Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management
Bali, Indonesia, January 7 – 9, 2014

Development of a Graphical User Interface as a Data Collection Tool for Cognitive Performance

Katherine G. Bagley and Eui H. Park
Department of Industrial Engineering
North Carolina A&T State University
Greensboro, North Carolina 27411, USA

Abstract

Technology has changed society's climate in nearly every aspect of life. From communicating, working, and even banking, technological applications have undergone a nearly unidentifiable transformation. To offset potential misuse or frustrations, designers and researchers have collaborated for years in conducting usability studies to gain insight on potential error, to anticipate user behavior and to predict performance. Some studies have failed because the selected technology was outdated, resulting in futile outcomes and replication issues. In other cases, studies were inadequate in capturing different aspects of performance.

This paper provides a systematic manner of capturing cognitive performance through the development of a graphical user interface. The interface mimics a familiar navigation screen, MapQuest™, requiring users to complete seven tasks. Additionally, the seven tasks were linked to four main task types: fine motor, communication, numerical and cognitive. The use of the graphical user interface (GUI) enhanced the capability of focusing on use patterns and the variability of specific individual differences (computer self-efficacy and age) during the experiment. Two heuristic evaluations were performed to establish ecological validity of the interface. Lastly, these results encouraged utilizing an automated source of data collection to focus on the test participant and individual deviations during usability tests.

Keywords
Industrial engineering, usability analysis, human computer interaction, heuristic evaluation, systems analysis

1. Introduction

The world constantly evolves and technological enhancements are no exception. Technological devices are ubiquitous in nearly every facet of society. There are substantial investments made in organizations on a daily basis to improve information technology. Since inception, technological devices were designed to provide assistance to users; unfortunately, some users view those enhancements as sources of frustrations and anxiety (Barber, 1998). Researchers are constantly striving to eliminate the dichotomy between those who perceive technology as an indispensable tool and those who view technology as a convoluted distraction.

It is critical to understand the unique human-machine interaction. An underlying factor in this understanding is enhancing the overall usability of the human-machine interface. Usability is defined as how "user-friendly" the machine is, or how well the design allows users to meet their intended goals (Nielsen, 1993). Factors that affect usability include task requirements, cognitive constraints and cultural biases.

This paper employs a structured framework to facilitate the assessment of cognitive performance relying heavily on the use of a graphical user interface (GUI) for automatic data collection.

2. Literature Review

Individual differences such as demographic variables, computer experience, cognitive abilities and personality factors have long been important in end user computing research (Zmud, 1979). In reviewing literature regarding technological tasks, Social Cognitive Theory (SCT) is often applied; more recently, computer self-efficacy has been on the forefront of research. In the 1970s, Albert Bandura posited that behavioral changes occurred from dynamic interactions between personal, behavioral, and situational factors (Bandura, 1982). The very core of SCT is rooted in self-efficacy. “Self-efficacy is defined as the strength of one’s conviction that he or she can successfully execute a behavior to produce a certain outcome” (Bandura, 1997). Computer self-efficacy (CSE) is defined as “an individual’s beliefs about his or her capabilities to use computers.”(Heinnesen, 2005).

Therefore self-efficacy classification (high or low) may serve as a predictor of success. Likewise, age has received increased attention in terms of technological use. “Age has been considered one of the most influential individual variables in terms of technology use” (Czaja, 1998). Very few published studies have investigated computer self-efficacy and additional demographics, reflecting the relative neglect of cultural factors in literature” (Barber, 1998). Currently, the notion of the Digital Divide, the gap between people with effective access to technology and those with very limited access (Hoffman and Novak, 1996), has been studied to gain understanding of individual differences in technology. Advancing these constructs, this paper analyzes the effects of those two sociocultural factors, age and computer self-efficacy, on cognitive performance through an online GUI.

3. Methodology: GUI Development

The major requirement for the GUI was not only to provide a realistic scenario but most importantly, one that provided the types of task complexity desired for future analysis. After an exhaustive literature review, it was revealed that a major trend in technological aids, regardless of the industry was associated with navigation, word processing or first aid applications. Navigation was selected for this task scenario. Next, subject matter experts were leveraged to construct a task analysis, establish a performance baseline, and select the appropriate task types.

