ABSTRACT: The abstract is a brief, comprehensive summary of the content of the proposal in about 150 words in plain language. Reviewers receive their first impression from this abstract. The information needs to be concise, well organized, self contained, and understandable to persons outside your academic discipline.

Over the last twenty years, computers have become a significant factor in nearly all that we do. With this rise in popularity, the need for easily usable software has increased as well. The study of Human-Computer Interaction (HCI) is devoted to understanding and exploring methods of creating usable software. One of the most common methods of testing software usability involves usability labs. A usability lab is a room or rooms in which a potential user can test software while under observation of the software development team. While there are common types of usability labs, all implementing various methods of observing the user and measuring usability, there is no set standard for how this is done. Some research has shown that certain methods used could have a psychological effect on the user. In our study, we will be looking at how different usability lab designs affect user anxiety levels and whether that impacts the number of mistakes made during a usability test. The results of this study could have significant impact on how usability labs are designed and the methods of testing that are used.

Upon submitting this proposal, I verify that this writing is my own and pledge to fulfill all of the expectations of the Undergraduate Research Academy to the best of my abilities. I understand that failure to do so may result in return of fellowship money to the University and forfeiture of academic credit and honors recognition.

Signature of the Student

I am able, willing, and committed to providing the necessary facilities and to take the time to mentor this student during this project. I verify that this student is capable of undertaking this proposed project.

Signature of the Faculty Mentor
This project is within the mission and scope of this department, and the department fully supports the faculty mentor and student during this venture.

Signature of the Department Chairperson

I testify that all necessary research protocols (human, animal, toxic waste) have been fulfilled, and I support this proposed faculty-student scholarly activity as within the mission of the College/School.

Signature of the Dean of the College/School
PROJECT TITLE: The Effects of the Testing Environment on User Performance in Software Usability Testing

ABSTRACT: The abstract is a brief, comprehensive summary of the content of the proposal in about 150 words in plain language. Reviewers receive their first impression from this abstract. The information needs to be concise, well organized, self-contained, and understandable to persons outside your academic discipline.

Over the last twenty years, computers have become a significant factor in nearly all that we do. With this rise in popularity, the need for easily usable software has increased as well. The study of Human-Computer Interaction (HCI) is devoted to understanding and exploring methods of creating usable software. One of the most common methods of testing software usability involves usability labs. A usability lab is a room or rooms in which a potential user can test software while under observation of the software development team. While there are common types of usability labs, all implementing various methods of observing the user and measuring usability, there is no set standard for how this is done. Some research has shown that certain methods used could have a psychological effect on the user. In our study, we will be looking at how different usability lab designs affect user anxiety levels and whether that impacts the number of mistakes made during a usability test. The results of this study could have significant impact on how usability labs are designed and the methods of testing that are used.

Upon submitting this proposal, I verify that this writing is my own and pledge to fulfill all of the expectations of the Undergraduate Research Academy to the best of my abilities. I understand that failure to do so may result in return of fellowship money to the University and forfeiture of academic credit and honors recognition.

Signature of the Student

I am able, willing, and committed to providing the necessary facilities and to take the time to mentor this student during this project. I verify that this student is capable of undertaking this proposed project.

Signature of the Faculty Mentor
This project is within the mission and scope of this department, and the department fully supports the faculty mentor and student during this venture.

Signature of the Department Chairperson

I testify that all necessary research protocols (human, animal, toxic waste) have been fulfilled, and I support this proposed faculty-student scholarly activity as within the mission of the College/School.

Signature of the Dean of the College/School

3-16-04
I. Introduction

With the arrival of desktop computers in the early 1980s and the decreasing cost of computer hardware over the last two decades, computers have been employed in nearly every aspect of our lives. They are being used to both increase work productivity and to provide entertainment. As a result, the primary computer user has shifted from the techno-savvy individual who understands how computers work, to the everyday person who does not know. However, the fallacy is that computer productivity has not risen at the same rate as computer popularity (8).

