Running head: COLLECTING SUBJECTIVE DATA IN REMOTE USABILITY

Collecting Subjective Data in Remote Usability Testing

Cassandra L. Holmes, B.A., Hons.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of
Master of Arts

Department of Psychology

Carleton University

Ottawa, Ontario

May, 2003

© 2003, Cassandra L. Holmes

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitons et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-89048-1

Our file Notre référence

ISBN: 0-612-89048-1

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this dissertation.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de ce manuscrit.

While these forms may be included in the document page count, their removal does not represent any loss of content from the dissertation.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The effects of the presence or absence of an experimenter in remote usability testing, as well as the effects of participants' input method (voice or typing) on the quality of data collected, was examined. Emphasis was placed on information that is normally spoken, such as thoughts, comments, and answers to questions, but consideration was also given to differences in the evaluation of the web site usability as a function of the remote testing conditions. Fifty participants were randomly assigned to one of five conditions: four experimental conditions, all of which were remote, or the local control condition. Differences among groups on test time, ease of use ratings, experimenter communications, and various aspects of the comments were examined. Based on the results of this study, if remote usability testing is to be used, the recommended method is attended/voice as opposed to attended/typing, unattended/voice, and unattended/typing.

Acknowledgments

There are several people I would like to thank for their involvement and assistance in the completion of my thesis. I would like to thank my thesis advisor, Dr. Richard F. Dillon, for his guidance, insight, and time spent providing comments on the copious drafts. I would also like to thank Dr. Andrew Patrick of the National Research Council for his support, contribution of the web site, and the many hours of work assisting me with this study. I would like to thank my committee members, Dr. Jo Wood and Dr. Gitte Lindgaard, for their advice and suggestions along the way, and Dr. Brian Greenspan for agreeing to be my external committee member on short notice. For his time, technical support, and wise words, I would like to thank Jing Liu. Finally, I would like to thank my friends for their patience and proofreading, and my family for their love and support.

Table of Contents

	Page
Abstract.....	iii
Acknowledgements.....	iv
Table of Contents.....	v
List of Figures.....	vii
List of Tables.....	viii
List of Appendices.....	ix
Introduction.....	1
Categories of Usability Tests.....	1
Usability Test Procedure.....	2
Number of Participants.....	4
Important Considerations.....	5
Advantages of Usability Testing.....	6
Disadvantages of Usability Testing.....	6
Remote Usability Testing.....	8
Advantages of Remote Testing.....	13
Disadvantages of Remote Testing.....	14
Think-Aloud Protocol and Comments.....	16
Current Study.....	18
Method.....	19
Independent Variables.....	19
Input method by participants.....	19
Experimenter presence.....	20
Experimental Conditions.....	20
Dependent Variables.....	22
Comment type.....	22
Comment quality.....	23
Relevance.....	24
Usability Problems with the Web Site.....	24
Ease of use ratings.....	25
Total test time.....	25
Experimenter communications.....	25
Participants.....	26
Materials.....	27
Procedure.....	29
Data Cleaning.....	30
Results.....	31
Data Analysis.....	31
Test Time.....	32
Usability Problems with the Web Site.....	32
Ease of Use Rating.....	34
Comments.....	35
Total Number of Comments.....	36

Direct and Supplementary Comments.....	37
Direct comments.....	37
Supplementary comments.....	38
Comment Completeness.....	39
Comments requiring no follow-up.....	40
Comments requiring follow-up.....	40
Relevance of Comments to Usability.....	42
Relevant comments.....	43
Irrelevant comments.....	44
Experimenter Communications.....	45
Discussion.....	46
Differences Due to Experimenter Presence or Absence.....	46
Differences Due to Input Type.....	48
Comparisons Between Remote Testing Conditions and the Local Control Condition.....	48
Technology Issues in Remote Usability Testing.....	50
Coding Comments.....	52
Practical Problems with Remote Testing.....	53
Generalizability.....	54
Conclusions.....	54
Future Research.....	58
References.....	59

List of Figures

Figure	Description	Page
1	Interaction Between Presence and Input Type on Number of Comments Requiring Follow-up.....	41

List of Tables

Table	Description	Page
1	Mean Number of Direct and Supplementary Comments for Attended and Unattended Usability Test Conditions.....	37
2	Mean Number of Direct and Supplementary Comments for Voice and Typing Input Usability Test Conditions.....	38
3	Mean Number of Relevant and Irrelevant Comments on Presence.....	43
4	Mean Number of Relevant and Irrelevant Comments on Presence.....	43
5	Comparisons of Remote Conditions to the Local Condition	56

List of Appendices

Appendix	Description	Page
A	Usability Test Screen Shots.....	62
B	Screen Shots of Desktop Arrangements.....	73
C	Informed Consent Forms.....	77
D	Participant Instructions.....	79
E	Trust Questionnaire.....	88
F	Written Debriefing.....	91
G	Tables of Means for Independent Variables and Experimental and Local Conditions.....	92

Collecting subjective data in remote usability testing

Within the usability evaluation literature, there has not been much emphasis on the collection (or quality) of subjective data in remote usability testing. This thesis examines the effects of the presence or absence of an experimenter in remote usability testing. Emphasis is placed on information that is normally spoken such as thoughts, comments, and answers to questions. The study also investigates the effects of input method – voice or typing -- on the quality of data collected.

The purpose of usability testing is to evaluate the user interface of a product with representative users. If users cannot achieve what they want, the site or product is not usable. Participant satisfaction ratings and opinions about the product's ease of use are also important information collected in a usability test. Usability testing emphasizes data based on participant performance and comments as opposed to alternative approaches (e.g., expert inspection methods, walkthroughs, etc.) that emphasize expert opinion about the quality of the interface or whether an evaluator *thinks* users will be able to achieve their goals.

Categories of Usability Tests

Usability tests can be broken down into two main categories: formative and summative (or acceptance). Formative evaluations are conducted as part of an iterative design cycle to help improve the interface throughout its creation (e.g., Nielsen, 1993). The evaluations should be started early in the design (Redmond-Pyle & Moore, 1995). Formative evaluations uncover which aspects of the interface design work and do not work. A working system is not required for these evaluations, which test prototypes of

only parts of the system. Formative evaluations address contentious or critical user interface design issues.

Summative evaluations are conducted when the product is completed or a comprehensive simulation is available. They test the overall quality of the interface before the product is released. Unlike formative evaluations where testing sessions are scripted, the tasks in summative evaluations are broad scenarios where users decide what goals they wish to accomplish. Both summative and formative evaluations can be used to compare two or more interfaces. Products that have not undergone formative evaluations are unlikely to pass summative tests.

Usability Test Procedure

There are many steps to conducting a usability test. Formative and summative usability testing typically use the same methodology. First, problem statements or test objectives must be developed (Rubin, 1994). Participants that match the intended users of the product are recruited to participate in usability testing and are often tested one at a time (Kantner & Rosenbaum, 1997), although some experimenters test participants in pairs.

Participants may or may not be chosen at random (Rubin, 1994). Due to time constraints, budget restrictions, and long distances from actual product users, companies will occasionally use convenience samples that, it is hoped, are representative of the user population for the product. For example, some companies occasionally use their own staff in an attempt to create a sample with representative knowledge and skills, not a sample of actual users, to participate in their usability tests. This approach may or may

not be acceptable, in terms of generalizability, depending on how similar the users behave and think compared to the users of the product.

Often, participants are tested in a controlled laboratory, but there are occasions when the work environment is an important part of the task. There are many different possible arrangements for the testing room. Some testing environments may have the user and tester in one room, while there are other people observing through one-way mirrors or by live video. Sometimes usability test observations are recorded with the use of video and/or audio recording equipment. Usability tests are conducted using scripts and are facilitated by the experimenter. The test administrator (or experimenter) is responsible for meeting/greeting participants, interacting with them during the test, and debriefing them after the test is completed (Rubin, 1994). Sometimes a second person records the relevant information obtained from observation (Kantner & Rosenbaum, 1997), because it can be difficult for the experimenter to be actively interacting with the participant and taking notes simultaneously. In some situations, a single person is the facilitator, observer, and recorder. Having a single person as an evaluator can be the case in companies with small usability teams.

It is important to construct a usability test plan that describes how you will test the product before testing. The test plan is reviewed by the members of the development team and interested stakeholders to make sure all testing needs (as outlined by the usability evaluator/team) will be met (Rubin, 1994). Creating the plan allows stakeholders to agree on the procedure. The test plan determines what resources are required (Rubin, 1994). The test plan also outlines the metrics of the test. These metrics

specify what is required for a pass or fail of performance on the task. These criteria determine when the desired level of usability is reached so that the iterative testing (in formative evaluations) can be stopped (Lindgaard, 1994).

It is important to conduct the sessions impartially, to be aware of body language and tone of voice, and not to help participants too soon when they have difficulties (Rubin, 1994). It is useful to ask participants to think aloud while they are completing the tasks and to probe participants for information about the difficulties they are having and what they are thinking when appropriate.

Information about their experience and opinions are collected using questionnaires, rating scales, or interviews upon completion of the test (Kantner & Rosenbaum, 1997). Participant thoughts and comments during testing are also recorded. After testing, the experimenters compile a report and provide recommendations for the improvement of the design of the product (Rubin, 1994).

Number of Participants

Often 4 to 10 participants are tested in a usability test (Rubin, 1994). Nielsen and Landauer (1997) suggest five users are sufficient, while researchers such as Spool and Schroeder (2001) argue that 5 users are insufficient. Spool and Schroeder performed what they considered to be a type of usability test to examine reliability of results with small sample sizes. The Spool and Schroeder usability test methodology was unusual. Spool and Schroeder gave participants money and asked them to go through a particular site and purchase whatever they wanted, although they were not required to make a purchase. This was the method used to identify usability problems. Spool and Schroeder

did not use a formal usability plan, which consists of specific tasks to perform. Therefore the usability test was not consistent with other usability tests so the results were not consistent with previous research, and do not lend themselves to understanding the number of participants required in a usability test. The method used by Spool and Schroeder, which encourages a variety of different experiences, leads to their conclusion that 5 users are insufficient to discover all problems. The number of participants required for usability tests is the subject of an ongoing debate.

Important Considerations

Usability testing is hypothesis-driven: tasks are chosen to address user interface issues that are identified as potential problems. One key part of conducting usability testing is knowing what to test and why. Usability specialists, and often relevant stakeholders (e.g., project managers and software engineers), identify what aspects of the interface they feel users might have difficulty with (e.g., through heuristic evaluation) and create tasks that will identify if users have trouble with those parts of the system or not. The tasks must be chosen carefully to ensure that they target the problem areas; it is possible to create tasks that do not address the issues. For example, if the tasks designed for the a usability test do not cover areas of the web site that contained problems, then these problems will not be uncovered by the participants in usability testing.

Usability testing teams should work closely with the stakeholders when identifying what issues to test, designing the tasks, and choosing the usability metrics. By having a close relationship with relevant stakeholders, including the development team, the usability tester can receive important information that may not have been

available otherwise. If the test plan is signed-off by stakeholders, it helps to ensure that critical questions will be examined (Maguire, 1997; Thomas & Beven, 1996).

Advantages of Usability Testing

Nickerson and Landauer (1997) claim that the information obtained from usability testing is the most often used information for making major usability improvements. It has been found that usability testing uncovers a great number of severe usability problems (Jeffries & Desurvire, 1992). Chapanis (1991, pp. 368) notes “Measuring usability means, almost by definition, measuring performance, that is, the behaviour of a person and a system while they are actually doing a job.” This is an advantage of usability testing because the results are based on user performance not on user self-reports. Measuring performance is superior to other methods that require the evaluators to infer usability problems and predicted impact on users. It is possible that the problems identified by evaluators are not problems with users, or that users may encounter problems that were not identified by evaluators.

Another advantage of usability testing is its impact on developers. Developers may not believe that a problem exists, but their minds can easily be changed by watching a participant struggle, an advantage that most other techniques including expert evaluation do not have (Jeffries & Desurvire, 1992). Usability testing directly shows the impact that user interface problems have on users (Jeffries & Desurvire, 1992).

Disadvantages of Usability Testing

The disadvantages associated with usability testing (and possibly remote testing) arise because this method conducts tests with users. Usability testing is considered

important to the user-centred design cycle, but in some circumstances it is not possible or feasible to evaluate a product or web site with this method. Due to the necessary experience, time, and cost, usability testing can be impractical.

As with other evaluation methods, experience and training of the evaluators are vital to conducting a good test. Also, interface evaluations may be subjective. In inspection methods, the subjectivity lies within the experimenter being solely responsible for identifying usability problems. The problem of subjectivity and inexperience can enter into every aspect of usability testing, for example, in the creation of the tasks. If the tasks are not created properly the experimenter will not get valid results. Kessner (2000) found very little agreement between evaluation teams on most of the usability problems identified in his study, but nevertheless, concluded that usability testing is still a useful technique because it identified serious usability problems, which helped to improve the design.

