Emotional and Attitudinal Responses to Remote Versus Co-located Usability Testing

Thesis

Submitted to

The College of Arts and Sciences of the

UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for

The Degree

Master of Arts in Psychology

by

Michael J Goliber

UNIVERSITY OF DAYTON

Dayton, Ohio

July, 2005

APPROVED BY:

Dr. Kenneth A Graetz, Ph.D.
Chairperson, Thesis Committee

Dr. Donald J. Polzella, Ph.D.
Thesis Committee Member

Dr. David W. Biers, Ph.D.
Thesis Committee Member

CONCURRENCE

Dr. David W. Biers, Ph.D.
Chairperson, Department of Psychology

ABSTRACT

EMOTIONAL AND ATTITUDINAL RESPONSES TO REMOTE VERSUS CO-LOCATED USABILITY TESTING

Goliber, Michael, J.
University of Dayton, 2005

Advisor: Dr. Ken A. Graetz

Current usability testing is often conducted via face-to-face interactions. This method can be costly, both in terms of timelines and budget. However, remote usability testing has been shown to be a viable alternative, in that performance scores have been shown to be quite similar to face-to-face methods. Although performance appears similar, remote usability testing may present challenges that threaten the validity and reliability of usability testing results. Rather than focusing on the performance of users in remote versus co-located conditions, the proposed study investigates the emotional and attitudinal responses of users engaged in software usability tests. The purpose of this study was to compare users' anxiety and satisfaction with communication in remote and face-to-face usability tests. It was hypothesized that participants in the remote condition would exhibit a lower level of anxiety and be less satisfied with the communication method. Multiple usability tasks were administered and measures were recorded at three time intervals. Responses on the Social Anxiety Thoughts (SAT) questionnaire and the Communication Satisfaction Inventory (CSI) were collected. Although there were no significant differences between the groups in terms of anxiety

and communication satisfaction, methodological limitations may have prevented the detection of differences and additional research is required to explore the strengths and weaknesses of remote usability testing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The concept of usability is applicable to any interaction that involves a human attempting to use a machine via some interface. This could be a person using something as simple as a can opener or as complex as a software application to debug a programming language. Many attributes of the interface are considered when attempting to understand the usability of a product, from color and layout, to ease of use and navigational design. The investigation of usability, an important segment of human factors research, began in the pre-World War II era with a focus on improving production lines and fitting humans to job type (Meister & O'Brien, 1996). During WWII, human factors testing and evaluation shifted toward pilot selection and training, and later to the modification of aircraft cockpits for the improvement of pilot performance and safety. After WWII, usability testing became a staple in the design world, bringing it more directly and centrally into the full system design process. The benefits of thorough usability testing are numerous. It can reduce development costs and prevent costly errors that increase distribution costs and support (Mauro, 1994). It can also improve marketing by influencing early adopters of new technology and reducing perceived training costs (Conklin, 1991).

Usability research has been conducted as an integral part of the design of televisions, VCRs, telephones, digital oscilloscopes, spreadsheets, and automobile

navigation systems (Dumas & Redish, 1993). More recently, computer technology has been the focus of usability testers. With ever increasing access to and usage of computers, testing the usability of computer hardware/software has become an enormous task. Not only does the computer user encounter hardware usability issues everyday, but the productivity and Internet applications delivered via computers present similar usability challenges. The present study compared two methods for conducting software usability testing: co-located versus remote.

Typically, usability testers are in the same room with users as they attempt to interact with a software application. The tester is on hand to guide the user, obtain information (via non-verbal cues, user comments, etc), and distribute questionnaires. Today, the widespread availability of collaborative networking tools or groupware that allows for real time, computer-mediated interaction promises to revolutionize the practice of usability testing over the next decade. Since the advent of the Internet, and its recent acceleration in speed, information can be transmitted all over the world with the click of a mouse. Recently, programs have been developed that allow users across the globe to share information with each other using collaborative whiteboards, video/audio conferencing, and other interactive tools. Such tools could be used to conduct usability tests remotely.

Tools are currently available that allow usability testers to observe the screen movements of remote users over an Internet connection. From their remote locations, users can be interacting with a software application, while communicating with the usability tester over the telephone, a videoconferencing connection, or via some computer-mediated channel (e.g., electronic chat, voice over IP). From his/her

computer, the usability tester can observe every mouse movement and every click made by the remote user in real time. Although the economic benefits of remote usability testing over the Internet are obvious, separating the user from the usability tester may have some impact on the response of the user. These ramifications of remote usability testing merit careful study.

Rather than focusing on the performance of users in remote versus co-located conditions, this study investigated the emotional and attitudinal responses of users engaged in software usability tests. This aspect of usability testing has been relatively under-investigated, but is of critical importance as more usability testing is conducted remotely. The major research question in this study centered on the hypothesized positive and negative emotional effects of remote usability testing. Specifically, it was hypothesized that users in a remote usability test will experience less anxiety than users in a co-located test. However, it was also predicted that remote users will be less satisfied with their interaction with the tester than co-located users. The remainder of this introduction provides background information related to the definition of usability, traditional usability testing methods, and remote usability testing. It concludes with a more detailed description of this study and the theoretical rationale for its hypotheses.

Defining Usability

Although the general concept of usability is easy to understand, a more rigorous definition is required for the purpose of this current research. A usable interface has been defined as one with the capability to be used by humans easily (to a specified level of subjective assessment) and effectively (to a specified level of human performance) (Galitz, 1997). This combination of positive evaluation and performance is a critical.

Not only do users need to perform well, they also need to perceive the software as usable.

Dumas and Redish (1993) and Nielsen (1993) have developed relatively specific assumptions about usability. Dumas and Redish (1993) assume the following:

1. Usability means focusing on the users.
2. People use products to be productive.
3. Users are busy people trying to accomplish tasks.
4. Users decide when a product is easy to use. (p. 4).

Dumas and Redish stress that any definition of usability needs to take these points into consideration. Nielsen (1993) argues that there are five attributes of software usability: learnability, efficiency, memorability, errors, and satisfaction. According to Nielsen, a learnable system is easy to learn and allows for rapid adoption. Once learned, an efficient system is one that can be used to achieve a high level of productivity. Memorable systems allow users to return to that system after a significant period of separation without having to relearn the system. Usable systems should also demonstrate low user errors and allow users to recover from errors easily. Additionally, Nielsen contends that catastrophic errors must not occur in usable systems. Finally, the system should be pleasurable to use and lead to a high level of user satisfaction. Presumably, the combination of all of these attributes in any software interface will result in a highly usable product. Other approaches also emphasize these attributes. Hix and Hartson (1993), for example, describe a combination of five user-oriented

characteristics that contribute to usability: ease of learning, high speed of user task performance, low user error rate, subjective user satisfaction, and user retention over time. Although all of these approaches emphasize both user satisfaction and user performance, the latter has been the traditional focus of usability testers.

Usability Testing

Rubin (1994) describes usability testing as a tool, deeply rooted in empirical methodology, which can range from quantitative (large sample sizes and objective measures) to qualitative (small samples and open-ended protocols). Traditional software usability testing often requires that users interact with the application while under observation by a tester and that they respond to questions about the application posed by the tester. The actual process often varies across tests, depending on what application is being tested.

Dumas and Redish (1993) list the following defining characteristics of usability tests:

> 1. The primary goal is to improve the usability of a product. For each test, you also have more specific goals and concerns that you articulate when planning the test.
> 2. The participants represent real users.
> 3. The participants do real tasks.
> 4. You observe and record what participants do and say.
> 5. You analyze the data, diagnose the real problems, and recommend changes to fix those problems. (p. 22)

There are three types of typical usability testing situations which involve co-location of users and the experimenter at the same site. The first two occur at the software developer's site. In one variation the experimenter sits right beside or next to the user ("over-the-shoulder") in the same room with the user and experimenter in direct communication. In the other variation, the experimenter sits behind a one-way mirror with the user alone in the next room. Communication in this case is via audio microphone and speakers. In some sense this situation is akin to remote testing (see below), with the exception that the user and experimenter are co-located. Despite differences in possible anxiety generated in these two situations, Hackman & Biers (1992) demonstrated that the presence of the experimenter in the same room versus an adjacent room had no effect on performance or the number and quality of thinking out verbalizations.

The problem with usability testing at the software developer's site is that the users are not often representative of the target population and the context for evaluation is contrived. Even though the on-site location may have a technologically advanced usability testing lab, there is a tendency to utilize participants based solely on their proximity, and not on their appropriate representation of the user group. As a consequence, usability evaluations are often conducted in the field at the customers or end-users site (e.g., contextual evaluation.). In this third co-location situation, the user and experimenter are usually situated in the same room. Although off-site testing is more likely to yield representative users and provide a more appropriate context, it is not without cost, the cost of travel for the software developer's employees.

Involvement of real users is the factor which differentiates usability testing from other forms of usability evaluation such as heuristic evaluation. Dumas and Redish (1993) also point out that real usability tests are conducted on a representative sample of the actual group of users who will be using that product. Conducting a test on programmers when the product to be tested is primarily intended for legal secretaries would not be an appropriate usability test. Similarly, Rubin (1994) describes usability testing as, "a process that employs participants who are representative of the target population to evaluate the degree to which a product meets specific usability criteria" (p. 25). This definition essentially rules out other forms of testing such as expert evaluations that do not require representative users. Expert or heuristic evaluations are, "usually performed by comparing the interface with the human factors criteria listed in the requirements specification and also with other human factors standards and guidelines" (Wickens et al., 1998; p. 66). Typically, this involves a usability professional examining the product to determine significant flaws. Nielsen (1993) suggests at least three, preferably five, professionals should conduct these sorts of tests to construct a reliable evaluation.

In addition to some general assumptions about usability testing, a number of specific techniques have emerged. Verbal reports/think-aloud evaluation involves the user speaking out loud, or verbally relating, his/her experience as he/she uses the application (Karat, 1997). Ericsson and Simon (1984) suggest that short-term memory is accessible by collecting verbal reports. They recommend conducting think-alouds concurrently with task performance when interacting with the software. This is a very effective method for gaining personal insight into the user's mental processes

pertaining to the tested product. Wickens, Gordon, and Liu (1998) describe this method as yielding, "insight into underlying goals, strategies, decisions, and other cognitive components" (p. 60).