3.1 GUI Functionality

The test bed mimicked a traditional MapQuest™ screen. The map display, embedded with seven questions, assessed the participant’s performance in four task types: communication, numerical, cognitive and fine motor. Because multiple mobile devices were not feasible for the experiment, a stationary online prototype was used. The developed interface was stored at the public domain, www.aggieuxstudy.org. The website was created to facilitate ease of use in the multiple testing facilities. The interface combines of JavaScript, HyperText Markup Language (HTML), Dreamweaver, Personal Homepage Tools (PHP) and an Access database.

The test bed was developed using both client side and server side languages to create an interactive interface. The programming languages used for the interface creation were PHP (server-side language) and JavaScript (client-side language); while HTML (Hyper Text Markup Language), Dreamweaver and Access were supportive technologies that provided connectivity and data storage.

As the server side language, PHP allowed the web page to be functional and dynamic. AJAX (Asynchronous JavaScript) was also used to send data to and retrieve data from the server without interfering with the display. AJAX functioned as a group of interrelated techniques to assist with client side functionality. The use of JavaScript adjoined with Dreamweaver (code storehouse) provided a familiar and realistic look to the GUI through the enhancements of text fields, command buttons, and forms. Interface features such as the map graphics, clear and search features were enabled as well. Lastly, the Microsoft Office Access functioned as the backend database management system of the test bed to store confidentiality codes, participant answers, total task times, and completion times.

The main screen of the test bed layout encompassed two address entry fields (first and second) and a map graphical display (see Figure 1). Unlike MapQuest™, typical advertisements and media streams were removed to provide a more realistic picture of a user’s perspective.

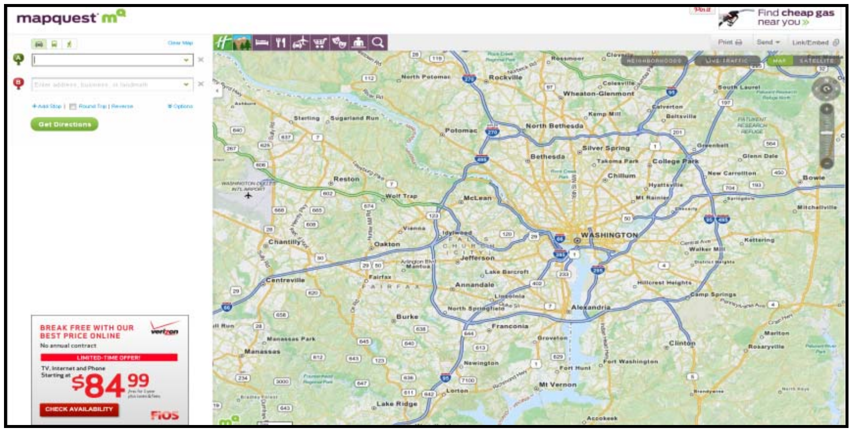

Figure 1: Navigational Interface

Functionality within the interface included: a Search Button, a Clear Button, a Help icon, an active timer, a Next Command (within questions), embedded pop-up style questions (see Figure 2).