Due to the reliance on testing with users, usability testing can be very time consuming, which places it at a disadvantage compared to other methods for identifying user interface problems (Jeffries, Miller, Wharton, & Uyeda, 1991). Usability testing is time consuming because participants are tested singularly or in pairs. Furthermore, usability testing has a greater cost than inspection methods, like walkthroughs (Jeffries & Desurvire, 1992) because of the increased time testing participants individually, and the cost of paying participants. Time and costs are also higher for the preparation of usability test materials.

In some cases the actual users of the product are not located near the testing facility and it is, therefore, expensive for participants to travel to the testing facility. It is also time consuming and expensive to send experimenters to the user's location. Also, it may not be convenient for the experimenter to transport the appropriate equipment needed for the usability test. An alternative to having users or experimenters travel is remote usability testing.

Remote Usability Testing

Remote usability tests differ from conventional usability tests because the experimenter is not physically present during testing, and participants are at a distant location. As with traditional usability testing, this type of evaluation attempts to recruit representative users of the product. Representative users participate in the usability test from a distant location.

Remote usability testing also differs from other remote evaluations. Other methods, aside from remote usability testing, are local evaluation at a remote site, remote laboratory testing, remote inspection, and remote questionnaire/survey (Hartson, Castillo, & Kelso, 1996). Some of these studies discuss techniques such as server and event logging and click-stream analysis, which are automated methods of capturing user behaviours. These techniques are not remote usability testing. Often events (e.g., which links a user chooses or what path a user takes through a site) are logged automatically. The users are often not assigned tasks. One disadvantage of not having any task is that the experimenter is not aware of the user's intentions; the experimenter does not know what the user has intended to do on the site and therefore does not know if the participant

chose a particular link for a reason or if they simply got lost. The experimenter also does not know if the user accomplished his/her task or not. One of the advantages of usability testing is that tasks are created to guide users to areas with potential problems. If users are not assigned tasks it is possible that they may never visit the area of the site where a potential problem was identified by the usability specialist/team. Also with these types of remote evaluations (e.g., server logging, automatic data collection, etc.) thoughts and comments from participants are usually not collected, which is a loss of potentially valuable data.

Remote usability tests are either attended or unattended. Attended remote usability testing is where participants do the usability test at a distant location with an experimenter that is in contact with the participant using various methods such as telephones, text-based conversations (“chat”), or videoconferencing. Attended usability testing can be conducted using window/application sharing software and electronic whiteboards. Window or application sharing allows both the remote user and the researcher to see the screens being used at the same time on their respective computer screens (Hammontree, Weiler, and Nayak, 1994). It is also possible to use shared whiteboards where participants can share a writing space (Hammontree et al., 1994). Computer-based videoconference applications allow the researcher to see live video of the remote participant completing the assigned tasks (Hammontree et al., 1994). The potential advantage of using computer-based videoconferences is that information such as participant reaction, gestures, and facial expressions are observed. Interaction with the researcher through videoconference may help create a rapport with the remote

participant, which may help with understanding performance on the tasks and the product (Hammontree et al., 1994). Sometimes a simple telephone connection improves the remote test by allowing verbal communication between the remote user and the researcher.

Unattended testing occurs if the participant has no contact with the experimenter during the usability test. Unlike local testing, where participants usually give verbal responses to questionnaires and rating scales, which are recorded by the experimenter, unattended remote usability test participants respond by recording their answers themselves by typing or writing. With this type of remote testing, the task scenarios, instructions, and questionnaires may be faxed or emailed to the participants in advance (Hammontree et al., 1994), included in the software that the participants work with (Macko, 1998), or built into an online test. In this situation it may also be useful to record verbal user comments in audio form (Hammontree et al., 1994).

There are companies that specialize in remote evaluation of web sites. One company, NetRaker, claims to evaluate customer experience and information relevant to retaining customers, thereby creating greater revenues (Ann Sangpan, personal communication, April 23, 2002). NetRaker programs record navigation patterns, user clicks, and time spent on the clients' site, or time spent completing a particular task. Observation is also possible through video capture, screen sharing, chat, and online whiteboards. Participants are recruited via email for these unattended tests. Once a person has agreed to participate, they are directed to a website where they are asked to complete tasks on the site. Experimenters are able to receive visual presentations of the

navigation paths users have taken as well as how long they have spent on each web page. As participants complete the tasks they are prompted with questions to answer regarding their opinions about their experience.

There is in fact currently little published work on remote usability testing, with the exception of studies that evaluate the use of specific software tools in remote testing. Also, there are few studies published examining the use of think-aloud protocols for evaluating web site. This point is further made by Van Waes (2000), who notes that more work in this area is needed.

In a study by Scholtz (2001), WebVIP was used to help automate the collection of data of user paths from an unattended remote usability test. Experimenters also video-taped user's screens and provided them with rating sheets. This study concluded that user's ratings were not reliable. Users in this study did not rate the tasks they missed, which should have been rated as the most difficult. Scholtz suggests that this is because the users believed that they had accomplished the tasks. An experimenter remotely attending such a test would follow up these occurrences.

A study by Hartson et al., (1996) compared traditional lab evaluation to remote usability evaluation using video conferencing to examine differences in quantity and quality of data collection. In addition to the video, a telephone connection was used for voice communication. Online questionnaires were provided to participants in the remote condition. Each participant was given five tasks to complete. Hartson et al. found no significant differences in the total number of usability problems, noted by the evaluator and his/her time and navigation logging instruments, between the local and remote

condition. Also, the groups did not differ in their responses to the nine-item questionnaire or the three-open ended questions. The researchers conclude that remote evaluation with desktop video conferencing is practical.

In two studies by Tullis, Fleischman, McNulty, Cianchette, and Bergel (2002), comparisons were made between local usability test and remote usability test of web sites. In these studies, the remote tests were presented online and were unattended. The web site was presented on the Internet to participants with two independent browsers open. One browser displayed the web site being evaluated and the other presented tasks and collected user feedback. The results of their first study showed that the remote participants completed more tasks and took less time to complete the test than local users. However the differences were not statistically significant. Tullis et al. found that the local test identified more usability problems than the remote test --15 issues not identified in the remote test. However, the remote test identified 5 problems not identified by the local condition. In a second study by Tullis et al., remote and local test participants were equivalent on task completion rates and task time. Again, there were differences in the number of usability issues identified but in this study the remote condition identified more issues. These researchers concluded that the behaviour of participants in both conditions is similar and that remote usability test participants provided valuable comments.

In 1998 Macko, compared attended and unattended usability testing. At that time, there were no published studies that compared attended and unattended usability evaluation. In Macko's study the attended condition was a traditional local test and the

unattended test was conducted via the Internet. The unattended participants in this study typed their comments based on questions asked in the web prototype, and recorded their own start and end times. Macko found no group differences in the forced-choice questions but did find differences in open-ended questions. He attributes the differences in the open-ended question to the high frequency of probing made by the experimenter in the attended condition. Overall, the quantity of comments collected in the unattended condition was less than in the attended condition, with two exceptions. There were no differences in mean number of process comments per question (comments that would require a response from the experimenter if he/she was present) and mean number of words per comment. In his study, Macko found that both groups discussed similar themes and observations. However, he found that the quality of the unattended comments was lower than the quality of comments in the attended condition. Specifically, unattended participants provided fewer unique comments per group and per participant. The attended and unattended participants made an equal number of irrelevant comments. Despite its shortcomings, Macko concluded that the unattended condition provided useful comments and that unattended usability testing is a viable alternative to attended usability testing.

Advantages of Remote Testing

Macko (1998) outlines many advantages to remote usability testing. First, he argues that, in unattended testing, there is a reduction in cost in implementing remote testing because a usability specialist is not required to be physically present during testing

or required to travel to the remote location. Costs are saved because companies do not need to pay participants to travel to the researcher's location.

A second advantage is that it may be difficult to get experienced users for some products locally. With remote testing, researchers are able to contact participants globally thereby increasing the number of experienced participants potentially available for a test. For example, the job searching service PISA (Patrick, <http://www.pet-pisa.nl>) which is being developed in North America for a European audience could be evaluated anywhere in the world. Third, typically with remote testing, users are participating in the test from their normal working environment under normal working conditions, which may help to increase the external validity of the test. Fourth, there may be less chance for tester biases to be introduced in unattended remote usability testing. For example, if the tester does not like the product they may subconsciously record more negative comments and not record positive comments. Also, hints will not be given accidentally when the usability specialist is not present, as in unattended testing. Remote testing is advantageous because participants can be recruited from all over the world, thereby, allowing clients to examine differences among cultures (9AM Solutions, 2002).

Disadvantages of Remote Testing

Macko (1998) also outlines some of the disadvantages associated with unattended remote testing. First, when remote evaluations are unattended and not observed by researchers, data is limited to self-report. Second, if the participant experiences some difficulties (such as not understanding the task, or being unable to complete the task) the tester is not present to answer questions or provide hints. This may be frustrating to the

remote participant. Third, there is a chance that participants may be dishonest and lie about how they performed on the tasks. Fourth, in addition to the problem of dishonesty there is the problem of accuracy. It is possible that participants may not understand the questions or they may race through the questionnaire just to finish the task and move on to other activities. Fifth, if a participant gives incomplete answers when a usability tester is present, the tester can probe the participant for a more detailed response. This is not possible if the tester is not present, physically or via telecommunication. Sixth, another difficulty with unattended remote usability testing is that users are required to type their responses instead of simply relaying their ideas verbally to an experimenter taking notes on what is said. Typing may be a hindrance to some because of poor typing skills, the extra effort to type responses instead of verbalizing them, and the difficulty of typing responses about a task while completing the task.

Bartek and Cheatham (2003), also discuss some possible disadvantages of remote usability testing. First, if audio or video is not used, the experimenter cannot rely on the non-verbal cues provided by participants to determine if they are confused, frustrated, or tired. Second, there may be issues of security. If the material being evaluated is confidential, the experimenter cannot be sure that the remote participant is not showing it to other people in the room or taking screen shots of the material. Third, there may be system performance issues which interfere in the test, for example, servers that contain the application may be slow or stop working during the test. Another concern is whether or not difficulties will be met if the participant has a firewall, which would disrupt the test. Fourth, they believe that remote tests may actually be more time consuming because

information must be put online, scheduling online meetings and conference calls, and there may be practice set-up sessions with the participants. The additional costs associated with preparing standalone remote usability testing materials may partly offset some of the savings with remote testing.

Think-Aloud Protocol and Comments

In a typical usability test, subjective data (think aloud data, unsolicited comments, subjective ratings, etc.) collected in usability testing can be useful. Think aloud protocol is a technique in which the experimenter asks participants to express out loud what they are thinking while they complete tasks. Participants voice their expectations about where they expect to find functions and instructions, how they think a system will react, and their response to the feedback provided by their actions. During tasks, the experimenter can also prompt the user to think aloud or to ask for more detail about the thoughts they have spoken. Think aloud data helps to identify misunderstandings created by interface elements and provides an understanding of not just what methods are used to complete the tasks, but why users choose these methods (Nielsen, 1993). Often, user comments are solicited at the end of each task along with subjective ratings of the tasks completed. These comments differ from the unsolicited comments and data collected in think aloud protocol because comments are retrospective accounts of how users felt about the task, the layout, the navigation, etc. Think aloud protocol data gives insight into the thought processes of the participants, for example, where users expect to find functions and how they feel about the tasks while actually performing them.

Ebling and John (2000) concluded that the verbal protocols collected in their study provided the most evidence of usability problems compared to quantitative data such as time required to complete an exercise. A total of 65 problems were logged, with 61 identified by the verbal protocols. Of these problems, 40 were only identified by the verbal reports. However, verbal reports identified more, less critical usability problems than the quantitative data. Despite this, the researchers conclude that think aloud protocols are valuable.

Verbalizations may be either concurrent or retrospective. Concurrent verbalizations consist of information that is verbalized while the participant is attending to the task, and retrospective verbalizations are comments from participants who are asked, after completing the tasks, about their cognitive processes during the task (Ericsson & Simon, 1980). In studying the differences between concurrent and retrospective verbalizations Ericsson and Simon (1984) found that retrospective verbalizations often lack the information obtained from concurrent verbalizations. Therefore, concurrent verbalizations are generally desired in usability testing.

With concurrent verbalizations, participants are asked to think aloud while they complete the assigned tasks. By contrast, retrospective verbalizations are accounts of thought processes that occurred during the usability test.

Ericsson & Simon (1984) developed a model to characterize three levels of verbalizations. In Level One verbalizations, users are processing verbal material so that the input and output are both verbal and no translation needs to take place in the user's head. At level three, the input material appears in another form such as graphs or other

images, and the user must translate what is seen into words. This translation process, Ericsson and Simon argued, is likely to interfere with the way the user performs the task at hand and will yield different results from the situation in which the user is not asked to speak while doing the task. Level Two verbalizations fall between these two extremes. Ericsson and Simon (1980, pp. 247) argue that verbal reports generated in Level One or Level Two provide can provide information that is both useful and valuable for understanding the individual's thoughts. In this study, verbalizations fell mainly into Ericsson and Simon's (1984) Level One category, with some falling into the Level Two category. Therefore, verbal protocols were expected to yield valuable information here.