A second method is simply observation. Wickens et al. (1998) describes this method as one of the best ways for the usability tester to obtain a solid understanding of task performance. By observing a user perform different tasks, under different scenarios, the tester can more fully understand that which the user might have difficulty describing. This method can be videotaped for later examination.

A third method, cognitive walkthroughs, (also referred to as "structured walkthroughs"; Dumas & Redish, 1993) are essentially question and answer sessions where "Users make guesses about what actions to take by comparing the expected outcome of the action towards their goals" (Dumas & Redish, p. 68). The users utilize this information as they move forward toward their goals. Essentially, the users are answering questions about each of the decisions they must make as they interact with a product. These questions relate to identifying the goals, the ease of identifying the outcome of a decision, and the ease of user evaluation of their progress towards their goals. Dumas and Redish point out that each set of questions is asked in correspondence with each task to be performed.

Although various methods have evolved, the testing context has remained virtually constant over the years. "Traditional user interface evaluation usually is conducted in a laboratory where users are observed directly by evaluators" (Hartson, Castillo, Kelso, Kamler, & Neale, 1996, p.1). This usually includes an observation room, a video camera, computer, monitor, and any other related PC device. "Much

traditional user interface evaluation is conducted in usability laboratories, where a small number of selected users are directly observed by trained evaluators" (Castillo, Hartson, & Hix, 1999, Remote Usability Evaluation At A Glance, p. 1). This method of testing puts the evaluator in the same room as the user and affords presumably uninterrupted communication. The evaluator can easily pose key questions to the user such as, "Why did you click that button?" and "Where did you expect that link to take you?"

Although this is the most common format for software usability testing, co-located testing does present a number of problems. The first problem involves access to appropriate participants. In many instances, experienced users are hard to find or cannot schedule time to participate in usability tests. Thus, usability testers often rely on users who are not as familiar with the system as others, simply because they are easier to access. This may result in errors due to lack of familiarity with the system versus poor system usability. A second problem is small sample size. Due to time constraints and product deadlines, typical industry practices often do not allow the usability tester to run a large number of co-located tests, which can take precious time and resources. Small sample sizes reduce statistical power and increase the likelihood of observing aberrant responses. The third problem is the cost of co-located usability testing. Factoring in travel costs and personnel downtime, the testing of expert users can be very costly. As an example, testing a software interface that is part of a nuclear defense system might involve traveling to the user's site, gaining access to a highly restricted area, and taking a skilled user away from his/her post for a relatively long period of time.

Although most developers can justify the high monetary cost of careful usability testing, it is more difficult to justify the time it takes to conduct co-located usability tests. The benefits of usability testing can be substantial, if properly put into figures (both in terms of productivity, and monetary). For example, Wixon and Jones (1996) attributed an 80% increase in product revenues to usability testing that improved a user interface. Similarly, IBM states, "For developers and manufacturers, the advantages of creating usable products far outweigh the costs. The rule of thumb: every dollar invested in ease of use returns $10 to $100" (IBM, 2001, Making it easy, p. 1). On the other hand, thorough, co-located usability testing can slow production cycles dramatically. In today's software industry, this delay may be unacceptable and may result in untested software with numerous usability problems being shipped to consumers. Clearly, a new method for conducting usability tests that makes it easier to access a relatively large number of expert users quickly and cost efficiently would be attractive to software developers. This is motivating interest in remote usability testing.

Remote Usability Testing

One alternative to the co-located usability test is remote evaluation or collaborative remote evaluation (Hartson et al., 1996). With this type of evaluation, the experimenter and the user are no longer located within the same physical space. By utilizing an Internet connection, the user and evaluator can be electronically synchronized in order to administer the usability test. With such aids as Microsoft NetMeeting®, Timbuktu Pro®, WebEx®, CuSeeMe®, and Lotus Sametime®, the two parties can share applications, as well as have real-time audio and video conferencing

capabilities. Testers can monitor mouse movements and other user activity as users interact with the software.

Hartson et al., (1996) have defined many different approaches to remote evaluation, including remote questionnaire/survey, remote-control evaluation, and videoconferencing as an extension of the usability laboratory. In the remote questionnaire/survey method, users interact with a software application and are periodically queried by automated questions, the responses to which are sent to testers electronically. In remote-control evaluation, the user and evaluator are separated in space and time, and their computers are linked via the Internet. In videoconferencing, users are located remotely (i.e., in a different state or country) and "can be connected to evaluators using network and video conferencing software, as an extension of the video/audio cable between user and evaluator" (p.4).

Although very few studies have been conducted to date, research indicates few differences between co-located versus remote contexts when examining traditional performance variables such as error rates (Hartson et al., 1996; Macko, 1998). Macko showed that remote and co-located users scored almost identically on experience variables (e.g., statistical knowledge, scenario expertise, and Internet browsing expertise) and forced-choice questions. Hartson et al. found no differences between remote and laboratory participants on nine semantic scales, discovered through questionnaire completion. Their conclusion was that remote evaluation was just as effective as the co-located method and that the co-located method could be replaced by the remote method without significant detriment to outcome. Before completely replacing co-located with remote testing, however, it is argued that a deeper

understanding of the cognitive and emotional responses to both methods is required. That is the focus of the present research.

The Present Study

Today's technology allows usability testers to observe and communicate with geographically distributed users. Is remote usability testing an effective substitute for co-located testing? Although the economic benefits can be substantial, is something lost by moving usability testing to a distributed, online environment? This study focused on the impact of remote usability testing on user emotions and attitudes. Are remote users as comfortable during the usability test as co-located users? Are they less anxious? Is the tester perceived as more detached and disinterested? This study explored the experiences of users involved in remote usability tests, comparing them with users engaged in co-located tests. This research is critical to the understanding of usability testing from a user's perspective and to the adoption of remote usability testing techniques. Much of the past research on usability testing focuses on behavioral outcomes (e.g., errors made by the user during the test). Less attention has been paid to the process and channel through which this data is collected and the impact on the user. This study begins to address this gap in the literature.

This study differs from previous remote usability studies in a number of important ways. First, the current study does not confound remote testing with unattended testing. In the current study, remote software usability testing is operationally defined as a usability test wherein (a) the user and the tester are separated geographically, (2) the user and tester are communicating via a two-way audio channel (e.g., telephone), and (3) the tester can observe the user's screen activity in real time

using application sharing technology. This definition does not include the use of videoconferencing that might allow the remote tester to observe the user's facial responses and body language. Although this may become part of the definition in the near future, this technology is still relatively expensive and not widely available. Thus, it is not a method that could be used by the majority of usability testers today and its inclusion would reduce the ecological validity of the current study. Additionally, research conducted by Lesaigle and Biers (2000) suggests that there is little to be gained by adding on a video communication line. When conducting usability testing, it is most efficient to keep the lines of communication simple. Instead of attempting to add video or some other method, this study focused on a remote testing technique that is currently available and could be used immediately for usability testing.

Second, the current study differentiates users' responses to the remote testing environment, focusing specifically on evaluation apprehension and communication satisfaction. It is argued that evaluation apprehension, or the anxiety produced by a concern over winning a positive evaluation from an observer or avoiding a negative evaluation, is common during usability tests. Typically, usability testers take steps to put the user at ease and reduce this anxiety. In the current study, users' anxiety was measured using the Social Anxiety Thoughts Questionnaire (SAT; Appendix C). In addition, the current study measured users' satisfaction with the communicative process using the Communication Satisfaction Inventory (CSI; Appendix D), a scale developed specifically for the study. Items on this questionnaire address users' overall level of satisfaction with the interaction with the tester, the level of interest and attention shown

by the tester, and the ease with which users were able to respond to the tester's requests and questions.

Finally, the current study predicted differences between remote and co-located conditions. Two hypotheses were advanced. First, it was predicted that users in the remote condition will experience less anxiety than users in the co-located condition. Basically, due to the proximity of the experimenter, remote users will feel less evaluation anxiety and performance anxiety. Second, it was predicted that remote users will be less satisfied with the interaction in the remote versus the co-located condition. Specifically, they will find it more difficult to respond to the tester's requests and questions and they will perceive the tester as more detached and less interested in their performance. Although this study does not focus on the actual behavior of the users during the usability test, it is expected that this study will confirm the findings from previous studies in that no significant differences in performance will be obtained.

These hypotheses are derived from research and theory in social psychology and communication. Usability tests are social interactions wherein a user is asked to perform tasks in the presence of an evaluator. Social psychological research indicates that the presence of an audience can have considerable impact on behavior, a process that is probably mediated by emotional arousal. Research on social facilitation indicates that well-learned behaviors are facilitated by the mere presence of others, while newly learned behaviors may actually be inhibited (Zajonc, 1965). Emotional arousal, most likely experienced as a form of anxiety, has been found to underlie this behavioral phenomenon. Specific to the evaluative social context is the notion of evaluation apprehension, of the anxiety caused by a concern on the part of the individual being

evaluated to win a positive evaluation or avoid a negative evaluation (Cottrell, Wack, Sekerak, & Rittle, 1968)

Although the level of anxiety undoubtedly varies across users, given that (1) many users are asked to engage in tasks that may be somewhat unfamiliar, (2) the user and the evaluator are in very close proximity, and (3) the user's behavior is being closely evaluated, the traditional, co-located usability test can be expected to elicit some level of evaluation apprehension on the part of the user. It is important to note that, unlike most psychological testing, it is the software that is being evaluated in usability testing, not the user. Even if this is communicated to the user, however, it is still likely that the user will internalize the situation, viewing it as an evaluation of his/her ability to use the software. This will lead to evaluation apprehension. In remote usability tests, users can be expected to experience less apprehension as the presence of the evaluator is not as salient. This reduction in anxiety should also allow users to be more open in their responses to evaluator questions.

On the other hand, research suggests that remote communication, whether it's via telephone or electronic chat, may be viewed as more awkward, less intimate, and generally satisfying by participants. Such communication is leaner and may prohibit the exchange of some nonverbal, paralinguistic messages. Unlike structured surveys and other more quantitative methods, the fidelity of the communication channel between the user and the usability tester is a key component of a successful usability test. The channel must allow the tester to identify quickly any problems that the user might be experiencing. Similarly, the user must be able to understand and respond to the requests of the tester quickly. What is of critical importance is that errors caused by

miscommunication not be attributed to the software under examination. Another important aspect of the communication channel is the degree to which it contributes to the perceived relationship between the tester and the user. During a usability test, it is presumably advantageous to create an atmosphere where the user feels comfortable with the tester. When communication occurs remotely, there is the possibility that users will perceive the tester as more detached and less interested in their responses. In the present study, it is expected that remote users would rate the communication during tests as less satisfying than co-located users. Thus, the current study was expected to reveal both positive and negative responses to remote usability testing, further clarifying the impact of adopting this new technique.