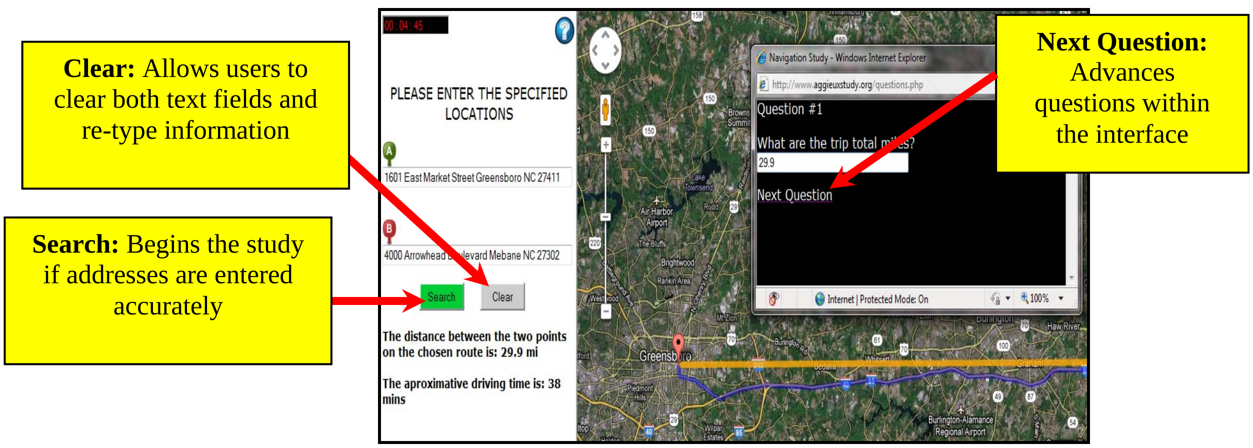

Figure 2: Test bed question display

The GUI functioned as a major source of data collection. The study identified specific performance metrics: total task time, task type time, error rate, and abandonment rate. While the GUI collected the objective data, the observer was able to monitor and record user comments, frustrations, trends, and/or patterns that may be missed during standard think aloud observations or even video. This capability to dissect and examine the user's actions immediately heightened understanding in individual differences.

3.2 Task Analysis

As mentioned earlier, this experiment required the participants to participate in four different task types which were linked to seven different questions. The association between the task types and the required questions is displayed in Table 1.

Table 1: Association of Task Type and Task Detail

<table>
<tr><th>Task Overview</th><th colspan="2">Task Type</th></tr>
<tr><td>1.0 Enter first address into text field</td><td rowspan="2">Fine Motor Taxon</td><td rowspan="9">Communication Taxon</td></tr>
<tr><td>2.0 Enter second address into text field</td></tr>
<tr><td>3.0 Locate total trip time and enter into pop-up field</td><td rowspan="4">Numerical Taxon</td></tr>
<tr><td>4.0 Locate total trip distance and enter into pop-up field</td></tr>
<tr><td>5.0 Convert total trip time into seconds</td></tr>
<tr><td>6.0 Convert trip distance into kilometers</td></tr>
<tr><td>7.0 Compare provided metric with total trip time to answer question regarding battery life</td><td rowspan="3">Cognitive Taxon</td></tr>
<tr><td>8.0 Identify specified target, and count total icons</td></tr>
<tr><td>9.0 Use click function to display desired screen</td></tr>
</table>

4. Methodology: Heuristic Evaluation

To ensure the GUI functioned accurately, a heuristic evaluation was conducted to make certain that no major usability problems existed within the design. This was a critical step since the majority of the study's data was collected automatically. "Heuristic evaluations are a usability engineering method for finding the usability problems in a user interface design so that they can be attended to as part of an iterative design process" (Nielsen, 1993). Heuristic evaluations are one of the most informal methods of usability inspection. Additionally, this technique was considered advantageous because it lacked complexity and monetary expenses. Lastly, heuristic evaluations only require one expert. Although heuristic evaluations can uncover major usability issues in a short period of time, a criticism with this technique is that the results are highly influenced by the knowledge and competency of the expert reviewer.

Even with the noted concern of the *"one-sided"* review, this method was selected and conducted with usability experts at the United States Army Research Laboratory (ARL) in Aberdeen, MD. The three experts, ARL Tool Development Lead, ARL Usability Analyst, and ARL Evaluation and Analysis Lead, evaluated the interface with the purpose of exposing any usability issues ensuring compliance with recognized usability principles (heuristics).These heuristics are general rules that describe common properties of usable interfaces. The evaluator team used Nielsen's Ten Heuristic principles (see Table 2).

Table 2: Nielsen’s Heuristic Principles

Name	Description
Visibility of system status	The interface should keep the user informed through appropriate feedback
Match between system and the real world	The interface should speak the users’ language, with their terminology, and it should follow real world conventions.
User control and freedom	Users should be able to leave unwanted state, or navigate as they wish.
Consistency and Standards	Users should not have to guess whether different words, situations or actions mean the same thing.
Error prevention	Errors should be prevented from occurring.
Recognition rather than recall	Minimize the memory load by supporting recognition and not requiring users to recall information from one screen to the next.
Flexibility and efficiency of use	Allow advanced users to speed up their interactions.
Aesthetic and minimalist design	Screen should not contain information that is irrelevant or rarely needed. Every extra element competes with something else.
Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language so that users can find a solution.
Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

To prioritize the list of exposed issues, an estimate for severity was required by each evaluator (see Table 3). For the assigned severity, the evaluators examined the frequency, persistency, and impact of the problem for each problem identified. Severity ratings provided a rough estimate of the need for additional usability resources and efforts. These descriptions allowed the evaluators to assess the various problems fairly easy.

