

AN IMPROVED SCALE FOR THE NOVICE-EXPERT RATIO METHOD

Tim J. Whalen

Submitted to the faculty of the School of Informatics
In partial fulfillment of the requirements
for the degree of
Master of Science in Human-Computer Interaction
in the School of Informatics
Indiana University

May 2008

Accepted by the Faculty of Indiana University,
in partial fulfillment of the requirements for the degree of Master of Science
in Human-Computer Interaction.

**Master's Thesis
Committee**

Karl F. MacDorman, Ph.D., Chair

Anthony Faiola, Ph.D.

Josette Jones, Ph.D.

© 2008
Tim J. Whalen

ALL RIGHTS RESERVED

Dedicated to
My wife, Jodi
and my kids, Jake and Kassi

ACKNOWLEDGEMENTS

I would first like to thank Dr. Karl F. MacDorman, my thesis advisor. His assistance and patience in helping me complete my thesis has truly been invaluable. This has been a long and difficult process but in the end well worth it thanks to Dr. MacDorman. Next, I would like to thank Dr. Anthony Faiola. Dr. Faiola was available with words of encouragement to keep me going and that is truly appreciated. I would also like to thank Dr. Josette Jones for agreeing to participate on this thesis committee. Many more thanks go to Chin-Chang Ho and Himalaya Patel. Their assistance with the numbers and the words are greatly appreciated.

Finally, I would like to thank my family. Without their love and support, none of this would have been possible. Let's go throw that ball, Jake.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
ABSTRACT.....	x
1. INTRODUCTION & BACKGROUND	1
Introduction to the Novice-Expert ratio Method.....	1
Importance of the Novice-Expert ratio Method	1
Intention of This Study.....	2
2. LITERATURE REVIEW	3
Heuristic Evaluation.....	3
Design Effects on Usability	4
Predicting Task Usability with GOMS	6
Novice vs. Expert Performance.....	8
Research Question.....	10
3. METHODOLOGY	12
Participants and Sampling.....	12
Data Collection Tools.....	13
Procedures	15
Statistical Analysis	17
4. RESULTS	18
Factor Breakdown	18
Factor Analysis Matrix.....	19
Correlation of Study Questionnaire by Clicks and Completion Time	22
Correlation of Factors.....	25
Correlation of Measures by Each Factor.....	26
Results Summary.....	29
5. DISCUSSION	31

Interface Design	31
Factor Analysis.....	31
Participant Performance by Measure	33
Implication of Results	34
6. CONCLUSION	35
Limitations	35
Future Research.....	35
Summary	36
REFERENCES.....	37
Appendix A: Consent Form	39
Appendix B: Demographic Information and Pre-Screening Questions	41
Appendix C: Post Study Questionnaire Data.....	43
VITA.....	47

LIST OF TABLES

Table 1: Nine Usability Heuristics	4
Table 2: Advantages and Disadvantages of the Four GOMS Techniques	6
Table 3: Total Variance Explained by Pre-Study Questionnaire	21
Table 4: Factor Matrix of Self-reported Pre-Study Questionnaire	21
Table 5: Total Variance Explained by Self-reported System Usability Questionnaire	21
Table 6: Factor Matrix of Self-reported System Usability Questionnaire	21
Table 7: Total Variance Explained by Performance Questionnaire	22
Table 8: Factor Matrix of Performance Questionnaire	22
Table 9: Correlation of Questions in Auto-Completion Experience, iTap Experience, Subjective Usability, Cellular phone Knowledge, iTap Knowledge, Clicks on Average, and Time on Average	24
Table 10: Correlation Matrix of Auto-completion Experience, iTap Experience, System Usability, Cellular Phone Knowledge, iTap Knowledge, Clicks on Average, and Time on Average	25
Table 11: Correlation of Auto-Completion Experience, iTap Experience, Usability, Cellular Phone Knowledge, iTap Knowledge, NE Ratio, Log, Median, Reciprocal, D , and Cohen's d	29

LIST OF FIGURES

Figure 1: The effect learning curves have on novices and experts	5
Figure 2: Phone Interface	14
Figure 3: Word 1, as it is presented to the participant	16

ABSTRACT

The Novice-Expert ratio Method (NEM) compares the time it takes a novice user to complete each step of a task with the time it takes an expert user to complete the same step. A high ratio indicates that novice users are performing a step far below expert performance. This allows researchers to identify aspects of a user interface that impede its usability, because they are counterintuitive. However, NEM has three major shortcomings. First, the merits of a ratio scale have never been appraised relative to other possible scales, such as log, median, reciprocal, D , and Cohen's d . Second, user error rates are not included in calculating the NEM ratio. Third, NEM does not specify how much time to allow for the completion of a step or what completion time to use for a step that was not completed.