CHAPTER II

METHOD

Participants

Forty undergraduate students, 20 male and 20 female, from the University of Dayton served as participants in the study in partial fulfillment of a research requirement in their Introductory Psychology course. All participants were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 2001).

Independent Variables

Co-presence Versus Distance

The major independent variable involved the manipulation of the physical proximity of the experimenter during the usability test. All participants engaged in a usability test with the experimenter serving as the usability tester (hereafter referred to as the tester). Participants were seated at a computer workstation and were asked to engage in various tasks by the tester. Participants were randomly assigned to either the co-located (COL) or distance (DIS) condition. Those in the COL condition engaged in the usability test in the presence of the tester. The tester was positioned immediately to the right and just behind the participant, giving the participant instructions and looking over the participant's shoulder at the computer monitor. The distance between the tester and participant in the COL condition was no more than 2 feet. This condition resembles the typical procedure used in actual usability testing.

In the DIS condition, participants did not meet the tester physically until after the experiment. In this condition, participants were escorted to the computer workstation by a testing assistant, where they communicated with the tester using a hands-free headset. The tester was located in an observation room with a one-way mirror, allowing the tester to see the participant but preventing the participant from seeing the tester. Using a computer in the observation room and an application called Lotus Sametime™, the tester was able to see the activity on the participant's computer monitor in real time. Sametime is a Web-based collaborative tool that allows distributed users to share computer applications in real time. Using Sametime, remote users can enter into a Sametime meeting with other users over an Internet connection using their Web browsers. Once a connection is established, remote users can decide to share any application that is currently running on their computer with the other users. The other users would be able to see on their computer monitor all of the activity from the remote user's screen as the remote user works with the application. This Sametime session was arranged prior to each DIS experimental session.

Subject Variables

Participants' gender and responses to a Computer Usage Questionnaire (Appendix B) developed for this study to assess participants' familiarity with computers in general and with Lotus QuickPlace specifically were collected. The Computer Usage Questionnaire included items related to self-reported frequency of computer use and perceived level of comfort and skill using a standard mouse and keyboard. Participants were also asked to rate their frequency of QuickPlace usage. Although no hypotheses were advanced related to these subject variables, they were

used to explore possible mediating and moderating effects and to detect participants who were unable to use a computer. Any participant who indicated a low level of experience and a high level of discomfort with computers was debriefed and dismissed.

Procedure

The tester introduced participants to the study after completion of the informed consent (Appendix A), the Computer Usage Questionnaire, and the Social Anxiety Thoughts (SAT) Questionnaire, in that order. All participants were then asked to complete two tasks: Task 1 and Task 2. Participants completed the SAT in between tasks and again after the second task. The Communication Satisfaction Inventory (CSI) was administered last, after all tasks and SAT completion. Participants were then debriefed and dismissed. Task 1 and Task 2 were counterbalanced to avoid order effects.

In both conditions, the tester provided instructions verbally. In the DIS condition, participants were seated at the workstation by a testing assistant and asked to wear a headset. From that point, all instructions came via the headset from the tester, located in an adjacent room. As part of the introduction, the tester led the participants through a brief training session on QuickPlace. This involved presenting QuickPlace to the participants, and showing them most of the features. After the training session, the tester began the first task. The verbatim protocols that were used for both tasks are included as Appendices E and F.

In Task 1, the participants were asked to create a new page in QuickPlace that included several elements. An example of this page is included in Figure 1. The participants were asked to give the page a title, and write two sentences about

themselves. The participants were then required to add to the page a hyperlink to a URL provided by the tester. Upon completion of this page, the participant was asked to publish the page, designating the tester as an additional editor and notifying the tester via e-mail that the page has been published. Participants then completed the SAT and moved on to the second task. Task 2 consisted of adding a new Task page to the QuickPlace. An example of this page is included in Figure 2. Participants first selected the "New Task Page" button to begin the task. Participants were required to give the task page a title, and enter the start date, due date, assigned to, and editors. Participants needed to select April 1 as the start date, April 15th as the due date, assign the task to the tester, and ensure that both the author and the tester had edit access. Finally, participants needed to enter two sentences of text describing a task and publish the task to the QuickPlace.

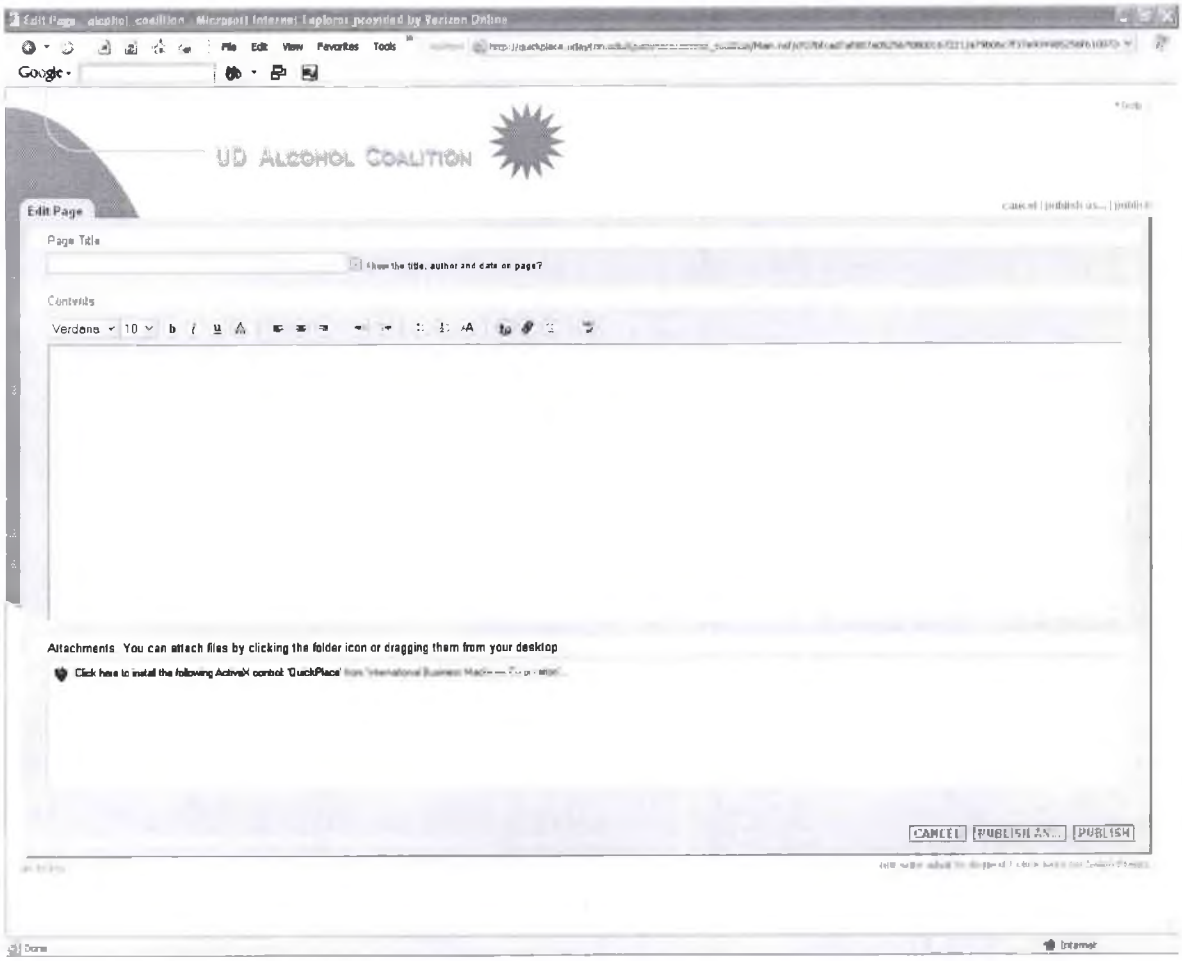

Figure 1. Screen Shot of a New Page in QuickPlace.