Shapiro (1994) indicates that there can be problems associated with think-aloud procedures. He notes that not all of the information that an experimenter requires goes through short-term memory and that a participant may not be able to report what is in his/her short-term memory adequately. Also, it is possible that the process of reporting thoughts can interfere with task performance.

Current Study

The purpose of this study is measure the effects of experimenter presence or absence in remote usability testing and to investigate the effects of participant input method on the quality of data collected. The focus, in terms of data, is on information that is normally spoken, including thoughts, answers to questions, and unsolicited comments.

Thoughts, comments, and answers to questions are collected in two forms: audio recording of spoken material and participant typing of thoughts and comments that would

normally be spoken. Two forms of experimenter presence, attended and unattended remote usability testing, is tested. To determine which method will provide more useful thoughts and comments, the verbal and typed responses of participants in both attended and unattended conditions is compared. There was also a control condition in the form of a traditional local attended test where the participant and the experimenter are in the same room during testing.

This study examines the differences between conditions on comment type, quality and relevance, web site performance, ease of use ratings, total test time, and experimenter communications.

Method

Independent Variables

Input method by participants. Two methods of input were used in the usability tests. Participants in the voice conditions provided their thoughts, comments, and answers to questions by speaking into a headset microphone. The audio for the experimenter and participants was recorded on the experimenter's hard drive by Camtasia recording software by TechSmith.

In the typing condition, participants provided their thoughts, comments, and answers to questions by typing into Microsoft NetMeeting's instant text messaging program (chat program) which was open on the participant's computer desktop. The chat sessions were saved as HTML files. For the control condition, which was a traditional local test, participants provided their comments by speaking to the experimenter who was sitting beside them.

Experimenter presence. Experimenter presence refers to whether or not the experimenter interacted and communicated with the participant during the usability test. In the unattended conditions the experimenter had no communications with the participant. In the attended conditions, the experimenter communicated with the participants remotely via an audio link using RAT (Reliable Audio Technology) program, using a headset microphone or by typing in the chat program. In both the attended and unattended conditions, the experimenter was able to observe actions and behaviours of the participant by watching the screen of the participant displayed on her own computer screen. This screen sharing was possible by using a software program developed at AT&T called Virtual Network Computing (VNC). Participants in the attended conditions were told prior to the onset of the test that the experimenter could see their screen from the other room. The purpose of informing participants that the experimenter could see their screen was to avoid confusion if the experimenter made references to participant behaviours during communications.

The control condition was also attended, with the experimenter seated beside the participant throughout the test. The experimenter communicated with participants in the attended conditions, remotely and locally, throughout the testing session in terms of asking for clarifications and reminding participants to answer questions.

Experimental Conditions

There were four experimental conditions (all remote) and one control condition (local):

- 1) Attended/Typing - In this condition, the participant had contact with the experimenter through the chat program. The participant provided all of his/her thoughts, comments, and answers to questions by typing into the chat program. The experimenter also communicated with the participant by typing into the chat program.
- 2) Attended/Voice - In this condition, the participant had voice contact with the experimenter via an audio link (RAT) and a headset with a microphone. The participant provided all of his/her thoughts, comments, and answers to questions by speaking into the microphone. The experimenter also communicated with the participant when appropriate by speaking into a microphone.
- 3) Unattended/Typing – In this condition, the participants had no contact with the experimenter and provided all of their thoughts, comments, and answers to questions by typing into the chat program. Participants were informed at the beginning of the test that no one would respond to the messages they typed into the chat program. The purpose of providing this information to participants was to avoid confusion in terms of using a two-way text-messaging program and receiving no responses.
- 4) Unattended/Voice - In this condition, the participants had no contact with the experimenter and provided all of their thoughts, comments, and answers to questions by speaking into the microphone.
- 5) Attended/Local testing (Control Condition) – This condition followed the protocol for a traditional usability test where the experimenter and participant are

located in the same room during the test. The participant's thoughts, comments, and answers to questions were provided verbally to the experimenter who was physically present during the test. The experimenter and participant's voices were captured by a microphone, placed on the participant's desk, and were recorded with the Camtasia software. The experimenter was able to ask the participant for more detailed responses, etc.

Dependent Variables

A comment is any statement made by a participant. Examples of comments are "Didn't I just fill out this privacy thing!", "There should be a drop-down box here." or "This is stupid." Utterances that were not counted as comments included when participants read the instructions out loud to themselves or when they said their personal information aloud as they filled in the text fields (e.g., "Name...Jane Doe...birthday...January 21st", etc.) were not counted as comments. If a participant made the same comment twice (for example, pointed out their dislike for the background colour on two separate web pages) it was counted as one comment.

Comment type. Comment type is made up of two variables, Direct Answers to Questions and Supplementary Comments. Direct answers to questions are solicited comments participants made in response to the questions presented to them in a frame on the left side of the screen in the experimental conditions or read to them by the experimenter in the local condition. Supplementary comments are unsolicited comments made by participants. For example, there was a text box in the job search website that

was labelled “job sector”, and participants were asked what job sectors are. When participants answered this question their comments were coded as direct answers to questions. Some participants continued to comment on that aspect of the interface by saying it should be a drop-down box with a list of possible job sectors instead of a blank text box. These types of comments were coded as supplementary comments.

Comment quality. Comment quality considers whether or not participant comments need to be followed-up by the experimenter. If the comment provided by the participant was unclear, the experimenter (when present) may need to ask the participant for clarification. Comments were coded as either “no need to follow-up”, “would have followed-up if present”, “should have followed-up and didn’t”, and “should have followed-up and did”.

The comments “would have followed-up if present” represent the number of unclear comments that participants in the unattended conditions made where it was not possible for the experimenter to ask for clarification. The comments “should have followed-up and did” are the number of unclear comments that participants in the attended conditions (including the local condition) made where the experimenter had the opportunity to follow-up (e.g., to ask for clarification) and did. The category “should have followed-up and didn’t” are the number of unclear comments participants in the attended conditions (including local condition) made that the experimenter should have followed-up but for various reasons did not. For example, in some occasions it was difficult for the experimenter to interrupt in the typing conditions. By the time the experimenter was able to type a comment to the participant they had already moved onto

another web page. Because of the instability of the prototype it was often a risk to ask participants to go back because the web site was known to crash. This category was included to allow a fair comparison between attended and unattended conditions. The three follow up comment variables were summed to create a measure called “follow-up needed” which reflected comments made by participants in both the attended and unattended conditions that needed some clarification.

Relevance. Relevant comments are the comments participants made that were related to the website that was being tested in the usability test. For example, a relevant comment would be “The colour is bland.” Irrelevant comments are comments made by participants that are not related to the website. For example, an irrelevant comment would be “I’m from New Brunswick.”

Usability Problems with the Web Site. During the usability test, participants were asked questions to determine whether or not they understood the purpose of various interface elements, terminology, and procedures (e.g., “What is the purpose of naming your task?”). Answers were coded in terms of whether or not participants understood. Participants received either “yes”, “no”, “partially understood”, “didn’t answer”, “not applicable”, “restated question or read from interface description”, or “can’t tell” (meaning that it could not be determined if the participant understood based on the answer provided). One of these designations was given for each of the questions asked. There were two behavioural indicators in the usability test, indicative of whether or not participants understood what they were doing. The experimenter observed which link participants initially chose on the main page (i.e., to the “Register Now” link, which was

for new users, or if they attempted to login directly from the main page despite the fact that they had not yet created a user account). The other indicator was whether or not participants were able to register. The designations outlined above (e.g., “yes”, “no”, etc.) were also assigned for these behavioural indicators. To obtain total performance scores, the total number of “yes” designations for each participant was summed. The highest possible score is 27.

Ease of use ratings. At the end of each testing session, participants were asked to rate the ease of use of creating their job search. They were asked to rate the ease of use on a scale of 1 to 7, with 1 being “very difficult”, 4 being “neither easy nor difficult”, and 7 being “very easy”. Ratings on this issue indicate the adequacy of the web site usability. One score was obtained for each participant.

Total test time. Total test time is the length of time participants took to complete the usability test.

Experimenter communications. Experimenter communications are instances where the *experimenter* spoke or typed to the participant. There are six categories of communications: queries to ask for clarification; confirmations to make sure the experimenter understood what the participant was saying; reminders to fill out information on the website or reminders to answer the questions; answers to participant’s questions; explanations of something to a participant; and probes to ask the participant if they want to make a comment. These measures apply only to the attended and local control conditions.

Participants

Fifty individuals were recruited with 10 participants assigned randomly to each of the five conditions. Participants were recruited from Carleton University through posters, the Psychology 100 notice board, and also by approaching individuals in common areas. Other volunteers were acquaintances of the experimenter or individuals that were referred by other participants.

The qualification for participation was familiarity and experience with the Internet, English as a native language, and the ability to type. Individuals were asked if they had experience using the Internet. All participants reported being familiar with the Internet and no one, as observed by the experimenter, encountered problems due to lack of Internet experience. The participants were native English-speakers to ensure that the differences in thoughts, comments, and answers to questions were due to the experimental conditions and not language ability. Participant's typing ability was tested before the session began to ensure that differences in thoughts, comments, and answers to questions were not due to a lack of typing skill. To maintain consistency among conditions, all participants completed a two-minute typing test using TypeMaster Typing Test 6.2, which is a free software program, downloaded from the Internet. Participants were required to obtain a minimum of 20 words per minute (gross) on the typing test. All participants met the typing criteria.

After the initial recruitment, a follow-up email was made to provide the time, date, and location of the test and contact information. In addition, participants were

contacted via email or phone one day prior to their testing session as a final reminder and confirmation of participation.

Participants were given an honourarium of 10 dollars for their participation in the study. One participant's testing time substantially exceeded one hour, so she was paid an additional 10 dollars.

Materials

The job searching web site, called MyAssistant (Patrick, <http://www.pet-pisa.nl>), used in the study employs a software agent to match the participant's personal and work history information with job postings on the Internet. The agent returns information to the users when appropriate matches have been found. The MyAssistant site is a major project concerning software agents. It was developed by Dr. Andrew Patrick from the National Research Council. He is interested in the evaluation of various aspects of the site such as usability, privacy, and trust. This thesis examines the usability aspects of the MyAssistant site.

Before the tasks and questions used in this experiment shown in Appendix A were decided upon, the experimenter examined the web site being evaluated and created a list of potential usability problems. The tasks and questions that were used in the test were created to help identify if these potential usability problems, were in fact, problems for actual users. Many non-performance questions were included to help identify whether or not participants understood what they were doing. In some situations the potential problem may not have been easily identified by observation alone so a question was added in addition to a task. The MyAssistant web site is large and is in its early stages of

development with many potential usability problems identified, so the whole site was not evaluated. Only one aspect of the site, creating a task, was evaluated. Eliminating the other aspects of the site from the usability evaluation allowed the experimenter to ask all necessary questions and tasks required to set up a job search, without creating a usability test that would take more than an hour to complete.

All testing was conducted at Carleton University's Human Oriented Technology Laboratory (HOT Lab). Two separate rooms were used to simulate distance for the remote testing; the participant was in one testing room while the experimenter was in another. The participant and experimenter workstations consisted of a computer using Microsoft Windows 2000 and 19'' monitors. Each computer also had Internet Explorer and NetMeeting installed. At each workstation there was also a headset with microphone, used in the unattended/voice and the attended/voice conditions.

AT&T's VNC screen sharing software was installed on both the participant's computer and the experimenter's computer. VNC allowed the experimenter to view the participant's screen on the experimenter's monitor in real time. The experimenter also had Notepad (a text editor) open on the screen for note taking during sessions. The monitor also displayed the NetMeeting chat program, used in the attended/typing condition. Audio and all information from the experimenter's monitor were recorded into an AVI file by TechSmith Camtasia recording software for later data analysis. Appendix B shows screen shots of experimenter and participant screens.

Microsoft Internet Explorer was used during sessions to view the web site being tested. For all conditions, the web browser displayed a job searching web site with

frames on the left side of the screen outlining task instructions and questions to be addressed by participants. In the local condition, the task instructions and questions were read to participants by the experimenter, which is typical in usability tests. Therefore, the frames containing the instructions and questions were hidden from participants in this condition. This was achieved by sliding the Internet Explorer window to the left until the frames were off the screen and no longer visible to the participants.

Procedure

The order of testing the five conditions was randomly determined within blocks of five participants.

Upon arrival, participants were greeted and asked to read and complete the informed consent form shown in Appendix C. After completing the form, participants were given the two-minute typing test. Once the test was completed, participants were given a brief overview of what was to take place during the testing session. Participants were also reminded to read the instructions thoroughly (with the exception of the control condition) and were asked if they had any questions. Before the test began, the experimenter started the screen recording software.