Table 3: Nielsen’s Severity Ratings

Description	Rating
This is not a usability problem at all	0
Cosmetic problem only: Need not be fixed unless extra time is available on project	1
Minor usability problem: Fixing this should be given low priority	2
Major usability problem: Important to fix, so should be given high priority	3
Usability catastrophe: Imperative to fix this before product can be released	4

Each individual evaluator received task instructions, GUI access, and a set of Nielsen’s heuristics and severity ratings. Next, they inspected the interface independently. This procedure was important to make sure that the assessments were unbiased. Once all evaluations were completed, the evaluators were allowed to communicate and aggregate their findings. This collaboration addressed the most relevant, severe, and reoccurring problems. The general task for the evaluators was to record all pertinent issues with the sites, taking approximately an hour. These problems were then compiled into a comprehensive list.

The first session produced 52 issues. The second session (collaborated discussion) reduced that figure to 39 issues. The majority of the issues were related to flexibility and ease of use and error prevention (see Figure 3). These results were not surprising considering the interface has a very rigid structure to test user ability.

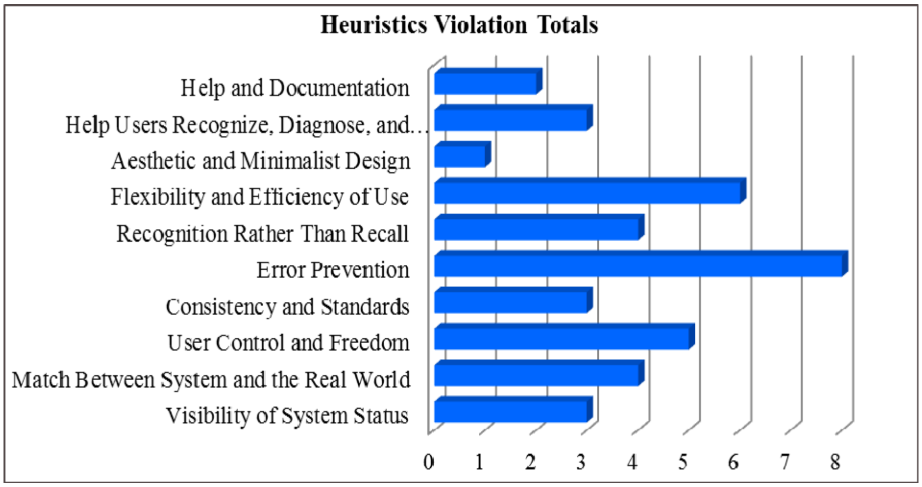

Figure 3: Heuristics violation totals

4.1. Second Heuristic Evaluation

Heuristic evaluations are iterative in nature. While the process is generally comprehensive in nature, it is stated that a good practice is to perform the evaluation an additional time to make certain that no major violations were present. As mentioned in the previous section, heuristic evaluations are heavily influenced by the knowledge and skill set of the expert. Therefore, if the expert misses a major violation, it is highly unlikely that it will be exposed within the group. To mitigate this concern, the evaluation was conducted twice with a different evaluation team for this interface.

The second evaluation was conducted in an academic setting: North Carolina Agricultural and Technical State University. Of the three reviewers, two were doctoral candidates in the Industrial and Systems Engineering program. The expert had more than five years of experience with web usability and the other reviewer had a strong focus in human robotic interaction. The last reviewer had over ten years of industry experience and was completing his Masters of Science in Computer Science specializing in Human Factors and Virtual Displays. The group used the same heuristic principles and severity ratings (Tables 2 and 3).

They followed the same regimented procedure of evaluating the interface independently while collaborating only after results had been documented. The results of the first session produced 38 issues. Once the group collaborated, that figure was reduced to 22 issues. Exactly half of the issues for this session were classified as minor usability problems while 41% were considered cosmetic issues (see Figure 4).