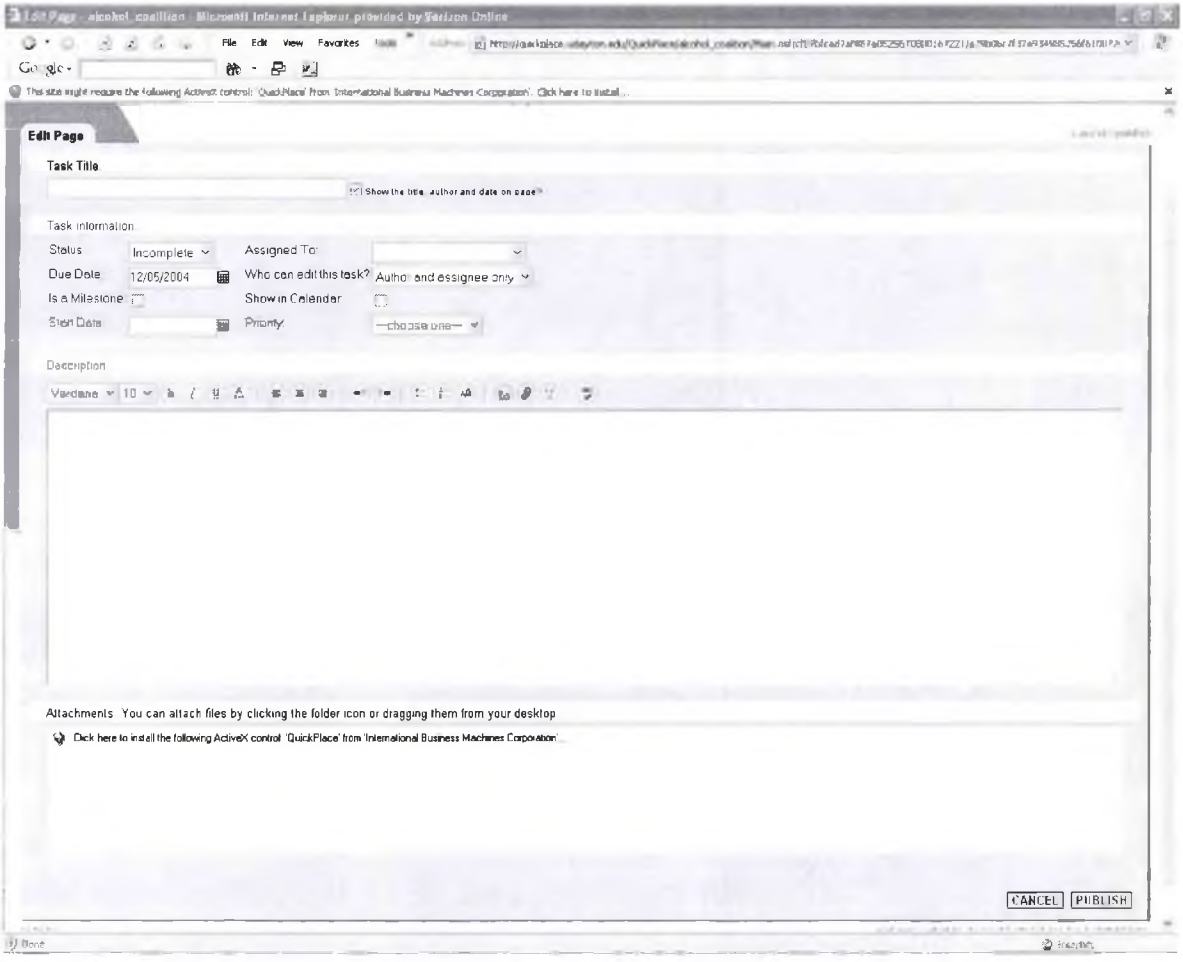

Figure 2. Screen Shot of a New Task Page in QuickPlace.

Following the completion of the second task, participants completed the SAT and the CSI. Participants were then debriefed and dismissed.

Dependent Variables

Social Anxiety Thoughts Questionnaire

A 21-item questionnaire with a five-point scale rating (Never = 1, Rarely = 2, Sometimes = 3, Often = 4, Always = 5) was administered. Scoring entailed summing the individual items, resulting in an overall score with a range of 21 to 105 where higher scores indicated a higher level of social anxiety. This instrument was used by permission of L. M. Hartman, Addiction Research Foundation (1984), and has

demonstrated good reliability and validity in past research, with significant correlations with the Social Avoidance and Distress Scale and the Fear of Negative Evaluation Scale. An analysis of the scores from the current study yielded a Cronbach's Alpha of .95.

Communication Satisfaction Inventory

This inventory is a 20-item questionnaire developed specifically for this study with a five-point rating scale (Strongly Agree = 1, Agree = 2, Neutral = 3, Disagree = 4, Strongly Disagree = 5). After reverse scoring, all individual items were summed, yielding a total score with a potential range from 20 to 100 where higher scores indicated a higher level of communication satisfaction. An analysis of the scores from the current study yielded a Cronbach's Alpha for CSI overall of .86.

Seven CSI subscales were also generated based on a priori construction of test items; comprehensibility, flow, depth, efficiency, respondent's engagement, partner's engagement, and comfort level (see Appendix D for further explanation). After reverse scoring of these subscales, each was individually summed with a higher score meaning a higher level of that subscale. Scores for the subscales of comprehensibility, depth, efficiency, and comfort level range from 4-20. Scores for the subscales of flow, respondent's engagement, and partner's engagement range from 6-30. Although the current study did not provide enough responses to allow for a confirmatory factor analysis of the CSI to support the subscale breakdown, the subscales showed moderate to low internal consistency. The Cronbach's Alphas for comprehensibility, flow, depth, efficiency, respondent's engagement, partner's engagement, and comfort level were

.73, .78, .42, .62, .76, .50, and .63 respectively. These internal consistency measures were in the low to moderate range.

CHAPTER III

RESULTS

Computer Usage

The percentage of participants in the current study describing their knowledge of computers as being Intermediate, Advanced, and Low was 60%, 22% and 18%, respectively. The percentage of participants describing their Internet/Web experience as Intermediate, Advanced and Low was 82%, 10%, and 8%, respectively. Based on their responses to this questionnaire, none of the participants were excluded from the study.

The Effect of Tester Proximity on Anxiety

It was hypothesized that users in the remote testing condition would experience less anxiety than users in the co-located testing condition. Means and standard deviations for SAT at each time period across tester proximity are listed in Table 1 and the means are illustrated graphically in Figure 3. A repeated measures analysis of variance (ANOVA) using SAT at Time 1, Time 2, and Time 3 as a three-level repeated measure and tester proximity as a two-level between-subjects independent variable revealed a significant effect for the repeated measure, $F(2, 37) = 6.33$, $p < .05$, but no significant effect for the interaction between the repeated measure and tester proximity, $F(2, 37) = .74$, $p = .48$, or the between-subjects effect for tester proximity, $F(1, 38) = .45$, $p = .51$.

Table 1

Means and Standard Deviations for Subject and Dependent Variables Across Location

	Means and SD	
	Remote	Co-located
SAT – Time 1	45.90	44.80
	(2.65)	(2.69)
SAT – Time 2	44.45	42.45
	(2.82)	(2.57)
SAT – Time 3	42.80	38.75
	(3.08)	(2.17)
CSI - Total	42.40	40.20
	(1.47)	(1.81)

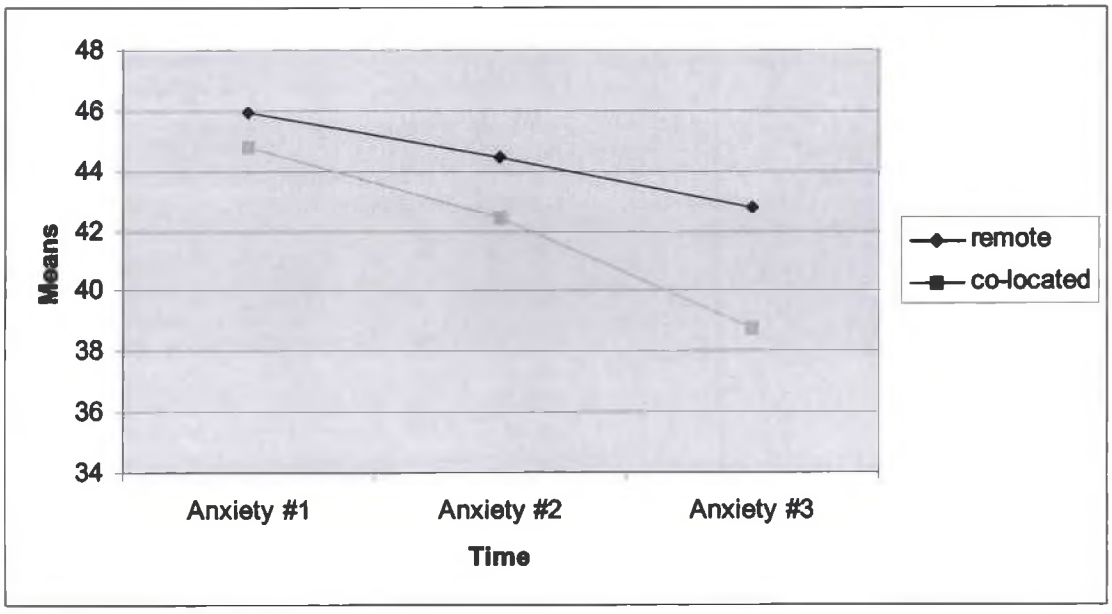

Figure 3. SAT Means Over Time.

The overall means (standard deviations) for SAT at Time 1, Time 2, and Time 3 were 45.35 (11.81), 43.45 (11.95), and 40.78 (11.94), respectively. A test of the linear trend in SAT across time was significant, $F(1, 38) = 12.17$, $p < .05$. The quadratic trend was not significant. Overall, SAT scores tended to decrease linearly over time.

Co-varying out SAT at Time 1, a repeated measures analysis of covariance (ANCOVA) was conducted using SAT at Time 2 and Time 3 as a two-level repeated measure, tester proximity as a two-level between-subjects independent variable, and SAT at Time 1 as a covariate. No significant between-subjects effect for tester proximity was obtained, $F(1, 37) = 1.04$, $p = .31$, and the repeated measure by tester proximity interaction was not significant, $F(1, 37) = 1.27$, $p = .27$. Parallel ANCOVAs using SAT at Time 2 and Time 3 as separate dependent variables and SAT at Time 1 as a covariate revealed no significant effects for tester proximity, $F(2,37) = .259$, $p = .614$ and $F(2,37) = 1.669$, $p = .204$, respectively.

Additional analyses were conducted to further examine any effects of tester proximity on anxiety. Parallel ANCOVAs using SAT at Time 2 and Time 3 as separate dependent variables, tester proximity as a two-level between-subjects independent variable, gender as an independent variable, and SAT at Time 1 as a covariate revealed no significant effects for tester proximity, $F(2,37) = .486$, $p = .490$ and $F(2,37) = .635$, $p = .431$, respectively. Additionally, separate Spearman's rho correlations between SAT at Time 2 and Time 3, and computer usage questionnaire question 2 revealed no significant correlations, $r = -.161$ ($p = .320$) and $r = -.168$ ($p = .301$), respectively.

The Effect of Tester Proximity on Communication Satisfaction

It was hypothesized that remote users would be less satisfied with the interaction during the usability test than the co-located users. Overall, the mean for CSI was 41.30 with a standard deviation of 7.35. Means and standard deviations for CSI overall and each of the seven subscales across tester proximity are listed in Table 1. Eight independent samples t tests were conducted using CSI overall and each of the

seven CSI subscales as dependent measures and tester proximity as a two-level independent variable. The results are listed in Table 2. Marginal tester proximity effects were obtained for both the Depth and Respondent's Engagement CSI subscales. In both cases, participants in the remote condition reported higher levels of satisfaction then participants in the co-located condition.