The main issue to increasing productivity is making software that is usable. Usable means that it fits naturally into a person’s daily work practice. As a result of being usable people maintain focus on their task at hand and not how the software implements their task at hand. The ultimate goal of usability is that there should be little or no learning curve, a user should never have to look at a manual, and the computer and software implementation should be invisible to the user (7). The area of research devoted to studying usability is Human-Computer Interaction (HCI). While HCI has made many advances since the advent of the graphical user interface beginning with Apple’s release of the Macintosh in 1984, methods to designing usable software still remain rather questionable (1).

One of the main reasons that software is often difficult to use is that the software developer has little or no contact with the people who will eventually be using the software. As a result the developer’s focus tends to be entirely on the design of the underlying system, with consideration to only what will make the software easier to implement rather than easier to use (2). It is with this concept that software designers have begun to incorporate the intended user into the design process. This is known as User Centered Design (UCD). UCD keeps the user as the focus of the software design process using a variety of methods including interviews, job shadowing, and usability testing (2).

Usability testing allows the user to experiment with early versions of the software or prototypes to provide designers with input about the design and ease of use. During a usability test, a potential user is given the software to try, usually through a set of tasks to complete. During the test, the user is monitored for performance and often encouraged to “think aloud” in order to give feedback as to how they are using the software. The “think aloud” technique is very common and “often considered to be the usability method” (7, p. 1).

Usability testing is generally conducted in a usability lab, specifically designed to test software. Usability labs are not uncommon in industry and implement/utilize a variety of different layouts. Figures 1, 2, and 3 are all images of usability labs in the St. Louis area. Figure 1 was taken at MasterCard Int., figure 2 at Tripos Inc., and figure 3 at Edward Jones Co. Most usability labs contain at least one kind of monitoring device. A few of the more common examples are cameras, voice recorders, and two way mirrors. As can be seen in figures 1-3, not all labs use all of these monitoring methods. The manner in which these methods are employed also varies. Some labs use small cameras mounted on the desk (figure 2), while others use cameras mounted on tripods around the user (figure 3).
However common the accepted methods of usability testing are, there still exist several questions about the way they are implemented. The first consideration is the effect of Electronic Performance Monitoring (EPM). EPM is the use of electrical monitoring devices such as cameras, voice recorders, or eye movement trackers in capturing or monitoring the performance of a subject (3). EPM is very common in usability testing as there is always the need to capture what the user is doing for review at a later time. According to psychological studies, EPM causes detrimental effects on workplace performance due to increased levels of anxiety (3). An open question remains: how do the psychological effects of EPM impact the evaluation of software design in a usability study?

The use of two-way mirrors is also a questionable user monitoring practice. The purpose of a two-way mirror is to observe the user while they are unable to see their observers. However when they look at the mirror they see themselves, which causes the user to become more self-aware. The self-awareness theory states that when we are forced to become inner focused, we concentrate on our own actions and then compare them to our personal standards (5). If, when we become inner focused, we find that we are not living up to our own standards, we try to change our behavior in order to match our standards. However, when a behavior change is not possible, we are left feeling uncomfortable (5). Such could be the case with usability testing, when users are unable to change their behavior. This could cause users to feel uncomfortable and suffer increased levels of anxiety that may also impact the result of evaluation of the software design.

The self-awareness theory does not only apply to mirrors, it also brings the popular “think aloud” technique into question. When users are told to “think out loud”, it usually implies that they will be explaining their every action and thought (7). However, when forced to examine their own thoughts, they once again become self-aware. Only this time, they are becoming aware of their thoughts instead of their actions. The “think aloud” technique presents more possible problems than just those associated with self-awareness. During a usability test, the users are concentrating on the tasks given to them. However, when told to “think out loud”, users are forced to split their attention
between completing the task at hand and focusing on their own thoughts. The divided attention theory states that when people are forced to split their attention between two or more tasks, results in decreased performance occur for all tasks (6).