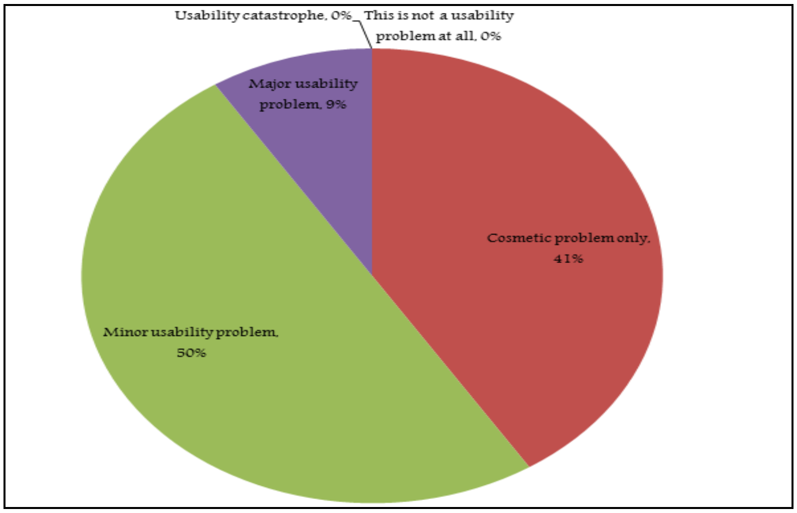

Figure 4: Second session heuristic evaluation severity ratings.

The violation totals varied within the second session as well. Those totals are provided in Figure 5.

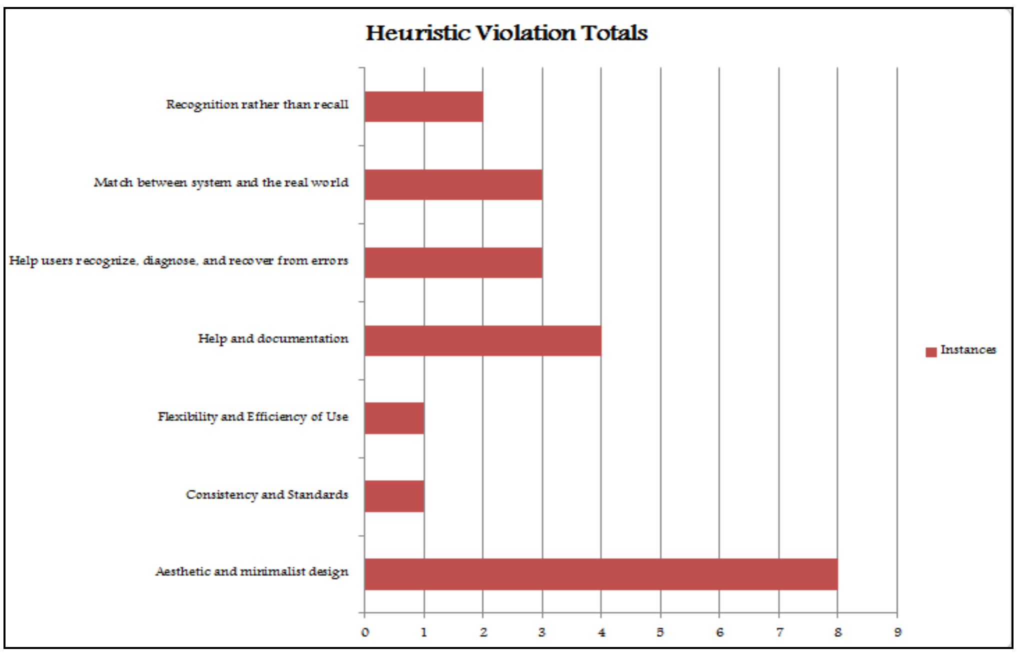

Figure 5: Second session heuristic violation totals

The ultimate concerns of the interface were the layout colors and visibility issues with the timer. While the timer was inserted to provide a mild level of anxiety, the original color scheme, made it difficult to read. Another concern was the figures within the Zoom feature. The team wanted to insert a method of inputting a percentage value associated with each click to aid the user with the final question that requires several clicks to display a desired screen (see Figure 6). The team also wanted to remove the pan function within the zoom feature as it was defined as confusing to the user. Lastly, the team wanted to eliminate the ability to scroll while in the interface to force users to only click. These were great ideas but unfortunately the majority of those concerns were not amendable. Those features were embedded in the Google Maps API software and could not be modified.