Table 2
Means and Standard Deviations for Subject and Dependent Variables Across Location and T-test for Equality of Means Comparing Tester Proximity and CSI Overall and Individual CSI Scales

	Remote	Co-located	t	Sig.
Comprehensibility	4.90 (1.33)	4.95 (1.82)	-0.099	.922
Flow	5.45 (1.54)	5.35 (1.38)	0.216	.830
Depth	7.20 (1.64)	6.30 (1.38)	1.877	.068
Efficiency	7.40 (1.64)	7.50 (1.76)	-.19	.853
Respondent	8.50 1.88)	7.25 (2.05)	.673	.051
Partner	3.45 (0.89)	3.45 (1.15)	.000	1.00
Comfort	5.50 (0.89)	5.40 (1.15)	.190	.851
CSI	42.40 (1.47)	40.20 (1.81)	.945	.351

Additional analyses were conducted to further explore the effect of tester proximity on communication satisfaction. Covarying out SAT at Time 1, a one-way analysis of covariance was conducted using overall CSI as the dependent variable, tester proximity as a two-level independent variable, gender as an independent variable, and SAT at Time 1 as a covariate to examine gender and condition interaction effects. No significant effects were obtained, $F(2,37) = .090$, $p = .767$.

Additionally, an ANOVA was conducted to examine any effects of gender using CSI as the dependent variable, tester proximity as a two-level between-subjects independent variable, and gender as an independent variable, revealed no significant effects for tester proximity, $F(2,37) = .186$, $p = .668$.

Multiple ANOVA's were conducted to examine any effects of gender using CSI and all CSI subscales as the dependent variables, tester proximity as a two-level between-subjects independent variable, and gender as an independent variable, revealed no significant effects for tester proximity (see Table 3).

Additionally, as shown in Table 4, separate correlations between the CSI subscales and computer usage questionnaire question 2 revealed no significant correlations.

Table 3

Univariate Analyses of Variance with CSI and subscales as the Dependent Variables, and Tester Proximity and Gender as the Independent Variable

	F	Sig.
Comprehensibility	.012	.912
Flow	.059	.809
Depth	3.325	.077
Efficiency	.000	.994
Respondent	3.247	.080
Partner	.010	.921
Comfort	.052	.821
CSI	.848	.363

Table 4
Spearman's rho correlations of overall CSI and all subscales with Computer Usage Questionnaire question number 2

	r	Sig.
Comprehensibility	-.251	.118
Flow	-.187	.248
Depth	.187	.248
Efficiency	-.286	.073
Respondent	-.004	.981
Partner	-.246	.127
Comfort	-.198	.222
CSI	-.182	.261

CHAPTER IV

DISCUSSION

Overview

The current study focuses on the impact of remote usability testing on user emotions and attitudes. Are remote users as comfortable during the usability test as co-located users? Are they less anxious? Are the tester and participant perceived as more detached and disinterested? Research related to these questions is critical to achieving an understanding of usability testing from a user's perspective and to determining the viability of the remote usability testing technique. A significant portion of the past research on usability testing in general and remote usability testing specifically has focused on behavioral outcomes (e.g., errors made by the user during the test). Considerably less attention has been paid to the processes and channels through which this data is collected, their impact on the user, and, most importantly, the value of the user data obtained. This study has attempted to address this gap in the literature and field.

In the current study, the effect of tester location was examined to determine if remote users were as comfortable and/or less anxious during the usability test as co-located users. It was predicted that users in the remote condition would experience less anxiety than users in the co-located condition. Due to the proximity of the tester, remote users were expected to exhibit lower levels of anxiety related to both evaluation and performance. It was also predicted that remote users would be less satisfied with

the interaction and related communication than the co-located users. Specifically, they would find it more difficult to respond to the tester's requests and questions and they would perceive the experimenter as more detached and less interested in their performance. In the current study, the reactions of the users toward the experimenter were examined to determine if they were perceived as more detached and disinterested.

Hypothesis 1: Anxiety

This hypothesis predicted that users in the remote condition would experience less anxiety than users in the co-located condition. Remote users were supposed to feel less anxiety due to lower evaluation apprehension given that the experimenter wasn't present in the room with them. The current data did not support this hypothesis. Remote users did not exhibit less anxiety over time than did their co-located counterparts. It is possible that this outcome is in relation to Zajonc's (1965) theory of social facilitation effects. It appears that due to the participants' knowledge of the product (or even perceived knowledge), the experimenter's presence and attention (even though in the remote location) facilitated their performance and lessened their anxiety levels. In fact, remote users showed a descriptively higher level of anxiety than co-located users at every measurement point. Barker & Biers (1994) suggest that laboratory environment will mediate the results of a usability test when dealing with participants that exhibit high levels of self consciousness (comparable to anxiety in this study). Perhaps the desired results would have been attained if only highly self conscious (i.e. highly anxious) participants were utilized. Although the difference was not significant, this trend should be explored further and may lead to a revision of the expectations related to remote usability testing. If it turns out that remote participants are more anxious than

co-located participants, this may be due to an increased sense of isolation and related discomfort in the remote condition. During the co-located condition, the experimenter was in the room with the participant guiding them through the foreign process of usability testing, and even more globally, their participation in a psychological experiment. Although there is no significant indication of this, it is possible that the immediate presence of the experimenter next to the participant (in the co-located condition) may have helped to reduce anxiety and evaluation apprehension. This presence may have afforded more non-verbal or testing cues that it was in fact the software being tested, and not the participant. Although it is not possible to determine the accuracy of this postulation here, further research should be conducted to examine if these aspects did in fact play a role. Conversely, remote participants were instructed to remain in the testing room with a headset on until the experimenter attended to them. Conceivably, this could have led to increased anxiety as the participants may have felt more scrutinized and 'under the microscope.' This initial burst of heightened anxiety may not have diminished quickly enough to normal levels before the experimental session had ended. Again, additional studies would need to be conducted to explore this interpretation.

A second potential explanation for the current results may be that the tasks presented in this study were not difficult enough to generate a high level of anxiety in the co-located condition. Although the tasks to be performed were deemed to be difficult enough to elicit emotions (such as apprehension and anxiety) during the course of the study, the level actually generated was below the midpoint of the anxiety scale on average. The scores on the SAT scale were at the lower end of the anxiety spectrum.

Actual scores ranged from 40-45, with potential scale limits of 21-105. These results may indicate a floor effect, with little room at the bottom end of the scale for scores to drop any further. A more difficult set of tasks may have raised these scores to a much higher level (i.e. 80 and above) which may have allowed anxiety to drop more precipitously in the remote condition while remaining high in the co-located condition as expected. This also could be attributed to social facilitation effects as described earlier by Zajonc (1965). Since the tasks may not have been difficult enough, the audience provided by the testing may have led to the participants' performance in the co-located group to be facilitated by the experimenter, thereby resulting in lower anxiety.

The downward trend in anxiety over time is not surprising. The most plausible explanation for this trend is that users became more accustomed to the testing procedure over time. Current data suggest that users in both conditions were slightly more nervous and apprehensive initially, perhaps because they felt as if they were being evaluated and they were not familiar with the surroundings. This would be comparable to any person arriving at a testing facility without sufficient explanation as to the extent and purpose of the test. Over time however, the user may realize that this test is not intrusive and may become more familiar and comfortable with the environment and procedures. Additionally, it is quite likely in this study that the participants were not ego involved, and were considerably unmotivated. Since this experiment was utilizing college students who were not dependent on the tool for their job or for life/career success, it is quite likely that they were unmotivated to perform well, and had very little apprehension (anxiety) in participating in the experiment.

Hypothesis 2: Communication Satisfaction

This hypothesis predicted that remote users would be less satisfied with their communication with the tester than the co-located users. Remote users were expected to find it more difficult to respond to the experimenter's questions and were expected to perceive the experimenter as more detached and less interested. The current study did not support this hypothesis. Remote users did not rate the experimenter as more detached and did not have more difficulty responding to questions posed. In fact, remote users showed a higher level of satisfaction than the co-located users at every measurement point. Although not significant, this finding could be tied to the earlier explanation. Because participants were isolated in the testing room, they may have formed a stronger relationship with the experimenter using the leaner communication channel in hopes of gaining feedback. As this was the only form of communication with the experimenter (i.e. lack of non-verbal cues), the participant may have attached greater importance to it. This reliance on the communication with the separated experimenter may result in higher levels of satisfaction.

Further Discussion

Although neither hypothesis was supported in this study, several conclusions can be drawn. There appeared to be no difference between the remote and co-located conditions when examining performance on the actual usability test. The two groups were very comparable on the amount of time to complete each task, number of errors made, number of clicks made, and quality and quantity of open ended responses. Although null results should be interpreted carefully, this finding may reinforce

previous conclusions that remote testing can be thought of as a viable alternative to the more traditional type of usability testing (i.e. face-to-face or co-located usability testing; Hartson et al., 1996; Macko, 1998).

Second, the appearance of reduced anxiety over time (albeit not based on condition) is an indication that initial apprehension can be alleviated over time, even in a remote usability test. Over time, users from both conditions exhibited a tendency to become more comfortable (lower anxiety) as the testing session progressed. This indicates that the longer a user participates in a session, the more comfortable they become, and conceivably the more genuine and non-biased their responses become. This finding can be attributed in a couple of different ways. First, since the user had been practicing using the product throughout the entire test, one could make a case that the practice had reduced their sense of evaluation apprehension (Sanna & Shotland, 1990). Since the participants evaluation apprehension had been reducing, their anxiety levels may have as well. Second, as the test progressed, the participant may have found the tasks to become easier, thus leading to a social facilitation effect (Zajonc, 1965). Simply by getting more comfortable with the tasks and having an audience present (either remote or co-located) may have led to better performance, and therefore lower anxiety. Although these findings can likely be attributed to most experimental situations, it provides an excellent example. Since most participants' anxiety levels will decrease with more experimental exposure, future research should keep this in mind when determining their methodology.

Since this study did not support the hypotheses, further research is clearly necessary to draw any valid conclusions regarding anxiety and communication

satisfaction during remote usability testing. Potential directions for future research are numerous. First, a more detailed and difficult usability testing procedure might cause more divided responses between groups of participants. The rigidity and depth of the actual usability test could be manipulated to ascertain potential effects. Additionally, the amount and value of the interaction with the participants could be examined. The script used for interactions with the participants was identical for each condition. Although these should not be different, the amount of interaction could be increased (or decreased) to elicit potential behaviors. Coupled with a more stringent testing protocol, participant/experimenter interaction could be strengthened to allow for enhanced communication. By operationalizing the experimenter's communication techniques, this methodological adjustment may uncover differences that were not detected.

