actions whose purpose is to develop a shared reality.

Rogers' [1995] theory of innovation diffusion defines attributes of technology that are important in general for innovation adoption and, perhaps, in the case of computer supported cooperative work, also important for the establishment and maintenance of a shared reality. Compatibility with individuals’ values, experiences and needs, as well as with individuals’ need to develop a shared reality, may be necessary for current and future collaboratory systems. If advantages over working face-to-face are not provided, individuals may not be sufficiently motivated to use technology and devise ways to overcome its deficiencies in order to create a shared reality and accomplish the tasks at hand. Without minimal complexity and support for trialability, individuals may not be able to easily experiment with the system or understand it. This experimentation and understanding can facilitate individuals’ ability to develop new strategies and joint actions to compensate for the lack of implicit cues and tasks made more difficult by the technology. Observability—observing how others use the innovation—can also reduce the effort needed to learn new strategies to compensate for lack of implicit cues and other tasks made more difficult by the technology. Thus we propose that when designing to support collaboration across distances, compatibility, complexity, advantages, trialability and observability to support the establishment and maintenance of a shared reality should be considered.

The results of our study also appear to support the use of collaboratories in science education as well as scientific research. Collaboratory systems have the potential to allow students at universities, community colleges and high schools without local access to specialized, state of the art scientific instruments to access scientific instruments remotely and conduct experiments in collaboration with students, faculty and staff at institutions that have them. An example of this type of application is the Bugscope educational outreach program for K-12 classrooms [Thakkar et al. 2000]. This project allows a class to remotely view their collection of bugs through a scanning electron microscope accessed over the Internet. The results reported in this paper suggest students could also successfully collaborate across distances to conduct complex scientific labs and projects.

As previously mentioned, these conclusions and implications need to be interpreted with caution. Replication of the controlled study, ideally with a larger number of study participants and over the four possible combinations of conditions, is needed to confirm or refute the conclusions and implications of this study.

Our future work plans include an analysis of videotapes from task sessions to investigate process differences and similarities between face-to-face and remote collaboration. We are currently conducting a longitudinal field study to investigate whether the results reported here are comparable to behavior and attitudes in professional scientific contexts. In the field study, the collaboratory system has been provided to scientists who have expressed an interest in conducting scientific investigations using the system. We are periodically conducting interviews and observations to learn about scientists’ use and non-use.
of the system and the scientists’ perceptions of the system’s impact on scientific outcomes and process over time. We plan to investigate the similarities and differences between scientists’ and study participants’ perceptions and use of the system to further our understanding of the impact of collaboratories on the scientific process and outcomes.

ACKNOWLEDGMENTS

Our thanks to the study participants; to Martin Guthold, Richard Superfine and Dorothy Erie who generously shared their natural science expertise and data in developing the scientific tasks and lab reports; to Leila Plummer, Ron Bergquist and Atsuko Negishi who helped run the experiment sessions; to Bin Li and Anita Creszenzi who helped code interviews; to the team who have and continue to build the nanoManipulator, including Frederick P. Brooks, Jr., Aron Helser, Tom Hudson, Kevin Jeffay, Don Smith, Russell M. Taylor II; to Christopher Wiesen who provided expert assistance with the statistical analysis; and to the anonymous reviewers whose comments helped us improve the final presentation of this work.

The development of the nM and this work have been funded by the NIH National Center for Research Resources, NCRR 5-P41-RR02170. The nanoManipulator project is part of the Computer Graphics for Molecular Studies and Microscopy Research Resource at the University of North Carolina at Chapel Hill.

REFERENCES


Evaluating a Scientific Collaboratory: Results of a Controlled Experiment


Received August 2001; revised May 2002; accepted February 2003