If all of this information says that the way we are performing usability testing is not necessarily the best, or in some way makes the user anxious and uncomfortable, then why is industry still using these methods? This study is designed to take the known side effects of usability testing and test them against modern usability lab designs. This information could greatly influence how usability labs are designed and what methods of monitoring are used for usability testing. After all, if software is being designed for the user, then does it not stand to reason that usability labs should be designed with the user in mind as well?

II. Literature Review

Much has been written about the different possible layouts of a usability lab. The Handbook of Usability by Jeffery Rubin is a book written entirely on the subject of usability testing (2). This book gives great attention to how usability labs are designed and even offers several standard types of usability labs outlining the advantages and disadvantages of each, (refer to Figure 4 for the different layouts).

However, little attention is paid to how each layout affects the user’s performance. All of the advantages or disadvantages mentioned center on how each layout allows for ease of data collection, cost of operating such a lab, and ease of use for the tester. The fact that a book as comprehensive to usability testing as The Handbook of Usability fails to accurately cover such an important topic suggests that more research is needed in this area before data on the subject can be published.

Usability testing is not the only subject to have been the focus of research. Significant research has been done in the area of EPM as well. Irving, Higgins, and Safayenima (1986) performed an investigation comparing survey responses of 50 clerical employees whose work was monitored by computer with the responses of 94 non-monitored employees who performed comparable jobs. Self-reported stress was higher among employees who were monitored (3).

Other important research covers the divided attention theory. Research in this area supports the idea that when someone divides his attention on multiple tasks his
performance suffers. Strayer and Johnston (2001) completed a study in which two different groups were told to try to keep a joystick aligned with a moving cursor in order to simulate driving. Visual cues were given throughout the test/experiment, including red and green lights. Subjects were told to complete simple tasks such as starting and stopping when the visual cues were given. One group was assigned to "drive," while the other group was given a cell phone on which to talk while "driving". Users carrying on cell phone conversations were twice as likely to miss a visual cue (7% compared to 3%) as those not on the phone. Response time was also 50 milliseconds longer for the people speaking on the cell phone than those that were not (6).

Finally, the article "Reflections on the Think-aloud Technique" offers insight into the use of think aloud (7). Neilsen, Clemmensen, & Carsten worked with students on think aloud and asked them to reflect on their experience. Their responses have raised a number of issues. Students complain that they think faster than they can speak, that their thought processes are much more complex than they can verbalize and that thinking aloud interferes with their interaction with the interfaces and the task. This supports the idea that although widely accepted and used, it might not be the best alternative. Later in the article, the authors also go on to say that "HCI research may benefit from experimental research [on "think aloud"]" (7, p. 1).

III. Hypothesis
Modern usability labs are designed in a variety of ways, all using different techniques for monitoring and recording the user's actions, performance, thought processes. These different methods all affect the user in one way or another. The goal of this research is to study the various usability lab settings to get a better understanding of what could help minimize the anxiety of the user and thereby reduce the number of mistakes made during a usability test. The research will also explore the possibility that the widely accepted "think aloud" technique causes the users to divide their attention and suffer poor test results when compared to those not using "think aloud." It is expected that camera position and the application of different monitoring techniques greatly influence the anxiety of the user, yielding an ideal set up for usability labs taking a user centered approach. At the same time, the "think aloud" technique will likely be discredited by the number of errors the users make.

IV. Materials and Methods
The study is intended to look at how known psychological side effects of techniques adapted to usability testing effect a user's performance during a usability test. Specifically, the study will examine how the lab is arranged and monitoring devices are used. It is also going to challenge the popular think aloud technique. The participants for the study will be chosen from Applied Computer Concepts (CS108), Aspects of Computer Science (CS111), and Introduction to Computing (CS140). The study will be done using six groups consisting of 25 participants each totaling 150 participants. The necessary clearance for human subject testing is on file and pending approval. Four of the groups will be tested using the SIUE Usability Lab in the Engineering Building, room 3048. The final two groups will be done using the half of the usability lab without cameras.
The usability lab is designed specifically for the purpose of testing software usability so it is ideal for the study. It is a split room with one room configured like an office building while the adjacent room is used for monitoring the office via closed circuit cameras. Groups A and B will be done using the lab in an altered state. Cameras will be mounted on tripods around the user and an audio recording device will be present on the desk in front of the user. Since a two-way mirror is too difficult to install, we will place a mirror near the user so that they will be able to see themselves and thus serve the purpose of making the user self-aware. Group A will be encouraged to think aloud and explain their actions as they go through the study. Group B will not be asked to think aloud. Instead they will be questioned in a post session interview to get their reasons and reactions on how they did. Groups C and D will be done once again using the usability lab only this time in its normal state. There will be no extra monitoring devices and no mirror in this study. Group C will be encouraged to think aloud while group D will be questioned in a post test interview. Groups E and F will be done in the half of the usability lab with no windows, mirrors, or recording devices present. Their audible remarks will be recorded using a small tape recorder that will be concealed from sight. Group E will be encouraged to think aloud while group F will be discouraged and questioned in a post session interview.