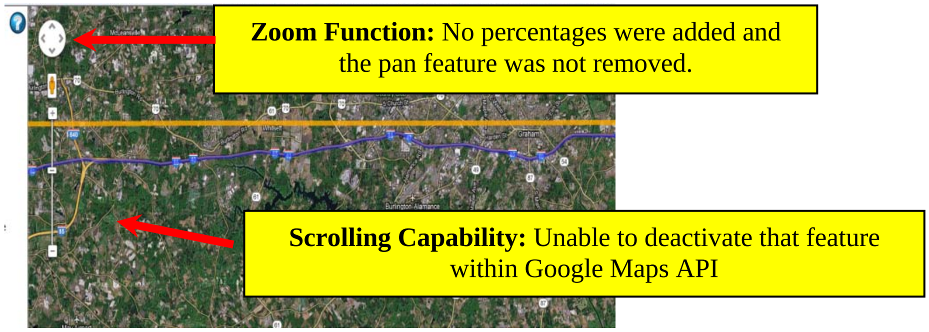

Figure 6: Zoom function in Google maps API

However the background color, font and timer were all changed as a result of the final evaluation (See Figures 7-8). Also, the help section was modified to include less precision in the digit display to help in the required numeric conversion tasks of the study. From this level of analysis, both teams, United States Army Research Laboratory and North Carolina Agricultural and Technical State University, felt confident launching the website for usability testing.

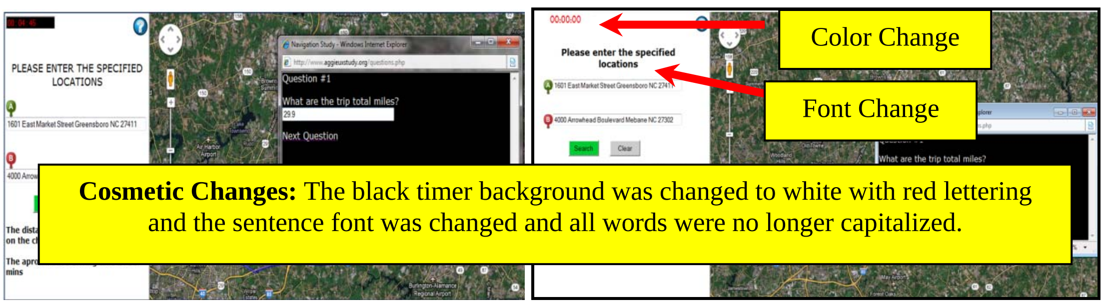

Figure 7: Cosmetic changes within navigational interface

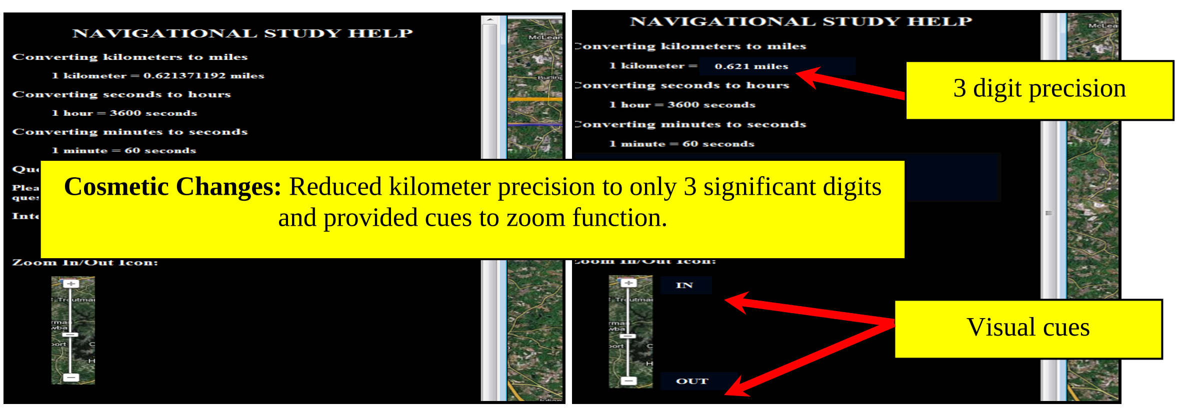

Figure 8: Cosmetic changes within Help section of navigational interface

5. Preliminary Results: Statistical Analysis

After the GUI was approved for use, a convenient sampling method of military participants was employed. 191 participants completed the study. The individual differences, computer self-efficacy and age, was captured through a self-reported tool prior to the online GUI. Because self-reporting instruments often reflect discrepancies due to blatant or tacit misreporting, the addition of empirical tests was conducted to increase the validity of the findings.

The experimental design used both a within-subjects design and a between-groups design to analyze five research hypotheses:

- H_1: Age is a predictor of mission success.
- $H_{2:}$ Computer self-efficacy is a predictor of mission success.
- $H_{3:}$ Age will have a statistically significant effect on fine motor time
- $H_{4:}$ The variability in computer self-efficacy is in some part is dependent on age in fine motor time.
- $H_{5:}$ Computer self-efficacy will have a significant effect on cognitive time.