This study tested the validity of the original ratio scale of NEM and other candidate scales to ascertain whether another scale based on the objective usability measures of completion time and steps required might better represent the concept of usability. Scale validity was tested by comparing how well each scale correlates with other kinds of usability measures, including the ability to answer questions accurately and self-reported ratings of usability. A dataset was constructed from the amount of time and number of clicks it took participants to enter words on a counterintuitive cellular phone interface as implemented on a webpage. Cohen's d measure correlated more highly with the number of clicks and completion time for novice-expert performance than the original NEM ratio.

CHAPTER 1. INTRODUCTION

Introduction to the Novice-Expert ratio Method

The Novice-Expert ratio Method (NEM), developed by Haruhiko Urokohara, Kenichi Tanaka, Kazuyoshi Furuta, Michiyo Honda, and Masaaki Kurosu, is described as an “evaluation method that generates a quantitative measure in terms of the usability of any specific system” (Urokohara et al., 2000). Essentially, NEM is used to compare the time it takes a novice user to complete each step of a task, to the time it takes an expert user to complete the same step. An expert user is believed to represent the minimum amount of time it takes to complete the step (Urokohara et al., 2000). A high ratio between novice and expert users indicates that the novice user is completing the step far below expert performance. This allows researchers to identify counter-intuitive aspects of a user interface. In the original NEM testing, the researchers recorded the time needed for users to locate a destination point using a car navigation system, to let the system determine the optimum route, and to display the traffic conditions on the highway. The ratio was developed based on the amount of time it took novice users to perform this task compared to the amount of time it took expert users.

Importance of the Novice-Expert ratio Method

Designing product interfaces that are easy to use is recognized as essential to gaining competitive advantage in the marketplace. This part of the development cycle

needs to be thoroughly planned and administered to avoid creating cumbersome and overly complicated designs. Card, Moran, and Newell (1983) believe the total performance of the system should be taken into account, including the psychological characteristics of the user and how they interact with each task and the computer itself.

Intention of This Study

NEM provides an excellent basis for gauging the difficulty that a typical user will have while completing a specified task. This study will test the validity of that basis and determine if any enhancements can be made. The validity of the scales and parameters used will be compared with its correlation to alternative measures of usability.

CHAPTER 2. LITERATURE REVIEW

Heuristic Evaluation

It has been noted that any system that is designed for people should be easy to learn and remember, effective in their work, and pleasant to use (Gould and Lewis, 1985; Molich and Nielsen, 1990). To accomplish these goals, an analysis of the usability of the system must be performed to determine any gaps that may exist. Heuristic evaluation was developed as a method for finding these gaps. In a heuristic evaluation, a group of evaluators examines a user interface and tests its adherence to recognized usability principles (Nielsen, 1992). Desurvire, Lawrence, and Atwood (1991) found that this method is most effective when the evaluators are usability experts. They used the results of laboratory tests to evaluate the performance of the heuristic evaluation method, comparing the “best guesses” of experts and novices to those of usability experts. Their results show that both approaches were able to predict the laboratory test results. Additional research has shown that expert evaluation is most effective when the evaluators are usability experts.

Originally, usability guidelines were very complex and consisted of over 1,000 rules for developers to follow. Evaluators relied primarily on their experience and common sense. Molich and Nielsen (1990) simplified these rules into nine basic principles, listed in Table 1, for performing a heuristic evaluation.

Table 1:
Nine Usability Heuristics (Molich and Nielsen, 1990)

Description
Simple and natural dialogue
Speak the user's language
Minimize user memory load
Be consistent
Provide feedback
Provide clearly marked exits
Provide shortcuts
Good error messages
Prevent errors

Design Effects on Usability

Usability concerns are not limited to novice users. Kjeldskov, Skov, and Stage (2005) studied the usability problems that affect both novice and expert users over time. They used three measures. The first was the effectiveness and efficiency of the test participants. The second was the number and severity of usability problems that the novice and expert test groups experienced. The third was the test participants' subjective workload. They concluded that while additional time does reduce or eliminate some usability problems, expert users are not more efficient on complex tasks over time than novice users. In his book *Usability Engineering*, Jakob Nielsen (1994) described how learning curves affect both novice and expert users as they learn a system (Figure 1). This figure depicts the learning curve for a novice user where the system is easy to learn but less efficient to use. The learning curve of the expert user displays a system that is more difficult to learn but more highly efficient over time. Nielsen also outlined specific steps needed to provide a highly usable user interface.