The first page encountered by participants in the experimental conditions was an instruction page. Appendix D contains all instruction pages. The instruction page provided information about the web site and its purpose, explained the purpose of the usability test, and outlined the importance of providing detailed thoughts, comments, and answers to questions. A detailed example, shown in Appendix D, was provided on the type of information that was needed from the thoughts and comments. The instruction

page also explained that the task instructions and questions would appear in a frame on the left side of the screen and how, depending on the specific method of input being used, participants were to provide their thoughts, comments, and answers to questions. The purpose of the instruction page was to ensure that all participants fully understood what type of data needed to be obtained and how to provide it.

At the bottom of the instruction page, participants were instructed to re-read the page if they did not fully understand. After participants read the instructions, they began the usability test by clicking a link at the bottom of the page. Clicking the link brought participants to the main page of the MyAssistant web site.

In the control condition (local condition), participants were read the instruction page by the experimenter from a paper copy. Participants were given a copy to follow along. After the instructions were read, the experimenter turned on the participant's monitor, where the MyAssistant main page was displayed, to start the test.

Participants were reminded, in the task instructions on the left side of the screen, to use fictional information when filling out personal data.

After completing the usability test, participants were given a 16-item questionnaire, shown in Appendix E, to complete prior to leaving the lab. This questionnaire for a separate study on trust, required participants to read the written debriefing form shown in Appendix F prior to receiving an honourarium.

Data Cleaning

The audio for two participants in the attended/voice conditions could be heard clearly and loudly, but the experimenter's voice did not record at all. When this occurred,

extrapolations of what experimenter communications were being made, was required. For example, if the experimenter heard “Do I use my own information?” followed by a pause and then “Ok, thanks” from the participant, the experimenter coded the pause as an answer to a question. It was clear from the participant's question and response that the experimenter was answering the participant’s question.

Results

Data Analysis

The design of this study is a 2X2 factorial with an additional control condition (the local condition) with data from a number of dependent variables. Each dependent variable required two basic analyses. The first series of analyses were factorial ANOVAs, which include the between-subject independent variables of presence of the experimenter (attended versus unattended) and input method by users (voice versus typing). The factorial ANOVAs provided information regarding the main effects and interactions among the experimental conditions, but did not include any information about the control condition. A second series of analyses incorporated all of the usability testing conditions as five levels of a between-subjects independent variable. This analysis allowed comparisons of the four experimental conditions with the control condition.

Significant main effects of the analyses involving the five conditions were followed by Dunnett’s test that compared all experimental conditions with a control condition.

Means on each dependent variable for each experimental and local condition are presented in Appendix G.

Test Time

It is of interest to see how the remote test conditions influenced the total amount of time required to complete the usability test.

Participants who had contact with the experimenter during testing (attended) took longer to complete the usability test ($M = 37.00$ minutes) than those who did not have contact with the experimenter ($M = 28.17$ minutes), $F_{(1, 36)} = 5.18, p < .05$. The main effect of type of input method (voice vs. typing) and the interaction between type of input and presence of the experimenter were not significant.

The analysis that included all five conditions resulted in a significant main effect of condition, $F_{(4, 45)} = 4.10, p < .01$. Participants in the local control condition took less time to complete the usability test ($M = 23.21$ minutes) than did the participants in the conditions that had contact with the experimenter and provided their comments by typing (attended/typing) ($M = 42.27$ minutes), $p < .01$. No other remote test conditions differed significantly from the local condition.

Usability Problems with the Web Site

To investigate whether or not the remote test conditions had an effect on the ability to find usability problems with the web site, usability performance scores were created. During the usability test, participants were asked questions to determine if the web site interface was clear. Could users understand the purpose of various interface elements, terminology, and procedures? For example, participants were asked, "What is

the purpose of naming your task?" Answers were coded in terms of whether or not participants understood these concepts. Participants received one designation for each of the questions asked, either "yes they understood", "no they did not understand", "they partially understood", "they didn't answer", or "the experimenter couldn't tell if they understood". Behavioural indicators in the usability test, indicative of whether or not participants understood what they were doing, "Did participants go to Register Now first?" and "Was the participant able to register?", also received the designations outlined above ("yes", "no", etc.). To obtain total performance scores, the total number of "yes" designations for each participant was summed. The highest possible score, indicating that the participant understood all web usability issues queried, was 27.

Participants who had contact with the experimenter during testing (attended) understood more web site concepts ($M = 18.90$) than those who had no contact with the experimenter (Unattended -- $M = 15.55$), $F_{(1, 36)} = 8.68$, $p < .01$. There was no significant difference in total performance scores for voice and typing input and no interaction between type of input and presence of the experimenter.

The analysis that included all five conditions found that none of the experimental conditions differed significantly from the local condition in total number of web site concepts understood.

The factorial and one-way analyses were also performed on the number of "no" designations that indicated potential problems with the web site usability. This measure reflects the number of answers that were wrong. There was no significant interaction and no significant main effects in the number of concepts that were not understood, smallest p

> .05. Subjects in all four remote usability testing conditions and the local control usability test missed similar numbers of web site concepts, $M = 4.96$.

In the unattended conditions, participants often failed to answer some of the questions and some participants simply restated the question. Because these were the unattended participants, the experimenter was not able to remind the participant to answer the question. To see if this happened differentially for the testing conditions, number of missing responses (“didn’t answer”) was analysed.

Participants in the unattended conditions had more missing responses ($M = 4.35$) than the participants in the attended conditions ($M = 1.30$), $F_{(1, 36)} = 19.15$, $p < .001$. No significant interaction and no main effect of voice versus typing input were found.

To summarize results for the usability of the web site, participants in the attended conditions performed better than participants in the unattended conditions in terms of the number of web site concepts that they understood. However, they did not differ significantly in terms of the number of web site concepts that were not understood. The unattended condition did, however, have significantly more missing data than the attended conditions.

Ease of Use Rating

At the end of each testing session, participants were asked to rate the ease of use of creating their job search on a scale of 1 to 7, with 1 being “very difficult”, 4 being “neither easy nor difficult”, and 7 being “very easy”. Ratings on this issue indicate the adequacy of the web site usability. One score was obtained for each participant (with two missing values). Analyses were conducted on ease of use to determine if the testing

conditions affected how individuals perceived the ease of use of performing a job search. No significant main effects of input type and presence of the experimenter, and no interaction between the two were found. In addition, there were no differences in rated ease of use for the four remote testing conditions versus the local control condition. The overall mean ease of use rating was 4.7 out of 7. Participants' ease of use ratings were not significantly affected by presence or input type, and the ratings did not differ significantly among the conditions.

Comments

Comments are an important aspect of this study. It is of interest to see if there are differences among the conditions on the number and quality of comments provided by participants. The subjective data collected in usability testing can be a useful part of a site's evaluation, and it is important to see how the use of remote testing affects the quality of data collected.

All of the comments that were made were divided into subcategories and reanalyzed from different perspectives. The total number of comments was divided into (a) direct comments and supplementary comments, (b) complete comments and comments that would need follow-up, (c) relevant and irrelevant comments.

Total Number of Comments

Total number of comments is the sum of all the participants' comments. These analyses determine if participants in the different conditions make different numbers of comments, and if the experimental conditions differ from the control condition (local condition) on the total number of comments.

Participants who had contact with the experimenter during testing (attended) provided more comments ($M = 41.30$) than those who did not have contact with the experimenter (unattended) during testing ($M = 31.75$), $F_{(1, 36)} = 9.02, p < .01$

Participants who spoke their thoughts, comments, and answers to questions during testing (voice input) provided more comments ($M = 42.65$) than those who typed their thoughts, comments, and answers to questions during testing ($M = 30.40$), $F_{(1, 36)} = 14.84, p < .001$. There was no significant interaction between input type and experimenter presence.

The analysis of all five conditions showed that participants in the local usability test condition made more comments ($M = 44.40$) than did participants in the condition that had no contact with the experimenter and provided their comments by typing (unattended/typing) ($M = 26.10$), $p < .001$. No other experimental conditions differed significantly from the local condition.

To summarize results for total number of comments, participants who had contact with the experimenter during testing made more comments than participants who did not have contact. Also, participants who provided their comments by speaking into the microphone gave more comments than those who provided their comments by typing.

The unattended/typing condition was the only remote testing condition that differed significantly from the local condition on the total number of comments made.

Direct and Supplementary Comments

To gain a better understanding of the comments that participants made, the total number of comments were divided into direct and supplementary comments. Direct answers to questions are solicited comments participants made in response to the questions presented to them in a frame on the left side of the screen in the remote testing conditions or read to them by the experimenter in the local condition. Supplementary comments are unsolicited comments made by participants. Separate ANOVAs were performed on direct and supplementary comments.

Direct comments. Mean numbers of direct and supplementary comments are shown for attended and unattended remote testing conditions in Table 1 and for Voice and typing input in Table 2. For reference, the mean number of direct and supplementary comments are shown for the local testing control condition in the final column of both tables.

Table 1

Mean Number of Direct and Supplementary Comments for Attended and Unattended

Usability Test Conditions

Type of comments	Attended	Unattended	Local control
Direct answers to questions*	25.10	22.05	26.40
Supplementary comments †	8.65	9.70	8.30

* There were significant differences on this measure for the attended and unattended test conditions.

† No significant differences on this measure for the Attended and Unattended test conditions.

Table 2

*Mean Number of Direct and Supplementary Comments for Voice and Typing Input**Usability Test Conditions*

Type of comments	Voice	Typing	Local control
Direct answers to questions †	23.85	23.30	26.40
Supplementary comments*	13.20	5.15	8.30

* There were significant differences on this measure for the voice and typing test conditions.

† No significant differences on this measure for the voice and typing test conditions.

A significant main effect of presence was found, $F_{(1, 36)} = 12.633, p < .01$.

Participants who had contact with the experimenter during testing provided more direct answers to questions ($M = 25.10$) than those who did not have contact with the experimenter during testing ($M = 22.05$). No significant interaction between presence of the experimenter and input type and no main effect of voice and typing input were found.

Multiple comparisons of the significant analysis that included all five conditions showed that participants in the local condition made more direct answers to questions ($M = 26.40$) than did the participants in the conditions that had no contact with the experimenter and provided their comments by typing (unattended/typing) ($M = 21.90$) and voice (unattended/voice) ($M = 22.20$), both $p < .01$. The two attended remote testing conditions did not differ significantly from the local condition.

Supplementary comments. Mean number of supplementary comments is shown in Tables 1 and 2.

A significant main effect of input type was found, $F_{(1, 36)} = 10.21, p < .01$.

Participants who provided their comments by voice during testing provided more

supplementary comments ($M = 13.20$) than those who typed their comments ($M = 5.15$).

There was no significant interaction between input type and experimenter presence and no main effect of experimenter presence.

None of the experimental conditions differed significantly from the local condition on the number of supplementary comments provided.

To summarize direct answers to questions and supplementary comments, participants in the attended conditions provided more direct answers to questions but a comparable number of supplementary comments than participants in the unattended conditions. Also, participants in the voice conditions provided more supplementary comments but a comparable number of direct comments than participants in the typing conditions. The only difference between the local control testing condition and the remote testing conditions was that participants in the unattended/voice and unattended/typing conditions provided fewer direct answers to questions than participants in the local condition.

Comment Completeness

The comments were re-examined to see if the completeness of the comments depended on the testing conditions. The dependent variable “comments requiring no follow-up” considers whether or not participant comments need to be followed-up by the experimenter. The dependent variable “comments requiring follow-up” considers whether or not participant comments need to be followed-up by the experimenter. For example, a comment that would not require follow-up might be “The font on this page is too small...it is difficult to read.” A comment that would require follow-up might be

“This section was pretty clear what they wanted.” In this situation, the experimenter, if present, would have needed to ask the participant what was meant by “this section” to clarify their comment.

Comments requiring no follow-up. A significant main effect of input type was found, $F_{(1, 36)} = 7.04, p < .05$. Participants who provided their comments by voice during testing provided more comments that required no follow-up ($M = 33.65$) than those who provided comments by typing ($M = 25.95$). No other significant main effect or interaction was found.

The analysis that included all five conditions found that participants in the local condition made more comments that required no follow-up ($M = 34.30$) than did the participants in the conditions that had no contact with the experimenter and provided their comments by typing (unattended/typing) ($M = 21.90$), $p < .01$. No other remote testing conditions differed significantly from the local condition.

Comments requiring follow-up. A significant interaction between presence of the experimenter and input type was found, $F_{(1, 36)} = 21.58, p < .001$. This interaction is illustrated in Figure 1. Number of comments that required follow-up is shown on the vertical axis with presence of the experimenter on the horizontal axis. Comments made by typing are indicated with squares and comments made by voice are indicated with circles. The significant interaction was followed up using a multiple comparison procedure based on the Bonferroni approach to protect against Type I errors. In this approach, the alpha of .05 is divided by the number of comparisons being made, in this

case four. This helps to protect from concluding there are differences between the groups when in fact there are none.