All groups will be tested using the same method. Users will be told they are doing a usability audit of the new Microsoft One Note program. The test will first consist of a written 25-question “quiz.” The users will be questioned on what menus they think certain options are under. The second half of the test will consist of ten different tasks. The users will have five minutes to complete each task but will be told that there is no bonus for completing the tasks ahead of schedule and therefore can take their time. However the time it takes to fully complete a task will be collected and analyzed along with the rest of the data. We will be using five minutes as to avoid a ceiling in our collected data. Five of the tasks will be considered “easy” tasks which anyone that has used other Microsoft programs can complete. The other five tasks will be considered difficult and only someone with prior experience using One Note will be able to immediately complete. The data collected from each study will consist of the amount of time it takes to complete a task (if completed) and the number of errors a user makes on a given task. An error will consist of anything diverging from the proper method to complete the task. In order to measure the users’ anxiety levels we will be using the State-trait Anxiety Inventory which gives a direct score that is used to gauge a user’s anxiety level.

Data collection will begin in the September 2004. The groups will not be done in sequential order (A-F), instead they will be split up in order to minimize the chance that the experimenters own practice skews the results. The testing will start with groups A and C and will be done throughout September and October while groups E and D will be tested throughout November and December. I will finish groups B and F in January and February. It is very important that groups A and B are tested at the beginning of the semesters because they require the most alteration to the normal state of the lab. The statistical analysis will consist of three 2x3 Analysis of Variance (ANOVA). There will be one ANOVA for each of the different dependent variables. The three dependent variables are amount of time, number of errors, and anxiety level.
V. References


V. Budget Justification

Commodities

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 – 6 hour VHS tapes for 150 participant tests (Target)</td>
<td>$38.97</td>
</tr>
<tr>
<td>2 boxes number 2 pencils for filling out questionnaires (Target)</td>
<td>$3.89</td>
</tr>
<tr>
<td>Samsung 50 pack DVD-Rs (<a href="http://www.newegg.com">www.newegg.com</a>)</td>
<td>$40.00</td>
</tr>
</tbody>
</table>

Contractual Services

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>175 questionnaire copies $0.15 per copy at SIUE print and copy (SIUE)</td>
<td>$26.25</td>
</tr>
<tr>
<td>STAI software and 175 users</td>
<td>$302.50</td>
</tr>
<tr>
<td>HyperCam screen capture software to record usability tests</td>
<td>$35.00</td>
</tr>
<tr>
<td>(<a href="http://www.hyperionics.com/">http://www.hyperionics.com/</a>)</td>
<td></td>
</tr>
</tbody>
</table>

Travel

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel to ACM conference to present paper.*</td>
<td>$353.99</td>
</tr>
</tbody>
</table>

Totals

<table>
<thead>
<tr>
<th>Category</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity</td>
<td>$82.86</td>
</tr>
<tr>
<td>Contractual Services</td>
<td>$363.75</td>
</tr>
<tr>
<td>Travel</td>
<td>$353.99</td>
</tr>
<tr>
<td><strong>TOTAL:</strong></td>
<td><strong>$800.00</strong></td>
</tr>
</tbody>
</table>

* Additional funds available to support student travel and will be used