Secondly, the inclusion of video may in fact yield important insights, as both conditions would then be closer to actually mimicking each other. Although video was not used in the current study, its use may be necessary for a more complete picture of the differences between conditions. With the inclusion of a video feed from participant to experimenter and vice versa, the remaining differences between conditions could theoretically be reduced to only that of physical proximity. Since both groups of users would have visual contact with the experimenter, it would reduce differences with respect to social effects and environmental pressures, namely solitude and isolation. This would then allow for a much stronger direct comparison between groups, with a greater focus on communication satisfaction and acceptance, as well as anxiety responses.

Third, the analysis of certain variables not studied in detail here could have conceivably altered the findings. Although the participants were certainly not QuickPlace experts, this user group as a whole probably could not be considered novice users. These participants rely heavily on technology not only for academic tasks, but also for personal experiences. This increased familiarity may have led to lessened levels of anxiety, less difficulty on task performance, etc. Additionally, this user group was relatively homogenous with respect to age, income, and education. This particular sample had unique characteristics that may distinguish them from a group of employees who could potentially be using QuickPlace for the first time. User groups contained within a corporate environment, for example, would likely have a higher variance in computer knowledge and efficiency, and thus may exhibit higher levels of anxiety and display decreased performance measures.

Although the current findings did not support the predicted hypotheses, additional studies in this area should be conducted. It is still not known to what extent anxiety and communication satisfaction play a role when conducting a remote usability evaluation. Knowing the role that these emotions may play in a successful usability test will allow us to determine the efficacy of remote usability testing in comparison to the traditional co-located method. It is possible that there is little difference between the methods, but it is imperative that all avenues be explored to determine if this is in fact the case. If there were to be some unknown effect that alters the results obtained from a remote usability test, would its use still be supported within the usability community? Subtle, indirect effects related to communication satisfaction and anxiety could play a detrimental role in remote testing threatening the validity and reliability of remote

usability test results. Although it has been demonstrated that remote testing yields similar results, testers must be aware of any possible challenges associated with this technique. Any suggestions of inefficiency should be studied and corrected if possible in order to achieve the utmost effectiveness of usability testing, both co-located and remote. In addition, future research should be fully aware of potential practice effects, personality differences, and social facilitation/interference aspects of any future research endeavor (Sanna & Shotland, 1990, Rosenberg, 1969, and Zajonc, 1965).

Although usability testing may be thought of as just one aspect of a product development process, it is an important one. Conducting usability tests on products to be used by consumers is essential for the quality and ease of use of the product. Without adequate usability testing, many products have suffered and many companies have lost significant revenues. By attempting to research as exhaustively as possible all aspects associated with usability testing, not only is the usability community as a whole contributing to our knowledge base, but they are also providing a great service to corporations and research facilities around the world. By examining usability in its most detailed form, we are improving the process, the findings, and the end results that impact the development process. By doing this, we are creating the best products possible that can be used with ease by consumers worldwide. As humans continue to rely more and more heavily on technology, it is essential that we continue to refine this field and these testing procedures, to ensure user experiences of the highest of quality for years to come.

APPENDIX A

INFORMED CONSENT FORM

Study Overview

Welcome to the study **Product Evaluation**. The following is a general description of the study and a reminder of your rights as a potential participant. As in any study, your participation is completely voluntary. If now, or at any point during the study, you decide that you do not want to continue participating, please let the experimenter know and you will be dismissed without penalty. Also, please remember that your name will not be associated with any of the information that you provide during the study. All of the information you provide is absolutely anonymous and confidential.

In this study you will be completing several questionnaires and performing two separate tasks on a computer. Although the experiment does not require that you have computer experience, we would like to identify those individuals who are either extremely uncomfortable or unfamiliar with them. If you fall into either of these categories, please let an experimenter know at this time.

In this study, you will be asked to use a particular piece of software. You will be taught the basic elements of operating the software you will use during the study. As you use the software, you will be observed by an experimenter and your voice and screen movements will be recorded. Your name will not be associated with these recordings and they will be used for data analysis and presentation purposes only. Your signature on this form indicates your consent to be recorded. If you have any questions or concerns about this, please direct them to the experimenter at this time.

For Further Information

The faculty member responsible for conducting this research is Dr. Ken Graetz. Dr. Graetz would be happy to address any of your questions or concerns regarding this study and he can be reached at 229-2168 or in his office at SJ 317. If you feel there is an ethical problem with this study or in any study that you have participated in, please contact:

Dr. Charles Kimble, Chair
Research Review and Ethics Committee
SJ 319
229-2167

If you would like to participate in this study, please sign in the space provided. Your signature indicates that you are aware of each of the following: 1) the general procedure to be used in this study, 2) your right to discontinue participation at any time,

3) the steps taken to insure confidentiality of the data you will provide during the study, and 4) your consent to be recorded.

Signature:_____ Date:_____

APPENDIX B

COMPUTER USAGE QUESTIONNAIRE

1. Which of the following best describes your computer knowledge?
   ___ None (Never use a personal computer, have no real knowledge)
   ___ Low (Little or no knowledge of jargon, can perform activities like basic data entry or word processing often need help)
   ___ Intermediate (Knowledge of jargon, can use some programs e.g., slide presentation, spreadsheets, or financial software, occasionally need help)
   ___ Advanced (Proficient with jargon, adept at learning new software, serve as a resource for other)
   ___ Expert (Knowledge of most computer platforms, can build computers and code them, have extensive knowledge of computer programming, including several languages e.g., C++, Perl, Java, etc.)

2. About how many hours per week do you use a computer?
   ___ 0-1 hour
   ___ 2-6 hours
   ___ 7-10 hours
   ___ More than 10 hours

3. Which of the following best describes your Internet/Web experience?
   ___ None (Never use the Internet, have no experience)
   ___ Low (Little or no knowledge of jargon, can perform activities like simple browsing, often need help)
   ___ Intermediate (Knowledge of jargon, can perform more complex activities like downloading files or online shopping, occasionally need help)
   ___ Advanced (Proficient with jargon, adept at using new applications, can build web pages, serve as a resource for others)

4. How many hours a week do you use the Internet?
   ___ 0-1 hour
   ___ 2-6 hours
   ___ 7-10 hours
   ___ More than 10 hours

On the back of this form, please provide any information that you think might affect your performance on a computer driven task.

1. How many e-mail addresses/accounts do you have?

2. When you have trouble or you have a question with your computer, where do you turn (be specific)?

3. How many windows or applications do you typically use at one time on your computer?

4. How do you switch between windows on your desktop?

5. What is TCP/IP?

6. What is HTML?

7. How many computers do you have (rent or own)?

8. Have you ever seen the inside of your computer (if yes, why)?

9. Do you tend to set preferences and customize your applications (if yes, please explain)?

10. What comes up when you start your browser?
    ___ Default page
    ___ A page I specified
    ___ A page someone else specified for me
    ___ Don't know

APPENDIX C

SOCIAL ANXIETY THOUGHTS QUESTIONNAIRE (SAT)

We are interested in the thoughts that people have in social situations. Listed below are a variety of thoughts that pop into peoples' heads in situations that involve being with other people or talking to them. Please read through each thought and indicate how frequently, if at all, the thought occurred to you over the last week. Please read each item carefully and, following the scale, indicate to the left of the question the number that best applies to you. Please answer each question carefully. In social or interpersonal situations during the past week, how often did you have the following thoughts?

1 = Never
2 = Rarely
3 = Sometimes
4 = Often
5 = Always

____ 1. I feel tense and uncertain.
____ 2. I don't know what to say.
____ 3. Maybe I sound stupid.
____ 4. I am perspiring.
____ 5. What will I say first?
____ 6. Can they tell I am nervous?
____ 7. I feel afraid.
____ 8. I wish I could just be myself.
____ 9. What are they thinking of me.
____ 10. I feel shaky.
____ 11. I'm not pronouncing well.
____ 12. Will others notice my anxiety?
____ 13. I feel defenseless.
____ 14. I will freeze up.
____ 15. Now they know I am nervous.
____ 16. I don't like being in this situation.
____ 17. I am inadequate.
____ 18. Does my anxiety show?
____ 19. I feel tense in my stomach.
____ 20. Others will not understand me.
____ 21. What do they think of me

APPENDIX D

COMMUNICATION SATISFACTION INVENTORY

Subscale Description

Comprehensibility (easy versus difficult to process). This refers to the self-perceived comprehension of the conversation. Was the respondent able to understand the conversational partner?

I found it easy to understand what the tester said. (+)
The tester spoke clearly. (+)
I understood everything that the tester said. (+)
At times, it was difficult to understand what the tester was trying to say. (-)

Flow (awkward versus flowing). This refers to the respondent's perception of the sequential flow of the conversation. Was it easy to transition from one speaker to another? Were there awkward pauses?

This was a smooth conversation. (+)
I always knew when the tester was done speaking. (+)
We tended to interrupt each other during the conversation. (-)
It was always clear to me when I was supposed to speak. (+)
There were some awkward pauses in this conversation. (-)
At times, it was hard to know when to speak. (-)

Depth (deep versus shallow). This refers to the respondent's perception of the depth of the conversation. Did the conversation drill down into various topics or did it focus on surface topics, moving from one to another.

Our conversation focused in depth on a few specific topics. (+)
Our conversation jumped from topic to topic. (-)
We didn't spend much time on any one topic. (-)
This was a very focused conversation. (+)

Efficiency (efficient versus inefficient). This refers to the respondent's perception of that the conversation accomplished its goal. Did it serve its purpose? Was the conversation directed or was there a lot of off-topic conversation?

The conversation stayed on track. (+)
This conversation served its purpose. (+)
There was a lot of extraneous (i.e., off topic) discussion. (-)
This conversation could have been more efficient. (-)

Respondent's Engagement (interesting versus boring). This refers to the respondent's perception of the conversation as interesting and engaging. Did the conversation hold your attention or was it boring?

The conversation was interesting. (+)
I really liked this conversation (+)
I paid attention to what the tester said. (+)
I was bored by this conversation. (-)
I really did not want to talk to the tester (-)
My attention drifted during this conversation. (-)

Partner's Engagement (attentive versus disinterested). This refers to the respondent's perception of his/her conversational partner as attentive or disinterested. Was the partner actively listening?