Figure 1: Interaction Between Presence and Input Type on Number of Comments Requiring Follow-up

Participants who spoke with the experimenter during testing (attended/voice) provided significantly more comments that required follow-up ($M = 14.30$) than participants who spoke when the experimenter was not present (unattended/voice -- $M = 4.00$), $p < .001$; and more than those who typed their comments to the experimenter (attended/typing -- $M = 5.10$), $p < .001$. When participants typed, there was no significant difference in the number of comments that required follow up whether the experimenter was present or not.

The analysis that included all five conditions found that participants in the local condition made more comments that required follow-up ($M = 10.20$) than did the

participants in the conditions that had no contact with the experimenter and provided their comments by typing (unattended/typing) ($M = 4.20$) and by voice (unattended/voice) ($M = 4.00$), $p < .01$. No other remote testing conditions differed significantly from the local condition.

In summary, participants who spoke their comments made more comments that did not require follow-up than participants who had to type, while participants who typed without the experimenter present made fewer comments that did not require follow up than the local testing condition. Participants tested with voice input with the experimenter available made more comments that required follow up than the other remote test conditions. Unattended participants with voice and typing input made fewer comments that required follow up than the local control test participants.

Relevance of Comments to Usability

All comments were re-examined to determine if they were relevant to usability of the web site. Relevant comments are comments that were related to the website being tested (e.g., “The colour is bland.”) and irrelevant comments are not related to the website (e.g., “I’m from New Brunswick.”).

Mean number of relevant and irrelevant comments are presented in Table 3 for the attended and unattended remote testing conditions, and in Table 4 for voice and typing input testing conditions.

Table 3

Mean Number of Relevant and Irrelevant Comments on Presence

Type of comments	Attended	Unattended	Local control
Relevant comments*	36.75	29.70	39.30
Irrelevant comments*	4.55	2.05	5.10

* There were significant differences on this measure for the attended and unattended test conditions.

Table 4

Mean Number of Relevant and Irrelevant Comments on Input Type

Type of comments	Voice	Typing	Local control
Relevant comments *	36.65	28.80	39.30
Irrelevant comments*	5.00	1.60	5.10

* There were significant differences on this measure for the voice and typing test conditions.

Relevant comments. Participants who provided their comments by voice during testing provided more relevant comments ($M = 36.65$) than those who provided comments by typing ($M = 28.80$), $F_{(1, 36)} = 12.22$, $p < .01$. Participants who had contact with the experimenter during testing provided more relevant comments ($M = 36.75$) than the participants who did not have contact with the experimenter ($M = 29.70$), $F_{(1, 36)} = 7.76$, $p < .01$. The interaction between type of input and whether the test session was attended was not significant for relevant comments.

The analysis that included all five conditions showed that participants in the condition that had no contact with the experimenter and provided their comments by typing (unattended/typing) made fewer relevant comments ($M = 25.30$) than did the

participants in the local condition ($M = 39.30$), $p < .001$. No other remote conditions differed significantly from the local condition on number of relevant comments.

Irrelevant Comments. Participants who provided their comments by voice during testing provided more irrelevant comments ($M = 5.00$) than those who provided comments by typing ($M = 1.60$), $F_{(1, 36)} = 16.90$, $p < .001$. Also, participants who had contact with the experimenter during testing provided more irrelevant comments ($M = 4.55$) than the participants who did not have contact with the experimenter ($M = 2.05$), $F_{(1, 36)} = 9.14$, $p < .01$. No significant interaction was found.

The analysis that included all five conditions, resulted in a significant main effect of condition, $F_{(4, 45)} = 6.84$, $p < .001$. Participants in the local condition made more irrelevant comments ($M = 5.10$) than did the participants in the conditions that had no contact with the experimenter and provided their comments by typing (unattended/typing) ($M = 0.80$), $p < .01$. No other experimental conditions differed significantly from the local condition.

In summary, the examination of the relevance of comments to the usability of the site showed that the participants in the voice conditions made more relevant and irrelevant comments than did the participants in the typing conditions. Also, participants in the attended conditions provided more relevant and irrelevant comments than did participants in the unattended conditions. Participants in the local condition provided significantly more relevant and irrelevant comments than did the unattended/typing condition.

Experimenter Communications

Experimenter communications are instances where the *experimenter* spoke or typed to the participant. There are six categories of experimenter communications: queries to ask for clarification; confirmations to make sure the experimenter understood what the participant was saying; reminders to fill out information on the website or reminders to answer the questions; answers to participant's questions; explanations of something to a participant; and probes to ask the participant if they want to make a comment. These measures apply only to the attended and local control conditions. The subtotals of each category were summed to obtain a total number of experimenter communications.

These communications only occurred in the attended conditions (attended/typing, attended/voice, and the local condition) where the experimenter had contact with the participants. It is of interest to see if the amount of communications made by the experimenter varied among the attended conditions and if the attended experimental conditions differed from the control condition (local condition).

For this analysis, a one-way ANOVA with three levels was performed. A significant main effect of all attended conditions was found, $F_{(2, 27)} = 5.80$ $p < .01$. There were more experimenter communications in the local condition ($M = 31.50$) than in the condition where participants had contact with the experimenter and provided their comments by typing (attended/typing) ($M = 19.50$), $p < .01$. The attended condition where participants provided their comments by voice had more experimenter communications ($M = 30.00$) than in the attended condition where participants

communicated by typing ($M = 19.50$), $p < .05$. However, the local condition and the attended/voice condition did not differ significantly on number of experimenter communications, $p > .05$.

Discussion

This study investigated the effects of the presence or absence of an experimenter in remote usability testing, as well as the effects of input method (voice or typing) on the quality of data collected. This study had four experimental conditions, all of which were remote, and one local control condition. Emphasis was placed on information that is normally spoken, such as thoughts, comments, and answers to questions, but consideration was also given to differences in the evaluation of the web site usability as a function of the remote testing conditions.

With one exception, type of participant input (voice vs. typing) and experimenter presence (attended or unattended) did not interact. If the type of input had an effect, the effect was the same whether the experimenter was present or not, and if the presence of the experimenter had an effect, the effect was comparable for voice or typing input. These effects are discussed separately.

Differences Due to Experimenter Presence or Absence

Participants in the attended condition provided more comments than participants in the unattended conditions. These findings suggest that when an experimenter is present (via typing or voice), participants type and speak more than if the experimenter was not present. Perhaps individuals feel obligated to make more comments when an experimenter is present than when he/she is not. Possibly they give multiple comments

for questions asked of them because they assume that the experimenter will tell them they are not providing enough information, which could not happen in the unattended conditions.

Participants in attended conditions made more relevant comments and more direct answers to questions than the unattended condition. Although the number of irrelevant comments and comments that required follow-up were low, they also made more irrelevant comments and, if they were tested with the voice input, more comments that required follow-up. The participants in the attended conditions took longer to complete the usability test. This finding of additional time is not surprising because of the experimenter communications that occurred in the attended conditions. These communications plus relevant and irrelevant comments would increase the total time of the usability test.

The web site interface evaluation would be considered better in the attended conditions because more concepts were understood by attended participants than their unattended counterparts. Participants in the attended conditions and the unattended conditions found an equivalent number of concepts that were not understood, but, although the number of missing responses was low, participants in the attended condition left out fewer answers. The attended and unattended conditions did not differ significantly on number of supplementary comments or comments that did not require follow-up, and did not differ in their ease of use ratings.

Everything considered, not having communication between participants and an experimenter sacrifices some important information and may change conclusions about the user interface.

Differences Due to Input Type

Participants in the voice and typing conditions did not differ significantly on number of direct answers to questions, ease of use ratings, total test time, or their understanding of concepts. Participants in the voice condition provided more comments and more relevant comments. They had more comments that did not require follow-up, and more supplementary comments than participants in the typing conditions. On the other hand, they also provided more irrelevant comments and, if they were tested with the experimenter present, more comments that required follow up. These findings suggest that participants provided more comments when they were speaking as opposed to typing but many of the comments were not quality comments and were not central to the usability test. The additional comments are likely due to the fact that it is much easier to speak. If it is easier to provide comments, participants might likely make more of them.

Comparisons Between Remote Testing Conditions and the Local Control Condition

The results discussed above all refer to relative comparisons among the four remote testing conditions. This section considers each of the remote conditions relative to the standard usability test control condition. Compared to the local control condition, participants in the unattended/typing condition provided fewer comments, fewer relevant comments, and fewer comments that did not require follow-up. The finding of fewer comments provided was also the case in Macko's study (1998), where participants in the

attended condition (similar to this study's local condition) provided more comments (with the exception of two variables) than those in the unattended condition (similar to this study's unattended/typing condition).

On the positive side, the unattended/typing participants had fewer irrelevant comments than participants in the local condition. This means that participants in the unattended conditions made fewer comments that were unrelated to the interface than the local participants. The finding that participants in the attended condition made more irrelevant comments is contrary to the finding by Macko (1998), who found that participants in the attended condition and unattended conditions were equivalent on the number of irrelevant comments made.

Participants in the local condition provided more comments that required follow-up than participants in the unattended/typing condition and more comments that required follow-up than participants in the unattended/voice condition. However, participants in the local condition provided more direct answers to questions than participants in the unattended/typing condition and more direct answers to questions than participants in the unattended/voice condition.

Participants in the attended/typing condition took longer to complete the usability test than participants in the local condition. This additional time to complete the test is likely due to the fact that participants were required to type their responses unlike the participants in the local condition who simply spoke.

There were only three conditions that received communication from the experimenter: attended/voice, attended/typing, and the local condition. The participants

in the local condition received more experimenter communications than participants in the attended/typing condition, but less than the participants in the attended/voice condition. This is likely due to the fact that in the local and attended/voice conditions the experimenter communications were verbal instead of typed.

None of the remote testing conditions differed from the local condition (control condition) on ease of use ratings, number of concepts understood or number of supplementary comments. This is a similar finding to Tullis et al. (2002) and Hartson et al. (1996), who found non-significant differences between the groups (remote and local) on number of usability problems identified.

Technology Issues in Remote Usability Testing

Data from the voice conditions was collected with audio streaming over the Internet and was recorded on the experimenter's hard drive. The unreliability of audio technology is a notable limitation of remote usability testing when audio is employed. There were situations, prior to the beginning of the usability test, in which the participants could hear themselves and the experimenter speaking, but the experimenter was unable to hear the participants. Fortunately, the audio difficulties were rectified with great difficulty before the testing began. Since the settings were not changed throughout the entire testing phase, the cause of this problem remains undetermined. There were also cases where the audio set-up was checked and confirmed to be functioning prior to the arrival of the participant, but failed prior to the testing. The only way to correct the problem was to restart the computers and restart the experimental set-up. These problems were corrected before data were collected and did not interfere with the usability tests

themselves, but they made the test session setup much more complicated. In this case, the implications for real-world remote usability testing with audio are serious. If these audio problems occurred in real testing situations, it would be the responsibility of the participant to correct the problem. Relying on users to correct audio problems would not be practical.

The other limitation of remote testing with using audio was that a few of the sessions did not record clearly. In one attended/voice condition session, it was very difficult to hear the experimenter and the participant because the audio was recorded too quietly. In another session (unattended/voice), the audio was recorded too loudly and it was difficult to understand what the participant was saying. All of the comments were distinguishable with difficulty in both of these situations and no data was lost.

In two attended/voice conditions, the audio for the participants could be heard clearly and loudly, but the experimenter's voice did not record at all. The experimenter extrapolated what experimenter communications were being made. This occurred for only two of 30 of the sessions recorded with audio. The experimenter is confident that the only missing data was instances where the experimenter made communications such as "Ok." and "Thanks", and is not detrimental to the overall results of the study. But some means must be found to avoid such problems if remote usability testing with voice is to be practical.

Another issue with technology concerns the chat program used. The sequence of responses in the attended/typing condition, were problematic. For example, there were situations where the experimenter was in the process of typing a comment or question to

the participant, but the participant typed a second comment prior to receiving a response to the first. The participant then acknowledged the comment or question from the experimenter for the previous exchange as a response to the most recent comment. This, of course, caused confusion and required additional effort to correct the misunderstanding. Coordination of experimenter communications in the attended/typing condition was also crucial when asking questions or providing reminders to the participants. Often, the participant had quickly moved on to a different page by the time the experimenter was able to ask or remind them about something on the previous page. In these situations, the experimenter had the participant click the back button to return to the page in question. A separate category was added, “should have followed-up and didn’t”, for situations in which the experimenter needed to follow-up or remind a participant but could not because of the confusion resulting from the asynchronous communication. As far as can be determined, this had very little impact on the results, but it made it much more difficult to run sessions and interpret results. These problems can be avoided if another chat program is used. For example, in the chat program, MSN Messenger, the recipient is aware that the other person is typing because of a message displayed in their chat window as it is typed. This functionality might have eliminated the problem of participants continuing to type or moving on to a new page. Participants would have been aware that the experimenter was trying to communicate with them long before the experimenter had finished typing. Future studies using typing as an input method should consider such a program.