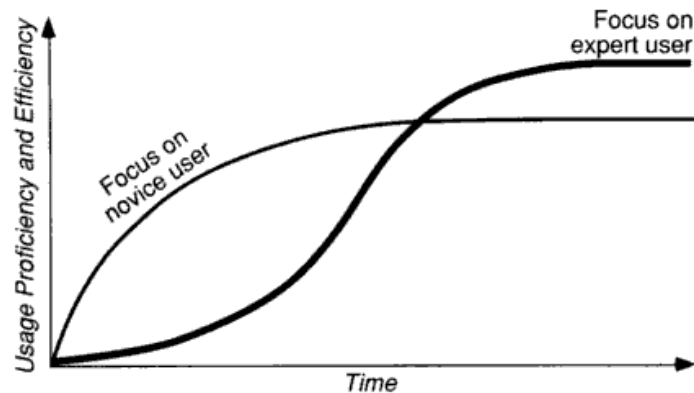


Figure 1: The effect learning curves have on novices and experts (Nielsen, 1994)

Users who try to climb a system’s learning curve face distinct challenges. Carroll and Rosson (1987) argued that many usability issues affecting novice computer users arise from conflicting motivational and cognitive strategies and not necessarily poor interface design. They attribute these conflicts to two types of biases: production biases and assimilation biases. Production biases arise when the only goal of users is task completion, and the user is not motivated to learn how to use the system. Assimilation bias occurs when users apply what they already know to interpret new situations. Fischer (2001) studied how user modeling has attempted to address the issue of coding a user-centered application to meet one user’s expectations and subsequently releasing it to millions of other users. The “value / effort” quotient is used to assess the payoff for user modeling. To determine the usefulness of user modeling, the “value” needs to be increased, showing that systems relying on user modeling are more usable and useful or that the “effort” put into creating the user model must be decreased.

Predicting Task Usability with GOMS

A structured understanding of users' mental processes can aid interface designers. In their book *The Psychology of Human-Computer Interaction*, Card, Moran, and Newell (1983) describe the four components of the user's cognitive structure: Goals, Operators, Methods for achieving the goals, and Selection rules for choosing among those methods. These components are more commonly known by the acronym GOMS.

Table 2:
Advantages and Disadvantages of the Four GOMS Techniques

Technique	Advantages	Disadvantages
KLM	<ul style="list-style-type: none"> - Simplest to use - Execution time is easily approximated by timing each individual method 	<ul style="list-style-type: none"> - Does not predict the method based on the task situation - Methods must be listed in sequence - Not practical for high-level analysis
CMN	<ul style="list-style-type: none"> - Can predict operator sequence and execution time - Can represent high-level goals - Easy to write 	<ul style="list-style-type: none"> - No description of the method representation or methods used in task execution - Vague and unspecified
NGOMSL	<ul style="list-style-type: none"> - Can represent high-level goals - Provides useful descriptions of a task at many levels - Provides reliable quantitative analysis at the keystroke level 	<ul style="list-style-type: none"> - User must already know the operators to provide meaningful predictions of learning and execution times
CPM	<ul style="list-style-type: none"> - Predicts execution time from component activities - Perceptual, cognitive, and motor operators can be performed in parallel 	<ul style="list-style-type: none"> - Primitive operators must be simple perceptual, cognitive, and motor acts - Assumes that the user is an expert