The tester paid attention to what I said. (+)
The tester seemed to really like this conversation (+)
The tester seemed interested in what I said. (+)
The tester did not seem to really want to talk to me (-)
The tester seemed bored by this conversation (-)
The tester's attention seemed to drift during this conversation (-)

Comfort Level (comfortable versus uncomfortable). This refers to the respondent's perception of his/her level of comfort during the conversation. Did the respondent feel at ease and comfortable during the conversation?

I felt comfortable during my conversation with the tester (+)
I really felt at ease talking with the tester (+)
At times, I felt uncomfortable talking with the tester (-)
This conversation was more uncomfortable than usual. (-)

Actual Scale

Instructions

Please answer the following questions by entering the number in the blank that corresponds to your feeling about the statement.

Strongly Agree (1) Agree (2) Neutral (3) Disagree (4) Strongly Disagree (5)

___1. Our conversation jumped from topic to topic.

___2. I really liked this conversation.

___3. There were some awkward pauses in this conversation.

___4. The tester seemed bored by this conversation.

___5. I really felt at ease talking with the tester.

___6. At times, it was hard to know when to speak.

___7. I paid attention to what the tester said.

___8. This conversation was more uncomfortable than usual.

___9. I felt comfortable during my conversation with the tester.

___10. I found it easy to understand what the tester said.

___11. Our conversation focused in depth on a few specific topics.

___12. This was a very focused conversation.

___13. This was a smooth conversation.

___14. At times, I felt uncomfortable talking with the tester.

___15. My attention drifted during this conversation.

___16. I really did not want to talk to the tester.

___17. The tester seemed interested in what I said.

___18. This conversation could have been more efficient.

___19. At times, it was difficult to understand what the tester was trying to say.

___20. The tester did not seem to really want to talk to me.

___21. We tended to interrupt each other during the conversation.

___22. The tester spoke clearly.

___23. It was always clear to me when I was supposed to speak.

___24. There was a lot of extraneous (i.e., off topic) discussion.

___25. I understood everything that the tester said.

___26. The tester paid attention to what I said.

___27. We didn't spend much time on any one topic.

___28. I always knew when the tester was done speaking.

___29. The conversation stayed on track.

___30. The conversation was interesting.

___31. I was bored by this conversation.

___32. The tester's attention seemed to drift during this conversation.

___33. This conversation served its purpose.

___34. The tester seemed to really like this conversation.

APPENDIX E

FACE TO FACE PROTOCOL

Meeting Participants

[Participant enters room and sits down.]
Interviewer comes in and greets subject.
"Hello, are you here for the PSY 101 study?" "Great", "let's get started."
"My name is Mike and I will be working with you today. I am going to have you do a couple of tasks on a computer, to see how well you do with a newly developed product. This is called a usability test. Don't think of it as a real test though, it's simply you interacting with this software to see how well it works. It is the product that is actually getting evaluated, not you."

Informed Consent
So, the first thing we need to do is have you fill out an informed consent form. This basically describes your rights as a participant, that I will not harm you in any way, and that you can stop at any time. If you have any questions, please feel free to ask, otherwise, please sign and date the bottom."

Introduction

"OK, let me tell you what you're going to be doing. First I'm going to have you fill out a computer experience questionnaire. After that, we're going to go over a little bit about the product you'll be using (it's called QuickPlace), and give you a little training (it won't take more than 5 minutes so don't worry). Then I'm going to have you complete a couple of tasks within QuickPlace. Lastly, you'll need to fill out another questionnaire, and then you can go. This should all take about 30 minutes. Do you have any questions before we begin? No? OK, let's get to it."

[Hand participant copy of Comp Exp Questionnaire.]

Comp Exp Quest
"Please take a minute or two to fill out the computer experience questionnaire. If you have any questions, please let me know. Thanks."
[Tester examines questionnaire to determine level of computer knowledge]

Move into testing room
OK, let's move in here and we can begin the study. [Participant sits at workstation; Tester sits next to and behind them].

Training
Go ahead and open up the window (at the bottom of your screen). This is a QuickPlace. Let me take control of your screen for a minute and show you around. You can think of a QuickPlace just like a web site. This screen you are looking at is the home page, the links on the side here are your navigation links, and this area here is the content area. That about the gist of what QuickPlace is. There is also an area over here [move mouse] called the tools panel that allows you to chat with other members of the QP. One big difference you may notice is this link over here called <members>. QuickPlaces are designed for groups of people to upload documents, have threaded discussions, etc, but you can't do any of that until you become a member. So if you go into <members>, you'll see that you can add/remove members. You can also customize what the QuickPlace looks like by clicking on the <customize> link here on the sidebar. One last thing you should be aware of is that you can add new pages to this QuickPlace, by clicking on the <new…> button up here at any time. Clicking this button gives you several options; you can create a regular new QuickPlace page, a new Microsoft page, tasks page, calendar page, etc. All of these options will provide you with step-by-step instructions. Do you have any questions before we begin?" [If Yes; answer them, If No; continue on.]

Task 1
"The first task I'm going to have you do is to create a new page in QuickPlace. [User clicks on <new…>] "Please give the page a title, it can be anything you would like. [User types in a title] Now please write two sentences about yourself in the Contents area. These two sentences can be anything you want" [User types two sentences] "Now I would like you to add a link to this page. There is an icon for adding a link there, and I would like you to add the link http://yahoo.com" [User adds the link] "Now you are ready to publish this page. Please click on the <Publish as…> button at the bottom of the page. Here I would like you to select Notify (via email) AND Add Editors. Also, I would like you to put this page into the "Put Page Here!" folder. Once you have completed these three steps, please click next. [User completes and clicks next and is brought to the Notify page] Please send this email to Michael Goliber. [The user must click on the "To:" link in order to bring up a page with all of the members of QuickPlace. After checking of the correct corresponding name, the participant will click the "Next" button that will bring them back to the "Notify" page. Clicking "Next" will then take the user to the "Add Authors" page] Now I would like you to add Michael Goliber as an author." [User clicks on the checkbox for Michael Goliber]. "Congratulations, you are finished with the first task. How did you feel about the task? Was it easy, hard? Could any aspect of this task have been made simpler for you? What types of things would you change about the functionality of QuickPlace?" The tester will reply to each answer with "good", "great", "OK", or some other affirmative response. After responding to these questions, the user is now done with Task 1.

Anxiety Questionnaire
"Good job. Now I'm going to have you take another questionnaire. It's labeled questionnaire B next to you. Please fill it out to the best of your ability. If you have any questions, please let me know".
[User will take the questionnaire at this point]

Task 2
"Thanks. Now we're ready for the second task. Please look at your monitor. You will now be adding a new Task page to the QuickPlace. Please go to the Tasks area and then select new task page. [User clicks on "Tasks" on the sidebar, then on "New Task Page" in the upper right] Please give this page a title, and again, it can be anything you want. [User gives page a title] Good. Please make the start date April 1st, and make the end date April 15th. Assign the task to Michael Goliber, and make sure that both the author and assignee can edit this task. [User completes the steps] Finally, type in 2 sentences of text in the description area that would describe a task of your choosing. These sentences can describe any task you want, but it needs to be 2 sentences long. [User completes description] Now you can simply click on publish (no need for publish as this time). [User clicks on publish] Very good job. You are now done with this task. "How did you feel about the task? Was it easy, hard? Could any aspect of this task have been made simpler for you? What types of things would you change about the functionality of QuickPlace?" The tester will reply to each answer with "good", "great", "OK", or some other affirmative response. After responding to these questions, the user is now done with Task 2.

Anxiety Questionnaire and CSI
"Good job. Now I'm going to have you take a couple more questionnaires. They are labeled questionnaires C and D next to you. Please fill them out to the best of your ability. If you have any questions, please let me know".
[User will take the questionnaires at this point]

Debriefing
"OK, you're completely done now. I wanted to let you know why I was conducting this study. The way you acted as a participant today was only one group of my study. The other group was when this study is conducted at a distance. Basically, the other group sat in here by themselves, and I interacted with them using headsets. What I am trying to examine here, is if there is a difference in user attitude and behavior between the two methods. Research suggests that in the distance group, participants typically give less detailed answers. I want to know why. That is why I had you fill out the last questionnaire, to determine how you actually felt about being a participant, and I can then compare that to your level of experience, your performance on the tasks, and other variables. Does this all make sense? OK, here is your credit slip, and if you have any other questions, my name is on the informed consent sheet you have, so please feel free to contact me. Thanks for your time."

APPENDIX F

REMOTE PROTOCOL

Meeting Participants

[Participant enters room and sits down.]
Interviewer comes in and greets subject.
"Hello, are you here for the PSY 101 study?" "Great", "let's get started."
"My name is _____ and I am going to get you started today. In a minute, I'm going to turn you over to Mike, who will be running through your session today. He's going to have you do a couple of tasks on a computer, to see how well you do with a newly developed product. This is called a usability test. Don't think of it as a real test though, it's simply you interacting with this software to see how well it works. It is the product that is actually getting evaluated, not you. Mike's going to be working with you remotely, which means that he won't be in the room with you, but you'll be able to talk to him, and he'll be able to see what you're doing."

Informed Consent
"So, the first thing we need to do is have you fill out an informed consent form. This basically describes your rights as a participant, that I will not harm you in any way, and that you can stop at any time. If you have any questions, please feel free to ask, otherwise, please sign and date the bottom."

Introduction

"OK, let me tell you what you're going to be doing. First I'm going to have you fill out a computer experience questionnaire. After that, we're going to go over a little bit about the product you'll be using (it's called QuickPlace), and give you a little training (it won't take more than 5 minutes so don't worry). Then I'm going to have you complete a couple of tasks within QuickPlace. Lastly, you'll need to fill out another questionnaire, and then you can go. This should all take about 30 minutes. Do you have any questions before we begin? No? OK, let's get to it."

[Hand participant copy of Comp Exp Questionnaire.]