The primary statistical analysis used was a logistic regression analysis to provide insight for performance prediction, then univariate analyses and factorial analyses were conducted to examine specific task types and their effects on performance. From the initial analysis, there was a statistically significant difference in mission success (completing all seven tasks successfully) between older and younger participants. Additionally, high computer self-efficacy individuals completed the mission 8.2% faster than low computer self-efficacy individuals and computer self-efficacy was a predictor in mission success. There was a statistical significant difference in older and younger participants' performance in fine motor tasks and numerical tasks. Notably older participants were more successful with the numerical tasks while younger participants performed better on the fine motor tasks.

6. Discussion

It is clear that technological pursuits will continue to permeate society and usability studies are important in advancing technology. Integrating cognitive theories, industrial and systems engineering, and usability principles to predict performance in an automated forum allows the enhanced effort of capturing use patterns in individual differences. For example, while gender was not a factor in the study, the majority of the females committed no errors with the fine motor tasks, while they consistently committed errors on the communication tasks. Also, even though the study only analyzed two age groups, there was a distinctive difference in the youngest population that made comments referring to SmartPhone applications during the numerical tasks and they noticeably struggled with the use of a standard calculator. This was supported in the statistical findings.

Additionally, this group commented on their frustrations with fine motor tasks, as they noted their familiarity with voice command software and more advanced technologies. Other trends that were captured were the rationale of abandonment percentages as well as exposure to language ambiguity within the GUI. These less tangible aspects of the experiment may have been lost if the investigator solely focused on capturing task time in the GUI.

Incorporating individual differences into experiments will lead to more meaningful findings in human performance modeling and pertinent evaluation strategies models. The GUI provided an innovative tool for storing various performance data. As technology continues to inundate society, it is critical to understand these differences in terms of performance prediction and error analyses.

References

Bandura, A. 1982. Self-efficacy mechanism in human agency. American Psychologist: 122-147.

Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman, New York.

Barber, W (1998). Culturabilty: The Merging of Culture and Usability. Human Factors and the Web. Atlanta: AT&T Labs.

Czaja, S. J., & Sharit, J. (1998). Age differences in attitudes toward computers. Journal of Gerontology, 5, 329–340.

Heinssen, Glass and Knight. "Assessing computer anxiety: Development and validation of the Computer Anxiety Rating Scale. Computers in Human Behavior. (1987) Vol 3 pp. 49–59.

Hoffman, Donna L., William .D. Kalsbeek and Thomas .P. Novak (1996), "Internet and Web Use in the United States: Baselines for Commercial Development," Special Section on "Internet in the Home," *Communications of the ACM*, 39 (December), 36-46.

Nielsen, Jakob. (1993) .Usability Engineering. San Diego: Academic Press.

Zmud, Robert (1979). *Individual differences and MIS success: A review of the empirical literature*. Management Science Vol. 25, No. 10 pp 966-977.

Biography

Katherine Bagley is an Industrial Engineer at Honda Aircraft Company where she is involved with ergonomic assessments, usability analysis, research endeavors and performance metrics. Dr. Bagley holds her Ph.D. in Industrial and Systems Engineering from North Carolina A&T State University. She is a member of Alpha Pi Mu, National Industrial Engineering Honor Society, the Human Factors and Ergonomics Society (HFES), and the Institute of Industrial Engineers (IIE). Her research interests include lean manufacturing, human robotics interaction, and usability analysis.

Eui H. Park is a professor in the Department of Industrial and Systems Engineering at North Carolina A&T State University. He had worked for Boeing Commercial Airplane Company as a senior engineer for four years from 1978, and returned to school for his doctorate with a Boeing Fellowship. Upon completion of his Ph.D., he joined North Carolina A&T State University and has since initiated and developed a successful interdisciplinary manufacturing program at the university as the Director of Manufacturing Initiatives. He is the founder of teaching factory, Piedmont Triad Center for Advanced Manufacturing and has also initiated and developed a successful Human-Machine Systems Engineering program at NC A&T. Dr. Park was also the Chairperson of the Industrial and Systems Engineering Department for 16 years from July 1990. He has been an IIE (Institute of Industrial Engineers) Fellow since 2000. His fields of research are Human-Machine Systems Engineering and Quality Assurance. He has been a principle investigator in 24 awarded funded research projects totaling over $12 million in the past ten years.