John and Kieras (1996) discussed four different GOMS techniques and how they are related. Table 2 lists the advantages and disadvantages of each. The Keystroke-Level Model (KLM-GOMS) estimates execution time when the order of the operations is known. In such cases this method also offers a simple quantitative analysis of the results. It works by listing the operators for the task in sequential order. The total execution time for each operator is used to estimate the execution time for the entire task. Card, Moran, and Newell GOMS (CMN-GOMS) predicts the operator sequence and execution time for a particular task. It uses a strict goal hierarchy and assumes nothing about how the methods are represented or executed. Natural GOMS Language (NGOMSL) is a structured natural language notation that predicts operator sequence, execution time, and time to learn the methods. NGOMSL represents methods in terms of cognitive complexity theory (CCT), which has provided good predictions of execution and learning time and transfer of procedural learning. The Cognitive-Perceptual-Motor GOMS (CPM-GOMS) uses operators that must be simple perceptual, cognitive, and motor acts. This is the only GOMS technique that allows operators to be performed in parallel. With proper parameterization, any GOMS method may be used to simulate expert performance. However, because experts are more likely to perform actions in parallel, CPM-GOMS has the added advantage of being able to model this.

Novice vs. Expert Performance

Researchers understand that testing novice and expert users on the same task or tasks provides additional information on those tasks (Goonetilleke et al., 2001). Uehling and Wolf (1995) applied this in their development of UsAGE: User Action Graphing Effort. UsAGE is a usability testing tool that compares the performance of a novice user to that of an expert user during a set of predefined application tasks. Recordings of the novice and expert users are compared graphically. Expert actions are listed horizontally in the top row. Arrows are used to show the order in which the novice user completed each action. Arrows with arced lines indicate actions that were performed out of sequence. Each out-of-sequence action is represented by a node placed underneath the expert actions.

Although UsAGE permits visual comparison of the results, two deficiencies stand out. One is the inability to compare how a group of novice users performed relative to the expert user; this system only compares one novice user at a time. While one novice user may have difficulty in completing a task, another novice user may not. So although UsAGE helps determine where individual usability problems exist, it does not provide an overall analysis of the usability of a system. Another deficiency is that the results do not show the severity of a given usability issue. They are able to show where novice users had problems with a particular task but not the level of difficulty of that step.

Dillon and Song (1997) developed a system for testing novice and expert user performance using a customized online search tool. The system tested the performance of both types of users with a text-based search interface and a graphical search interface.

The study measured the participants in three categories: completion time, accuracy, and navigation. The mean and standard deviation of the task completion times for both groups were recorded. Accuracy was calculated by assigning a point value for the answers to each of the search questions. Two points were awarded for a correct answer, one point for a partially correct answer, and zero points for an incorrect answer. Experts were slightly better than novices at finding accurate information. There was a slight trend favoring the graphical interface, but this trend was not statistically significant. Navigation was calculated by counting the transitions between various nodes. It was assumed that the more transitions a user made the more difficulty the user had in locating the correct information. From this assumption it was determined that the participants had a much higher navigation rating with the text-based search interface than with the graphical interface.

While novice users showed a slight improvement in their ability to perform searches with the graphical interface, expert users showed no increase in performance between the text and graphical interfaces. However, by capturing and analyzing only the time needed to complete all the tasks, the researchers may have masked problems related to the search questions and not related to the participant's ability. Timing each search question and then totaling the amount of time needed to complete all of the tasks could have revealed individual searches that were problematic.

The Novice-Expert ratio Method (NEM) provides an objective measure of user performance to assist engineers in pinpointing usability problems. These problems are identified by comparing the time a novice user needs to complete a subtask with the time an expert user needs to complete the same subtask. The larger the ratio, the more likely a

usability problem exists. Urokohara, Tanaka, Furuta, Honda, and Kurosu (2000) developed this quantitative measurement, because engineers can easily accept it.

However, to evaluate new products, expert users are not always available. Training novices until they become experts is time consuming and costly. Even when a few experts are available, perhaps those involved in product development, using developers as experts introduces its own kind of bias: Their usage patterns are determined by development and testing requirements rather than use in the field, and they have insider knowledge about the internal workings of the system. They will therefore be using somewhat different mental models from those of expert users. There is also some variance in the individual performance of experts. This necessitates a reasonably large sample size for results that are generalizable. Stable estimates of expert performance can be rapidly calculated using KLM-GOMS tools, such as CogTool (Luo and John, 2005), based on reliable models (e.g., models of human movement, such as Fitts's law). KLM-GOMS can complement data from expert users in estimating expert performance.