Coding Comments

One advantage of the use of typing input was the ease in coding responses. The AVI screen sharing video files of the usability tests were played back in Microsoft Windows Media Player. In the typing conditions, it was possible to quickly scroll through the videos, watching participant behaviours and stopping only at participant and experimenter (in the attended condition) comments. This allowed for faster coding and more efficient use of time, because it was not necessary to watch all of the videos in real time. In the voice conditions the experimenter had to watch all of the videos in real time, because she had to wait and listen for the comments to occur while watching the participants perform their tasks. Listening and watching in real time took longer than the scrolling, which could be done in the typing conditions. The other advantage of the typing conditions over the voice conditions is that the experimenter can simply hit the pause key to view the comment instead of having to repeatedly rewind and replay the file to hear a missed comment. Faster coding is an advantage, but ease of coding should not be chosen if the typing condition sacrifices data quality.

Practical Problems with Remote Testing

There are differences in a laboratory setting from actual remote usability testing. The experimenter set up the usability test for herself and the remote participants. In actual remote usability tests, the installation of software and computer configuration would be the responsibility of participants. It is likely that participants would need specific instructions and extra help setting up remote usability tests.

In this study, the remote participant was not actually remote. Participants all had an opportunity to meet the experimenter in person before they began the test. The

“remote” aspect was simulated by having the participant in one room and the experimenter in another. Results might be different if the participants are actually at a great distance from the experimenter. The participant may feel more isolated and react differently to the remote test. It is possible that responses, especially typed responses, might be much briefer without contact with an experimenter. In addition, voice and typing latencies might be longer, and communications perhaps less reliable over the Internet.

Companies that choose remote usability testing as a cheaper and faster way to collect data from a larger audience should give careful thought to the real savings. The tasks and questions in the remote usability tests take as long to create as in the local condition, perhaps longer. The set-up time for remote testing is comparable to local testing and may in fact take longer if remote participants are unable to set up their own equipment effectively. Also attended/typing remote tests take significantly longer to complete.

Generalizability

The generalizability of these results should be considered. In this study, the emphasis was on the comments provided by the participants and not on their performance. This emphasis on comments may not be typical in standard usability tests. Also, there are potentially more comments in this study because participants were asked many questions, especially regarding their understanding of the interface.

Conclusions

It is not appropriate to make a definitive statement about what remote testing method is best considering there were advantages and disadvantages of each method. Table 5 summarizes the significant results when each of the experimental conditions was compared to the local control conditions. The measures are listed in the far left column and the experimental conditions are listed across the top of the table. The differences found are listed in the table. For example, the first entry in the table shows that participants in the unattended/typing condition provided fewer comments than participants in the local condition. The dashes indicate no significant difference between the experimental condition and the local condition.

Table 5

Comparisons of Remote Conditions to the Local Condition

	Unattended/Typing	Unattended/Voice	Attended/Typing	Attended/Voice
Number of Comments	fewer than local	---	---	---
Direct Answers to Questions	fewer than local	fewer than local	---	---
Supplementary Comments	---	---	---	---
Relevant Comments	fewer than local	---	---	---
Irrelevant Comments	fewer than local	---	---	---
No Follow-up Needed	fewer than local	---	---	---
Follow-up Needed	fewer than local	fewer than local	---	---
Experimenter Communications	---	---	fewer than local	---
Understanding of Interface	---	---	---	---
Ease of Use Rating	---	---	---	---
Total Test Time	---	---	longer than local	---

The summaries in Table 5 indicate a greater number of disadvantages for the unattended/typing condition compared to the local usability tests, and compared to the number of disadvantages for the other three experimental conditions. The goal of remote usability testing is to have at least the same level of data quantity and quality as the local condition; otherwise, it is not a viable option. Therefore, unattended/typing is not a recommended approach for remote usability testing.

The unattended/voice and attended/typing condition would not be considered unacceptable because they only differed from the local condition on two measures each, however, they did not fare as well as the attended/voice condition. For example, typing when someone is interacting with the participant makes the test takes longer. Based on the results of this study, if remote usability testing is to be used, the recommended method is attended/voice. In this study the results suggest that the attended conditions were superior to the unattended conditions and this was also the case for the voice condition over the typing condition. It is important to note that the attended/voice condition did not differ significantly from the local condition on any of the measures examined. This finding is the basis for the recommendation to use an attended/voice condition in remote usability testing.

Another finding to note is that none of the experimental conditions differed significantly from the local control condition on usability as measured by the understanding of interface concepts and ease of use ratings. This is an important issue to consider when deciding whether or not to use remote usability testing. It would not be in the best interest of a usability practitioner to use a method which affected the results of his/her usability test, and therefore, the conclusions about the product's ease of use.

Although, the recommended remote usability testing method is attended/voice, this is probably the most difficult and least practical method. The condition that is not recommended, unattended/typing, is probably the method that would be employed most often due to reduced cost and ease of conducting these tests.

Finally, if representative users are too far away for a local usability test, then using a remote usability test seems to be a viable option because organizations can still collect behavioural and subjective data. However, they should consider presence and input type when choosing their remote testing method. In this case, remote usability testing can be a useful tool for an organization.

Future Research

The objective of this study was to learn whether, and under what conditions, remote usability testing could yield as many comments as conventional usability tests. The data show that remote testing is a viable option for evaluating usability. Because the focus was on comparing the volume of information participants contributed, no content analysis was performed on the data. Since “more words”, whether spoken or written, does not necessarily imply “better value”, it would be interesting also to compare the relative value of participants’ comments for informing the further development of the application under scrutiny. A future study should therefore measure the extent to which participants in the different conditions were addressing the same interface issues, how many comments were unique, how many comments identified issues that the performance measures did not. Qualitative content analyses could also have provided information about areas of the interface that needed to be redesigned, which may not necessarily have been identified by other means.

References

- 9AM Solutions (2002). *Methodology*. [Brochure].
- Bartek, V., & Cheatham, D. (2003, January 1). Examine study results on the benefits and downside of remote usability testing. Retrieved from <http://www-106.ibm.com/developerworks/web/library/wa-rmustrs1/>.
- Chapanis, A. (1991). Evaluating usability. In B., Shackel & S. Richardson (Eds), *Human factors for informatics usability* (pp. 359-422). Cambridge: Cambridge University Press.
- Ebling, M.R., & John, B.E. (2000). On the contributions of different empirical data in usability testing. *Symposium on Designing Interactive Systems*, 289-296.
- Ericsson, K.A., & Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87 (3), 215-251.
- Ericsson, K.A., & Simon, H.A. (1984). *Verbal reports as data*. London: MIT Press.
- Hammontree, M., Weiler, P., & Nayak, N. (1994). Remote usability testing. *Interactions*, July, 21-25.
- Hartson, H., Castillo, J.C., & Kelso, J. (1996). Remote evaluation: The network as an extension of the usability laboratory. *In proceedings of CHI '96 conference on human factors in computing systems*, 228-235.
- Jeffries, R., & Desurvire, H. (1992). Usability testing vs. heuristic evaluation: Was there a contest? *SIGCHI Bulletin*, 24 (4), 39-41.
- Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.M. (1991). User interface evaluation in the real world: A comparison of four techniques. *In proceedings of CHI '91 conference on human factors in computing systems*, 119-124.
- Kantner, L., & Rosenbaum, S. (1997). Usability studies of WWW sites: Heuristic evaluation vs. laboratory testing. *SIGDOC 97 Proceedings*, 1-8.
- Lindgaard, G. (1994). *Usability Testing and System Evaluation*. London: Chapman & Hall.

- Macko, J. S. T. (1998). Remote Evaluation: A comparison of attended and unattended usability testing via the Internet. Unpublished Master's thesis, Carleton University, Ottawa, Ontario, Canada.
- Maguire, M. (1997). RESPECT user requirement framework handbook (version 2.2). HUSAT Research Institute.
- Nickerson, R.S., & Landauer, T.K. (1997). Human-computer interaction: Background and issues. In Helander, M.G., Landauer, T.K., & Prabhu, P.V. (Eds.), *Handbook of Human-Computer Interaction* (pp. 3-31). Amsterdam: Elsevier.
- Nielsen, J. (1993). *Usability Engineering*. San Diego: Morgan Kaufmann.
- Nielsen, J., & Landauer, T.K. (1993). A mathematical model of the finding of usability problems. *In proceedings of ACM INTERCHI '93 conference*, 206-213.
- Patrick, A. (2002, November 29). Privacy Incorporated Software Agent: Building a privacy guardian for the electronic age. Retrieved from <http://www.pet-pisa.nl>.
- Preece, J. (1993). *A Guide to Usability: Human Factors in Computing*. Wokingham: Addison-Wesley.
- Redmond-Pyle, D., & Moore, A., (1995). *Graphical user interface design and evaluation: A practical process*. Great Britain: Pearson Education Limited.
- Rubin, J. (1994). *Handbook of Usability Testing*. New York: John Wiley & Sons, Inc.
- Scholtz, J. (2001). Adaptation of traditional usability testing methods for remote testing. *34th Annual Hawaii International Conference on System Sciences*, 5, 5030-5042.
- Shapiro, M. A. (1994). Think-aloud and thought-list procedures in investigating mental processes. In A. Lang (Ed.), *Measuring psychological responses to media messages* (pp. 1-15). Hillsdale, NJ: Erlbaum.
- Thomas, C., & Beven, N. (1996). Usability context analysis: A practical guide (version 4.02). Teddington, Middlesex: National Physical Laboratory.
- Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., & Bergel, M. (2002). An empirical comparison of lab and remote usability testing of web sites. *Usability Professionals Association Conference, July 2002*.

Van Waes, L. (2000). Thinking aloud as a method for testing the usability of websites: The influence of task variation on the evaluation of hypertext. *IEEE Transactions on Professional Communication*, 43 (3), 279-291.

Appendix A - Usability Test Screen Shots

Note: Task instructions and questions apply to the corresponding screen shot.

MyAssistant Demonstration - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media

Question:
Comment on the graphics and colours used on this page.

Activity:
Begin to use MyAssistant to search for a job.

MyAssistant
Welcome

New to MyAssistant?

Register Now!

[Help](#)

Registered Users

User name: [Help](#)

Password:

(secure access)

GlobalSign
TRUST ON THE NET

PISA
Privacy Incorporated Software Agent

[Terms of Use](#) [Privacy Policy](#) [Contact Us](#)

COLLEGE BESCHERMING PERSOONSGEGEVENS

Question: Comment on the graphics and colours used on this page.

Activity: Begin to use MyAssistant to search for a job.

MyAssistant: Register - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media

Question:
What do you think of the fonts used on this page?

MyAssistant
Register

Please provide the required registration information

Please choose a user name and a password. Your password should be at least 8 characters long, easy for you to remember, but hard for others to guess.

User name: [Help](#)

Password:

(secure access)

Question: What do you think of the fonts used on this page?

Registration successful - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media

Question:
What does
this message
mean?

----->

MyAssistant

Thank you for registering

A new personal assistant has been created by MyAssistant.com, and assigned to jane doe.
It will be used for the purposes of:

Personal Assistance

Login to MyAssistant:

User name: [Help](#)

Password:

(secure access)

Question: What does this message mean?

Question: Explain the purpose of each of these menu options, starting with "Create Task".

Activity: Use MyAssistant to search for a job.

Create a Task - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

MyAssistant

Create a Task

[Home](#) **Create Task** [Modify Task](#) [Track Task](#) [Task Results](#) [Help](#) [Logout](#)

Step 1 (of 6): select a task type, and name your task

Task	☒ Search for a job ☐ Search for music (not available yet) ☐ Search for a car (not available yet) ☐ Search for software (not available yet) ☐ Search for news (not available yet)
Name Your Task	<input type="text"/>

Question:
Why do you
need to name
your task?

Question: Why do you need to name your task?

Search Instructions - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media

NOTE:
Use fictional information to complete these forms.

Question:
What are "Job Sectors" and "Company Types"?

Question:
What is this section for? ----->

Question:
What information are you protecting here?

MyAssistant

Search Instructions

Please enter your search instructions for the task: HCI Job.

Step 2 (of 6): Provide your task with 'Search Instructions'

Job Sector:	<input type="text"/>
Company Type:	<input type="text"/>
Salary Expectations:	<input type="text"/>
Desired Locations:	<input type="text"/>
Current Employer:	<input type="text"/>

Set Your Privacy Preferences

This section allows you to set privacy preferences for the information you have entered. Use the check boxes to set your preferences, or use one of the convenient preset options.

Preference Category	Search Instructions	Privacy Presets:
Intended Purposes		☒ No Presets
• Employment Services	☒	☐ Very Private
Other Allowed Purposes		☐ Moderate
• Fundraising	☐	☐ Very Public
• Advertising sent by us	☐	
• Advertising sent by others	☐	
		Help:

Note: Use fictional information to complete these forms.