Comp Exp Quest
"Please take a minute or two to fill out the computer experience questionnaire. If you have any questions, please let me know. Thanks."
[Tester examines questionnaire to determine level of computer knowledge]

Move into testing room
"OK, let's move in here and we can begin the study. [Participant sits at workstation; Tester walks participant in] I'm going to hand you off to Mike now. Please put on your headset so you can talk to each other" [Tester waits until participant and Mike are speaking, then leaves]

Training
"Hi (fill in name here), I'm Mike and I am going to be working with you today. Let's get started shall we? Go ahead and open up the window (at the bottom of your screen). This is a QuickPlace. Let me take control of your screen for a minute and show you around. You can think of a QuickPlace just like a web site. This screen you are looking at is the home page, the links on the side here are your navigation links, and this area here is the content area. That about the gist of what QuickPlace is. There is also an area over here [move mouse] called the tools panel that allows you to chat with other members of the QP. One big difference you may notice is this link over here called <members>. QuickPlaces are designed for groups of people to upload documents, have threaded discussions, etc, but you can't do any of that until you become a member. So if you go into <members>, you'll see that you can add/remove members. You can also customize what the QuickPlace looks like by clicking on the <customize> link here on the sidebar. One last thing you should be aware of is that you can add new pages to this QuickPlace, by clicking on the <new…> button up here at any time. Clicking this button gives you several options; you can create a regular new QuickPlace page, a new Microsoft page, tasks page, calendar page, etc. All of these options will provide you with step-by-step instructions. Do you have any questions before we begin?" [If Yes; answer them, If No; continue on.]

Task 1
"The first task I'm going to have you do is to create a new page in QuickPlace. Go ahead and click on the "new" link in the QuickPlace and select new page [User clicks on <new…>, selects new page] "Please give the page a title, it can be anything you would like. [User types in a title] Now please write two sentences about yourself in the Contents area. These two sentences can be anything you want" [User types two sentences] "Now I would like you to add a link to this page. There is an icon for adding a link there, and I would like you to add the link http://yahoo.com" [User adds the link] "Now you are ready to publish this page. Please click on the <Publish as…> button at the bottom of the page. Here I would like you to select Notify (via email) AND Add Editors. Also, I would like you to put this page into the "Put Page Here!" folder. Once you have completed these three steps, please click next. [User completes and clicks next and is brought to the Notify page] Please send this email to Michael Goliber. [The user must click on the "To:" link in order to bring up a page with all of the members of QuickPlace. After checking of the correct corresponding name, the participant will click the "Next" button that will bring them back to the "Notify" page. Clicking "Next" will then take the user to the "Add Authors" page] Now I would like you to add Michael Goliber as an author." [User clicks on the checkbox for Michael Goliber]. "Congratulations, you are finished with the first task. How did you feel about the task? Was it easy, hard? Could any aspect of this task have been made simpler for you?

What types of things would you change about the functionality of QuickPlace?" The tester will reply to each answer with "good", "great", "OK", or some other affirmative response. After responding to these questions, the user is now done with Task 1.

Anxiety Questionnaire
"Good job. Now I'm going to have you take another questionnaire. It's labeled questionnaire B next to you. Please fill it out to the best of your ability. If you have any questions, please let me know".
[User will take the questionnaire at this point]

Task 2
"Thanks. Now we're ready for the second task. Please look at your monitor. You will now be adding a new Task page to the QuickPlace. Please go to the Tasks area and then select new task page. [User clicks on "Tasks" on the sidebar, then on "New Task Page" in the upper right] Please give this page a title, and again, it can be anything you want. [User gives page a title] Good. Please make the start date April 1st, and make the end date April 15th. Assign the task to Michael Goliber, and make sure that both the author and assignee can edit this task. [User completes the steps] Finally, type in 2 sentences of text in the description area that would describe a task of your choosing. These sentences can describe any task you want, but it needs to be 2 sentences long. [User completes description] Now you can simply click on publish (no need for publish as this time). [User clicks on publish] Very good job. You are now done with this task. "How did you feel about the task? Was it easy, hard? Could any aspect of this task have been made simpler for you? What types of things would you change about the functionality of QuickPlace?" The tester will reply to each answer with "good", "great", "OK", or some other affirmative response. After responding to these questions, the user is now done with Task 2.

Anxiety Questionnaire and CSI
"Good job. Now I'm going to have you take a couple more questionnaires. They are labeled questionnaires C and D next to you. Please fill them out to the best of your ability. If you have any questions, please let me know".
[User will take the questionnaires at this point]

Debriefing
"OK, you're completely done now. I wanted to let you know why I was conducting this study. The way you acted as a participant today was only one group of my study. The other group was when this study is conducted at a distance. Basically, the other group sat in here by themselves, and I interacted with them using headsets. What I am trying to examine here, is if there is a difference in user attitude and behavior between the two methods. Research suggests that in the distance group, participants typically give less detailed answers. I want to know why. That is why I had you fill out the last questionnaire, to determine how you actually felt about being a participant, and I can then compare that to your level of experience, your performance on the tasks, and other variables. Does this all make sense? OK, here is your credit slip, and if you have any

other questions, my name is on the informed consent sheet you have, so please feel free to contact me. Thanks for your time."

APPENDIX G

DEBRIEFING FORM

Thank you for participating in the study. This is the debriefing section of the study. You are now finished with the experiment, so let me tell you a little bit about why this study was being run. The purpose for this study was to determine if users (you) had varying levels of anxiety and comfort of communication with the experimenter. What you didn't know was that there was a different condition in this study. You were either in the face-to-face condition (the experimenter sat right next to you), or you were in the remote condition (the experimenter was in a completely different room and spoke with you via headsets). The purpose for having these two conditions was to see if there was any difference between the two. We hypothesized that in the remote condition, users' anxiety would be lower due to a lack of perceived evaluation apprehension. In other words, users wouldn't feel as pressured by having someone sit and watch over their shoulder. We also hypothesized that in the remote condition, users comfort with communication would be lower, due to the conversation using headsets and not having the experimenter in the room with them.
That is the crux of the experiment. If you have any questions, please feel free to ask me right now, or in the future (Mike Goliber, golibemj@notes.udayton.edu). Also, if you feel that you have been mistreated in any way, please feel free to contact either the faculty member in charge of this experiment (Ken Graetz, ken.graetz@notes.udayton.edu) or the Chair of the Ethics Committee (Charles Kimble, charles.kimble@notes.udayton.edu).

Thank you for your participation in this study, and have a great day.

Mike Goliber

REFERENCES

Barker, R. T., & Biers: D. W. (1994). Software usability testing: Do user self-consciousness and the laboratory environment make any difference? *Proceedings of the Human Factors Society, 38th Annual Meeting* (pp. 1131-1134). Santa Monica, CA: Human Factors and Ergonomics Society.

Castillo, J. C., Hartson, H. R., & Hix. D. (1997a). Remote Usability Evaluation at a Glance. *Technical report from Virginia Tech: TR-97-12.* Retrieved May 23, 2001, from http://miso.cs.vt.edu/~usab/remote/docs/TR_remote_evaluation.pdf

Conklin, P. (1991). Bringing usability effectively into product development. In M. Rudisell, C. Lewis, P. B. Polson, and T. D. McKay (Eds.), *Human-computer interface design: Success stories, emerging methods, and real-world context* (pp. 367-374). San Francisco, CA: Morgan Kaufmann.

Cottrell, N. B., Wack, D.L., Sekerak, G. J., & Rittle, R. M. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology, 9,* 245-250.

Dumas J. S., & Redish, J. C. (1993). *A practical guide to usability testing.* Norwood, NJ: Ablex Publishing Corporation.

Ericsson, K. A. & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: The MIT Press.

Galitz, W. O. (1997). *The essential guide to user interface design.* New York: John Wiley & Sons, Inc.

Hackman, G. S., & Biers, D. W. (1992). Team usability testing: Are two heads better than one? *Proceedings of the Human Factors Society, 36th Annual Meeting* (pp. 1205-1209). Santa Monica, CA: Human Factors and Ergonomics Society.

Hammontree, M., Weiler, P., & Nayak, N. (1994). Remote usability testing. *Interactions. 1(3),* 21-25. Retrieved May, 21, 2001, from the Digital Library, http://www.acm.org.

Hartson, H. R., Castillo, J. C., Kelso, J., Kamler, J., & Neale, W. C. (1996). Remote evaluation: The network as an extension of the usability laboratory. *Proceedings of CHI'96 Human Factors in Computing Systems,* 228-235.

Hix, D. & Hartson, H. R. (1993). *Developing user interfaces: Ensuring usability through product & process.* New York: Wiley.

IBM (2001). *Cost justifying ease of use.* Retrieved October 3, 2001, from http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/23

Karat, J. (1997). User-centered software evaluation methodologies. In M. G. Helander, T. K. Landauer, & P. V. Prabhu (Eds.), *Handbook of human-computer interaction.* (pp. 689-704). Amsterdam; New York: Elsevier.

Lesaigle, E. M., & Biers, D. W. (2000). Effect of type of information on real-time usability evaluation: Implications for remote usability testing. *Proceedings of the IEA 2000/HFES 2000 Congress*, 6, 585-588.

Macko, S. J. (1998). *Remote evaluation: A comparison of attended and unattended usability testing via the internet.* Unpublished master's thesis, Carleton University, Hamilton, Ontario, Canada.

Mauro, C. L. (1994). Cost justifying usability in a contractor company. In R. G. Bias & D. J. Mayhew (Eds.), *Cost-justifying usability* (pp. 123-142). New York: Academic Press.

Meister, D. & O'Brien, T. G. (1996). The history of human factors testing and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 3-11), Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nielsen, J. (1993). *Usability Engineering.* Boston: Academic Press.

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. Rosnow (Eds.), *Artifact in behavioral research.* New York: Academic Press.

Rubin, J. (1994). *Handbook of usability testing: How to plan, design, and conduct effective tests.* New York: Wiley.

Sanna, L. J., & Shotland, R. L. (1990). Valence of anticipated evaluation and social facilitation. *Journal of Experimental Social Psychology, 26,* 82-92.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). *An introduction to human factors engineering.* New York: Addison-Wesley Educational Publishers.

Wixon, D. & Jones, S. (1996). Usability for fun and profit: A case study of the design of DEC Rally Version 2. In M. Rudisell, C. Lewis, P. B. Polson, & T. D. McKay (Eds.), Human-computer interface design: Success stories, emerging methods, and real-world context (pp. 3-35). San Francisco, CA: Morgan Kaufmann.

Zajonc, R. B. (1965). Social Facilitation. *Science, 149,* 269-274.