Research Question

The goal of this study is to appraise the validity of the Novice-Expert ratio Method as a measure of usability by performing a correlational analysis with other kinds of usability measures, such as self-reported usability ratings and declarative knowledge about the interface. Proposed alternatives to the ratio scale include log, median, reciprocal, D , and Cohen's d . (Completion time data typical have extended upper tails, because some participants can take a long time to complete a task or even fail to complete

it. The log and reciprocal scales are frequently used to improve central tendency estimates by removing the positive skew from the data. Although the median is seldom used for completion time, it provides a useful baseline for gauging the quality of other measures. Magnitude differences among experimental groups are typically correlated with the underlying variability of the data. Dividing means by the standard deviation adjusts for this. D divides the difference between novice and expert means by the standard deviation of both groups, whereas Cohen's d divides the mean of the novice group by the standard deviation of the novice group and the mean of the expert group by the standard deviation of the expert group before taking the difference.) After the data are collected, these alternative novice-expert scales will be used to determine whether it is possible to improve on the validity of NEM.

R1: Which scale has the highest validity for measuring the usability of an interface based on stepwise completion time?

Completion time is only one measure of usability. Users are more likely to feel frustrated with an interface, if its design causes them to make many errors. Although completion times often increases with error rate, this is not necessarily the case, especially when hastily performed actions can be quickly undone (e.g., using the backspace key when typing). Therefore, in addition to completion time, it is important to consider the number of steps required to complete a task.

R2. Which scale has the highest validity for measuring the usability of an interface based on the number of steps required to perform a task?

CHAPTER 3. METHODOLOGY

Participants and Sampling

Participants for this study were recruited by e-mail using a random sample from a list of undergraduate students and recent graduates from eight campuses administered by a Midwestern university. Minors were excluded from participation. The participants did not receive compensation. Participants with dial-up and satellite Internet connections or those who had repeated the testing process were also excluded.

There were a total of 337 participants. Of those, 62.3% were female ($N = 210$) and 37.7% were male ($N = 127$). The age range of the participants was 18 to 57. The plurality of participants were aged 19 and 20, which accounted for 35.6% of all respondents ($N = 120$). The range of 23 and older made up the next age range at 32.3% ($N=109$). 21 and 22 year olds accounted for 19.3% ($N = 65$), and 18 year olds accounted for 12.8% ($N=43$) of the participants. The mean age of the participants was 23 with a standard deviation of 7. 92.3% of the participants had at least some college experience ($N =311$). 83.4% reported having some college ($N = 281$). 29 were college graduates and 1 had a master's degree. Participants with some high school made up 0.6% ($N = 2$) and 7.1% were high school graduates ($N = 24$). The United States was the birthplace for 93.5% of the respondents ($N = 315$) with the next being the Virgin Islands with 0.9% ($N = 3$). China, Germany, and Mexico each had 2 participants and 13 other countries accounted for the remaining participants. Only 8 of the participants reside outside of the United States.

Data Collection Tools

Participants were asked to spell a series of words on a web-based, text interface tool that mimics iTap word prediction software. They were able to complete the tasks in an environment of their choosing, although an area free from distractions and interruptions was recommended. While the site was only tested using Microsoft Internet Explorer and Mozilla Firefox, any web browser that could properly access the website could be used.

The interface (Figure 2) was based on common cellular phone design and layout, although no particular cellular phone was used as a basis. The layout of the interface is designed in 3 sections. The smaller box at the top displays the letters as they are being chosen by the user. The second, larger box displays the partial spelling of the word based on the corresponding number selection and the letters previously selected. The third section of the interface is the input section. As numbers are selected on the keypad interface, the letters associated with the numbers are displayed in the middle text box. The first letter from the associated number appears in the upper text box. If this letter is the next letter in the word being typed, the participant can simply select the next number. If the letter is not correct, the participant must select it from the list of partially spelled words by clicking on it. Once clicked, the selection will appear in the above box. This process continues until the word is spelled correctly. The user then clicks the submit button to confirm that the spelling is correct and to continue to the next word. If the spelling is incorrect, the user is prompted to go back and fix it. A delete button is provided so that incorrect letters can be removed.