Question: What are "Job Sectors" and "Company Types"?

Question: What is this section for? ----->

Question: What information are you protecting here?

Question: Do the yellow boxes that appear help you to understand the preference settings?

Contact Information - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

MyAssistant Contact Information

Step 3 (of 6): Please provide your contact information for this job search.

The following fields represent your 'contact details'. Please enter the information that is required.

E-Mail:

Name:

Address:

City:

Province:

Postal Code:

Phone:

Country:

Gender: ☐ Male ☐ Female

Birth Date: (ex. 1980)

Set Your Privacy Preferences

This section allows you to set privacy preferences for the information you have entered. Use the check boxes to set your preferences, or use one of the convenient preset options.

Preference Category	Contact Information	Privacy Presets:
Intended Purposes		☒ No Presets
Development/Support	☐	

Question:
What
information
are you

Question: What information are you protecting here?

Question: What is the meaning of the labels "Intended Purposes" and "Other Allowed Purposes"?

Question: What is the meaning of the labels "Retention Period" and "Required Tracking"?

Question: What is the meaning of the labels "Required Data Controls" and "Allow International Transfer"?

Create a Resume - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media

Reminder:
Use fictional information to complete these forms.

Question:
Was it clear what you were required to fill-in in this section?
----->

Note:
Pretend you are part of a Union ("ECU Local 349").

MyAssistant

Create a Résumé

Step 4 (of 6): Provide your Résumé details

Education:	*Be sure to include your highest level of education completed
Employment History:	*Indicate all relevant past work experiences, with dates, and the positions you have held
Skills:	*List all relevant skills you possess
Additional Information:	*You may enter any additional information here, or use this space to write a brief summary of yourself to present to potential employers
Trade Union Membership:	none

Set Your Privacy Preferences
This section allows you to set privacy preferences for the information you have entered. Use the check boxes to set your preferences, or use one of the convenient preset options.

Preference Category	Résumé	Privacy Presets:
Intended Purposes		☐ No Presets

Reminder: Use fictional information to complete these forms.

Question: Was it clear what you were required to fill-in in this section? ----->

Note: Pretend you are part of a Union ("ECU Local 349").

Question: What is the purpose of the blue box that appeared?

Question: What information are you protecting here?

Confirm Instructions - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search Favorites Media

MyAssistant

Confirm Instructions

Step 5 (of 6): Confirm you task's instructions

Task Label:	HCI Job
Task:	Job Seek Task

Search Instructions

Job Sector:
Company Type:
Salary Expectations:
Desired Locations:
Current Employer:

Contact Details

Your Address

Your City, Your State/Province, Your Postal Code

Phone: Your Phone Number E-mail: Your e-mail address

Résumé:

Education:
*Be sure to include your highest level of education completed

Employment History:

Question: What is the purpose of the small blue box that appeared? What did you agree to?

Question: Please comment on the ease of creating a search task and rate it using a scale of 1 to 7. 1 = Very Difficult to 4 = Neither Ease or Difficult to 7 = Very Easy.

Activity: This part of the experiment is now complete.

Appendix B – Screen Shots of Desktop Arrangements

Screen Shot of Desktop Arrangement of the Experimenter

Screen Shot of Desktop Arrangement of Participant in Voice Condition

Screen Shot of Desktop Arrangement of Participant in Typing Condition

Screen Shot of Desktop Arrangement of Participant in Local Condition

Appendix C - Informed Consent Forms

Informed Consent Non-Students

The purpose of an informed consent is to insure that you understand the purpose of the study and the nature of your involvement. The informed consent must provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Present study: Remote Usability Testing Study.

Research personnel: The following people are involved in this research project and may be contacted at any time: Cassandra Holmes (Principal Investigator, 520-2600, ext. 6628) and Dr. Dick Dillon (Supervisor, 520-2600, ext. 6629). Should you have any ethical or other concerns about this study then please contact Dr. M. Gick, Chair, Carleton University Research Ethics Committee for Psychological Research, 520-2600, ext. 2664) or Dr. K. Matheson (Chair, Dept. of Psychology, 520-2600, ext. 2648).

Purpose: The goal of the study is to examine the effects of various factors on the quality of data collected in remote usability testing.

Task requirements: You will complete tasks on a job search web site and will be asked to provide comments about the ease of use and their satisfaction with the site. In addition, you will be asked questions regarding your experiences using the site.

Duration and locale: The experimental session will last **1 to 1^{1/2} hours** and you will be **paid \$10 per hour** for your time. Testing will take place in Room 210 of the Social Science Research Building, Carleton University.

Potential risk/discomfort: There are no potential physical or psychological risks in this experiment. Your participation in this experiment is completely voluntary; please do not feel obligated to participate.

Anonymity/confidentiality: The data collected in this experiment are confidential. All data are coded such that your name is not associated with the data. In addition, the coded data are made available only to the researchers associated with this project.

Right to withdraw: You have the right to withdraw from the experiment or not answer specific questions without penalty.

Signatures

I have read the above description of the Remote Usability Testing study and understand the conditions of my participation. My signature indicates that I agree to participate in the experiment.

Participant's Name (please print): _____

Participant's Signature: _____ Date: _____

Informed Consent for Students

The purpose of an informed consent is to insure that you understand the purpose of the study and the nature of your involvement. The informed consent must provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Present study: Remote Usability Testing Study.

Research personnel: The following people are involved in this research project and may be contacted at any time: Cassandra Holmes (Principal Investigator, 520-2600, ext. 6628) and Dr. Dick Dillon (Supervisor, 520-2600, ext. 6629). Should you have any ethical or other concerns about this study then please contact Dr. M. Gick, Chair, Carleton University Research Ethics Committee for Psychological Research, 520-2600, ext. 2664) or Dr. K. Matheson (Chair, Dept. of Psychology, 520-2600, ext. 2648).

Purpose: The goal of the study is to examine the effects of various factors on the quality of data collected in remote usability testing.

Task requirements: You will complete tasks on a mock-up of a job search web site and will be asked to provide comments about the ease of use and your satisfaction with the site. In addition, you will be asked questions regarding your experiences using the site.

Duration and locale: The experimental session will last **1 to 1^{1/2} hours**. Testing will take place in Room 210 of the Social Science Research Building, Carleton University.

Potential risk/discomfort: There are no potential physical or psychological risks in this experiment.

Anonymity/confidentiality: The data collected in this experiment are confidential. All data are coded such that your name is not associated with the data. In addition, the coded data are made available only to the researchers associated with this project.

Right to withdraw: You have the right to withdraw from the experiment or not answer specific questions without penalty.

Signatures

I have read the above description of the Remote Usability Testing study and understand the conditions of my participation. My signature indicates that I agree to participate in the experiment.

Participant's Name (please print): _____

Participant's Signature: _____ Date: _____

Appendix D - Participant Instructions

Instructions for Participants in the Attended/Voice Condition

You are going to help us test a new service on the Internet called *MyAssistant*. *MyAssistant* helps individuals search for many different things on the Internet, such as jobs, cars, or music. Today you will use *MyAssistant* to find a new job. *MyAssistant* asks you to enter information so that it can search the Internet for you to find jobs that you might be eligible for. *MyAssistant* then reports back if it finds anything relevant.

You will do several things on the *MyAssistant* web site and answer questions about the job searching service and your experiences using it. The *MyAssistant* service that you are using is a mock-up and not all areas of the site are working.

Please do not move, resize, or close any of the windows that are shown. Please do not change any settings.

We are testing the web site, not you or your abilities. If you have problems or do not like something, we will use that information to improve the *MyAssistant* site.

Input

During this test you will provide your thoughts, comments, and answers to questions by speaking into a microphone (attached to a headset). The experimenter will be listening to what you say and observing what you type and may ask you questions for clarification or to expand upon what you have said. The microphone should be placed approximately 3-4 cm in front of your mouth and the headset comfortably on your ears. Your voice and the experimenter's voice will be recorded throughout the test and what you type will be observed. You are not required to start the recording; it will begin automatically as soon as you start the test.

Activities and Questions

Throughout the test there will be activities for you to do and questions for you to answer. The instructions for the activities and the questions will appear on the left side of the screen and look like this:

Activity: This is an activity instruction.	Question: This is a request for comments.
---	---

Please answer the questions (which appear in blue) *before* you continue with the activities. You will answer the questions by speaking into your microphone. You will then continue with the activity explained on the left side of the screen.

Thoughts and Comments

Please speak out your thoughts as much as possible while you are doing things on the web site. We learn a lot from what you are thinking. This information is being recorded to help improve the web site. For example, if you find something difficult to access or to read, make comments that indicate this. Be critical but fair. Be open and honest about how you feel about any aspect of the web site. It is important that you make comments about what you are doing and how you feel about the site throughout the whole test.

Here is an example of what we are looking for in terms of you expressing your thoughts and comments. In this example, a woman named Martha was asked to **print** a document in Microsoft Word with **draft** quality. Be sure to read the entire example.

Martha said:

"OK, I have to print this document. I think the printer is under Tools. I will look under tools.....oh, I can't find it under Tools. I guess I was wrong. OK, I'll try something else. Maybe I'll start at the first item in the toolbar at the top of the screen and go through them all. OK, the first one is File, I'll look in there. Oh, there it is! Great! I'll click on it."

After Martha clicked on "print", a box popped up, and she said:

"I don't see the print quality though. Why is that not on this screen with the rest of the information? Maybe it is here under Options. Nope...not there. OK, I only see one other button called Properties. I will try it. Ah...here it is. I will click on draft. I think all I do now is click "OK" and it will print. Oh, I guess not. I have to click "OK" one more time. There...done."

It is important that you speak your thoughts and provide comments throughout. Once you understand these instructions, [click here](#) to begin.

Instructions for Participants in the Attended/Typing Condition

You are going to help us test a new service on the Internet called *MyAssistant*. *MyAssistant* helps individuals search for many different things on the Internet, such as jobs, cars, or music. Today you will use *MyAssistant* to find a new job. *MyAssistant* asks you to enter information so that it can search the Internet for you to find jobs that you might be eligible for. *MyAssistant* then reports back if it finds anything relevant.

You will do several things on the *MyAssistant* web site and answer questions about the job searching service and your experiences using it. The *MyAssistant* service that you are using is a mock-up and not all areas of the site are working.

Please do not move, resize, or close any of the windows that are shown. Please do not change any settings.

We are testing the web site, not you or your abilities. If you have problems or do not like something, we will use that information to improve the *MyAssistant* site.

Input

During this test you will provide your thoughts, comments, and answers to questions by typing into NetMeeting's chat program which is open in the window on the far right of the screen. You will type your responses in the box labelled "Message:" and hit the enter key to send the response. The experimenter is the person you will be "chatting" with. The experimenter will be watching what you type and may ask you questions for clarification or to expand upon what you have said (by typing back to you). What you type will be observed and recorded throughout the test.

Activities and Questions

Throughout the test there will be activities for you to do and questions for you to answer. The instructions for the activities and the questions will appear on the left side of the screen and look like this:

Activity: This is an activity instruction.	Question: This is a request for comments.
---	---

Please answer the questions (which appear in blue) *before* you continue with the activities. You will answer the questions by typing into the chat window and hitting enter. You will then continue with the activity explained on the left side of the screen.

Thoughts and Comments

Please type your thoughts as much as possible while you are doing things on the web site. We learn a lot from what you are thinking. This information is being recorded to help improve the web site. For example, if you find something difficult to access or to read, make comments that indicate this. Be critical but fair. Be open and honest about how you

feel about any aspect of the web site. It is important that you make comments about what you are doing and how you feel about the site throughout the whole test.

Here is an example of what we are looking for in terms of you expressing your thoughts and comments. In this example, a woman named Martha was asked to **print** a document in Microsoft Word with **draft** quality. Be sure to read the entire example.

Martha typed:

"- i think the printer is under tools [enter]
- i can't find it under tools [enter]
- i'll try the first item in the toolbar at the top of the screen and go through them all [enter]
- the first 1 is file...i'll look there [enter]
- there it is...great...i'll click on it [enter]"

After Martha clicked on "print", a box popped up, and she typed:

"- i don't see print quality[enter]
- maybe it's under options [enter]
- nope [enter]
- i only see 1 other button called properties....i will try it [enter]
- here it is...i will click on draft [enter]
- i think all i do now is click ok and it will print...i guess not [enter]
- i have to click ok one more time...done [enter]"

Do not worry about spelling, punctuation, grammar, or sentence structure. Feel free to use bullets and short forms. Just be sure that the experimenter will understand what you are trying to say.

It is important that you type your thoughts and provide comments throughout.

Once you understand these instructions, [click here](#) to begin.