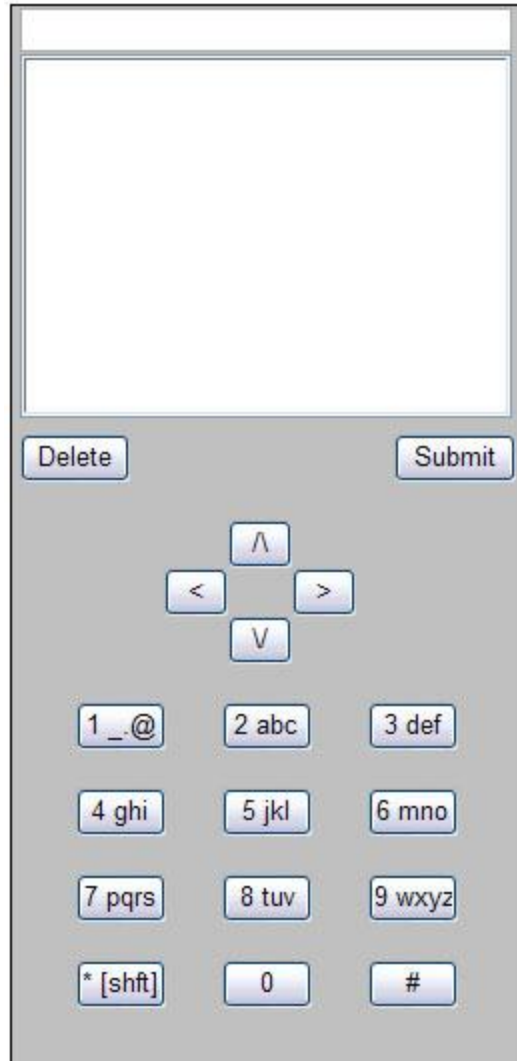


Figure 2: Phone Interface

Data collection consisted of the number of button clicks each participant made per word and the amount of time it took to successfully complete the spelling. Data collected from this study was stored in an SQL Server 2005 database.

Procedures

A website was created to perform the data collection portion of this study. The data collected came from three sections: a pre-study questionnaire, the data collection phase, and a post-study questionnaire and evaluation. All sections were required. Any results not fully completed were not used in the analysis.

Participants were first given a brief overview and description of the study and the technologies involved. After consenting to the study information sheet (Appendix A), each participant was required to complete a pre-study questionnaire that asked for demographic data and assessed their previous experience with iTap and automatic text completion technologies. A listing of the information collected and the pre-study questions can be found in Appendix B. Instructions and a sample interface were provided to give the participants a brief introduction on how to interact with the interface. The participants were then presented with the first of the 10 words to spell. Figure 3 shows the interface with the word to be spelled. The words used were *phone*, *normal*, *comic*, *bench*, *pancake*, *enable*, *ring*, *tunnel*, *detach*, and *focus*. These words were chosen to vary the numerical patterns used in their spelling on the interface. This was done to insure the time to complete the spelling would accurately reflect normal use of the interface and not be influenced because words had similar spellings. The words were selected from a dictionary consisting of 45,402 words.