Instructions for Participants in the Unattended/Voice Condition

You are going to help us test a new service on the Internet called *MyAssistant*. *MyAssistant* helps individuals search for many different things on the Internet, such as jobs, cars, or music. Today you will use *MyAssistant* to find a new job. *MyAssistant* asks you to enter information so that it can search the Internet for you to find jobs that you might be eligible for. *MyAssistant* then reports back if it finds anything relevant.

You will do several things on the *MyAssistant* web site and answer questions about the job searching service and your experiences using it. The *MyAssistant* service that you are using is a mock-up and not all areas of the site are working.

Please do not move, resize, or close any of the windows that are shown. Please do not change any settings.

We are testing the web site, not you or your abilities. If you have problems or do not like something, we will use that information to improve the *MyAssistant* site.

Input

During this test you will provide your thoughts, comments, and answers to questions by speaking into a microphone (attached to a headset). It may seem strange to talk to yourself while doing activities, but it is important that you do so throughout the test. The microphone should be placed approximately 3-4 cm in front of your mouth and the headset comfortably on your ears. Your voice will be recorded throughout the test and what you type will be observed. You are not required to start the recording; it will begin automatically as soon as you start the test.

Activities and Questions

Throughout the test there will be activities for you to do and questions for you to answer. The instructions for the activities and the questions will appear on the left side of the screen and look like this:

Activity: This is an activity instruction.	Question: This is a request for comments.
---	---

Please answer the questions (which appear in blue) *before* you continue with the activities. You will answer the questions by speaking into your microphone. You will then continue with the activity explained on the left side of the screen.

Thoughts and Comments

Please speak out your thoughts as much as possible while you are doing things on the web site. We learn a lot from what you are thinking. This information is being recorded to help improve the web site. For example, if you find something difficult to access or to read, make comments that indicate this. Be critical but fair. Be open and honest about how you feel about any aspect of the web site. It is important that you make comments about what you are doing and how you feel about the site throughout the whole test.

Here is an example of what we are looking for in terms of you expressing your thoughts and comments. In this example, a woman named Martha was asked to **print** a document in Microsoft Word with **draft** quality. Be sure to read the entire example.

Martha said:

"OK, I have to print this document. I think the printer is under Tools. I will look under tools.....oh, I can't find it under Tools. I guess I was wrong. OK, I'll try something else. Maybe I'll start at the first item in the toolbar at the top of the screen and go through them all. OK, the first one is File, I'll look in there. Oh, there it is! Great! I'll click on it."

After Martha clicked on "print", a box popped up, and she said:

"I don't see the print quality though. Why is that not on this screen with the rest of the information? Maybe it is here under Options. Nope...not there. OK, I only see one other button called Properties. I will try it. Ah...here it is. I will click on draft. I think all I do now is click "OK" and it will print. Oh, I guess not. I have to click "OK" one more time. There...done."

It is important that you speak your thoughts and provide comments throughout.

Once you understand these instructions, [click here](#) to begin.

Instructions for Participants in Unattended/Typing Condition

You are going to help us test a new service on the Internet called *MyAssistant*. *MyAssistant* helps individuals search for many different things on the Internet, such as jobs, cars, or music. Today you will use *MyAssistant* to find a new job. *MyAssistant* asks you to enter information so that it can search the Internet for you to find jobs that you might be eligible for. *MyAssistant* then reports back if it finds anything relevant.

You will do several things on the *MyAssistant* web site and answer questions about the job searching service and your experiences using it. The *MyAssistant* service that you are using is a mock-up and not all areas of the site are working.

Please do not move, resize, or close any of the windows that are shown. Please do not change any settings.

We are testing the web site, not you or your abilities. If you have problems or do not like something, we will use that information to improve the *MyAssistant* site.

Input

During this test you will provide your thoughts, comments, and answers to questions by typing into NetMeeting's chat program which is open in the window on the far right of the screen. You will type your responses in the box labelled "Message:" and hit the enter key to send the response. Even though "chatting" is usually between two people, in this test no one will respond. Just continue to type your comments and send them without response from anyone. What you type will be observed and recorded throughout the test.

Activities and Questions

Throughout the test there will be activities for you to do and questions for you to answer. The instructions for the activities and the questions will appear on the left side of the screen and look like this:

Activity: This is an activity instruction.	Question: This is a request for comments.
---	---

Please answer the questions (which appear in blue) *before* you continue with the activities. You will answer the questions by typing into the chat window and hitting enter. You will then continue with the activity explained on the left side of the screen.

Thoughts and Comments

Please type your thoughts as much as possible while you are doing things on the web site. We learn a lot from what you are thinking. This information is being recorded to help improve the web site. For example, if you find something difficult to access or to read, make comments that indicate this. Be critical but fair. Be open and honest about how you feel about any aspect of the web site. It is important that you make comments about what you are doing and how you feel about the site throughout the whole test.

Here is an example of what we are looking for in terms of you expressing your thoughts and comments. In this example, a woman named Martha was asked to **print** a document in Microsoft Word with **draft** quality. Be sure to read the entire example.

Martha typed:

"- i think the printer is under tools [enter]
 - i can't find it under tools [enter]
 - i'll try the first item in the toolbar at the top of the screen and go through them all [enter]
 - the first 1 is file...i'll look there [enter]
 - there it is...great...i'll click on it [enter]"

After Martha clicked on "print", a box popped up, and she typed:

"- i don't see print quality[enter]
- maybe it's under options [enter]
- nope [enter]
- i only see 1 other button called properties....i will try it [enter]
- here it is...i will click on draft [enter]
- i think all i do now is click ok and it will print...i guess not [enter]
- i have to click ok one more time...done [enter]"

Do not worry about spelling, punctuation, grammar, or sentence structure. Feel free to use bullets and short forms. Just be sure that the experimenter will understand what you are trying to say.

It is important that you type your thoughts and provide comments throughout.

Once you understand these instructions, [click here](#) to begin.

Instructions for Participants in the Local Condition

You are going to help us test a new service on the Internet called *MyAssistant*. *MyAssistant* helps individuals search for many different things on the Internet, such as jobs, cars, or music. Today you will use *MyAssistant* to find a new job. *MyAssistant* asks you to enter information so that it can search the Internet for you to find jobs that you might be eligible for. *MyAssistant* then reports back if it finds anything relevant.

You will do several tasks on the *MyAssistant* web site and answer questions about the job searching service and your experiences using it. The *MyAssistant* service that you are using is a mock-up and not all areas of the site are working.

Please do not move, resize, or close any of the windows that are shown. Please do not change any settings.

We are testing the web site, not you or your abilities. If you have problems or do not like something, we will use that information to improve the *MyAssistant* site.

Input

During this test you will provide your thoughts, comments, and answers to questions by speaking aloud to the experimenter sitting beside you like you normally would in a conversation. A microphone on the desk will record your voice and the experimenter's

voice. The experimenter will also take notes. The experimenter will be listening to what you say and may ask you questions for clarification or to expand upon what you have said. You are not required to start the recording; the experimenter will start the recording before you enter the room.

Activities and Questions

Throughout the test there will be activities for you to do and questions for you to answer.

Thoughts and Comments

Please speak out your thoughts as much as possible while you are doing things on the web site. We learn a lot from what you are thinking. This information is being recorded to help improve the web site. For example, if you find something difficult to access or to read, make comments that indicate this. Be critical but fair. Be open and honest about how you feel about any aspect of the web site. It is important that you make comments about what you are doing and how you feel about the site throughout the whole test.

Here is an example of what we are looking for in terms of you expressing your thoughts and comments. In this example, a woman named Martha was asked to **print** a document in Microsoft Word with **draft** quality.

Martha said:

"OK, I have to print this document. I think the printer is under Tools. I will look under tools.....oh, I can't find it under Tools. I guess I was wrong. OK, I'll try something else. Maybe I'll start at the first item in the toolbar at the top of the screen and go through them all. OK, the first one is File, I'll look in there. Oh, there it is! Great! I'll click on it."

After Martha clicked on "print", a box popped up, and she said:

"I don't see the print quality though. Why is that not on this screen with the rest of the information? Maybe it is here under Options. Nope....not there. OK, I only see one other button called Properties. I will try it. Ah...here it is. I will click on draft. I think all I do now is click "OK" and it will print. Oh, I guess not. I have to click "OK" one more time. There...done."

It is important that you speak your thoughts and provide comments throughout.

Once you understand these instructions, begin the test.

Appendix E - Trust Questionnaire

Please answer the following questions:

1. I am comfortable sending personal information on the Internet.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

2. I feel that most Internet services would act in customers' best interest.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

3. I am comfortable relying on Internet vendors to meet their obligations.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

4. The Internet has enough safeguards to make me feel comfortable using it to transact personal business.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

5. I believe that MyAssistant would act in my best interest.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

6. MyAssistant is truthful in its dealings with me.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

7. I would characterize MyAssistant as trustworthy.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

8. When I search for jobs, I would feel comfortable depending on the privacy provided by MyAssistant.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

9. If I were looking for a job, I would want to use MyAssistant again.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

10. I would be willing to provide information like my name, address and phone number to MyAssistant.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

11. I would be willing to provide my resume to MyAssistant.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

12. I would be willing to share my salary expectation with MyAssistant.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

13. Among my peers, I am usually the first to try out new Internet sites or services.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

14. Overall, this site worked very well technically.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

15. Visually, this site resembled other sites I think highly of.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

16. This site was simple to navigate.

1	2	3	4	5
Strongly disagree	Disagree	Undecided	Agree	Strongly Agree

Appendix F - Written Debriefing

Debriefing

The study will determine if remote testing is a good alternative to traditional local usability testing by comparing comments given by participants. Sometimes the most common way of usability testing, where the experimenter and the participant are in the same room together, is not possible. It is possible that the type of person that uses the product being tested lives far away from the testing lab. It is expensive to have the participants or the experimenters travel. Remote usability testing may be the solution. If we determine that remote tests produce the same quality of results as the current method of usability testing (participant and experimenter physically together) then there may be no reason that these types of tests should not be used.

The purpose of this experiment is to see if there is a difference in the quality of data obtained if the participant has contact with the experimenter or not. Also, it is important to examine if there is any difference in usefulness of the comments between different methods of input (typing or voice). You participated in one of 5 conditions. There are two conditions in which the participant interacts with the experimenter by either typing to them in a chat program or speaking to them over the Internet using a headset and microphone. In the other two conditions the same procedure is followed but the experimenter does not respond. The final condition is the control condition in which the experimenter and the participant are in the same room together and the participant can speak directly to the experimenter who is sitting beside them.

If you have any questions or concerns please contact Cassandra Holmes (Principal Investigator, 520-2600, ext. 6628) and Dr. Dick Dillon (Supervisor, 520-2600, ext. 6629). Should you have any ethical or other concerns about this study then please contact Dr. M. Gick, Chair, Carleton University Research Ethics Committee for Psychological Research, 520-2600, ext. 2664) or Dr. K. Matheson (Chair, Dept. of Psychology, 520-2600, ext. 2648).

Appendix G – Tables of Means for Dependent Variables
and Experimental and Local Conditions

	Attended	Unattended	Voice	Typing	Local
Total Number of Comments	41.30	31.75	42.65	30.40	44.40
Direct Answers to Questions	25.10	22.05	23.85	23.30	26.40
Supplementary Comments	8.65	9.70	13.20	5.15	8.30
Relevant Comments	36.75	29.70	37.65	28.80	39.30
Irrelevant Comments	4.55	2.05	5.00	1.60	5.10
No Need to Follow-up	31.95	22.65	33.65	25.95	34.30
Follow-up Needed	9.70	4.10	9.15	4.65	10.20
Understand of Interface	18.90	15.55	16.30	18.15	18.60
Ease of Use Rating	4.63	4.68	4.43	4.88	4.90
Total Test Time	37.00	28.17	29.66	35.51	23.21

	A/T	A/V	U/T	U/V	Local	Overall
Total Number of Comments	34.70	47.90	26.10	37.40	44.40	38.10 (SD = 1.79)
Direct Answers to Questions	24.70	25.50	21.90	22.20	26.40	24.14 (SD = 3.11)
Supplementary Comments	6.10	11.20	4.20	15.20	8.30	9.00 (SD = 7.99)
Relevant Comments	32.30	41.20	25.30	34.10	39.30	34.44 (SD = 9.13)
Irrelevant Comments	2.40	6.70	0.80	3.30	5.10	3.66 (SD = 3.38)
No Need to Follow-up	30.00	33.90	21.90	33.40	34.30	30.70 (SD = 9.36)
Follow-up Needed	5.10	14.30	4.20	4.00	10.20	7.56 (SD = 5.20)
Experimenter Communications	19.50	30.00	N/A	N/A	31.50	16.20 (SD = 5.38)
Understand of Interface	20.30	17.50	16.00	15.10	18.60	17.50 (SD = 3.89)
Ease of Use Rating	4.56	4.70	5.20	4.15	4.90	4.70 (SD = 1.38)
Total Test Time	42.27	31.73	28.74	27.59	23.21	30:42 (SD = 2:30)