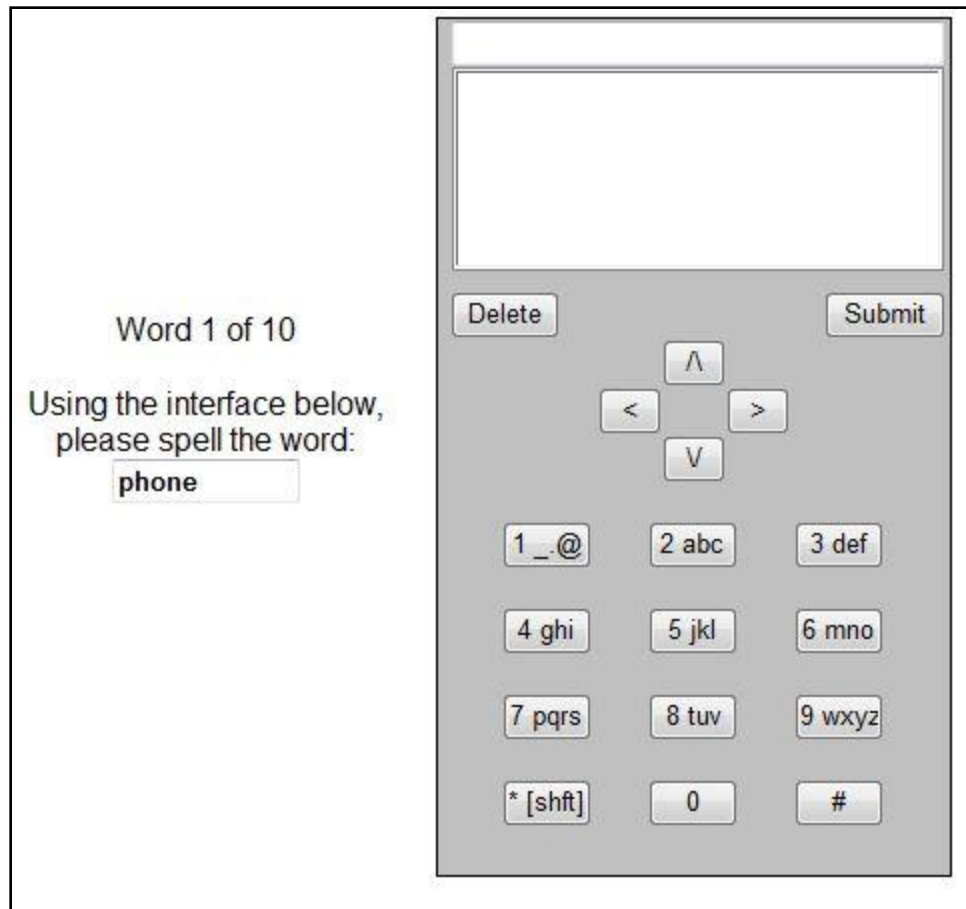


Figure 3: Word 1, as it is presented to the participant

Post-study questions consisted of two sections. The first section gathered information on the usability of the system. The second section tested the participant's knowledge of how the interface works. The post-study questionnaire was based on a subjective usability measurement used in a study by James R. Lewis (Lewis, 1995). A complete listing of the post-study questionnaire can be found in Appendix C.

Statistical Analysis

The data collected was analyzed using a series of database queries written in T-SQL and SPSS. Factor analysis, correlation, and regression modeling were applied. A correlational analysis was performed on the number of clicks versus the amount of time it takes to complete the spelling of the word. The correlation was based on the age and gender of the participant as well as the answers provided in the pre and post-study questionnaire. The log, median, reciprocal, D , and Cohen's d (effect size) were calculated for the number of clicks it took the participants to spell the word, and it was calculated for the amount of time it took to complete the spelling successfully. These calculations were based on the auto-completion experience, iTap experience, self-reported system usability, cellular phone knowledge, and iTap knowledge sections of the questionnaires. The D measure calculates the equal-weight average of the means of the minimum and maximum clicks and completion time. The novice data for the NE ratio was collected from each of the words that did not meet the minimum number of clicks or completion time. The expert data was collected by calculating the minimum number of clicks and the by selecting the quickest completion time for each word. The Novice-Expert (NE) ratio was then calculated and compared with the other scales for the data collected with respect to the number of clicks and the amount of time it took to complete the spelling. The statistical significance used in this study is 2-tailed. The statistical significance is noted as either .01 or .05.

CHAPTER 4. RESULTS

This chapter analyzes the results of the participant's responses. Initially, the total variance was calculated for the self-reported pre-study questionnaire, the system usability questionnaire, and the post-study task questionnaire. These results were then used to calculate a factor matrix on each section. Next, the responses from the pre-study questionnaire were correlated with the average number of clicks and the average amount of time for each word and were grouped by each factor. The correlation among all factors was calculated and with the number of clicks and the amount of time. Finally, these results are summarized and grouped by each factor. A correlation is performed among NE ratio, log, median, reciprocal, D , and Cohen's d to determine which measure had the highest validity. This summary is reported for clicks and time, and a review is performed for the significant differences.

Factor Breakdown

The data collected was grouped into a series of factors. These factors were based on the responses from the pre-study questionnaire and the post-study questionnaire. The auto-completion experience factor was determined from the responses to the questions regarding expertise and frequency of use with automatic text completion on any kind of device and cellular phone. iTap experience was determined from the participant's self-assessed expertise with iTap technology and the number of times the participant has used this technology on a cellular phone. Following the data collection phase, participants were asked their opinion on the usability of the interface. The results from these

questions make up the system usability factor. The final task for the participants was to complete a test on the performance of the cellular phone interface. Participants were quizzed on certain scenarios and asked to respond with a multiple choice answer. Answers to these questions were grouped into two separate factors. The first factor generated was iTap knowledge. These questions centered on the letters that were being displayed in the interface and the interactions needed at specific times during the spelling of a word. The second factor generated was cellular phone knowledge. This concentrated on the actual pressing of keys and the numbers and symbols associated with those keys. A complete listing of the pre-study and post-study questionnaires can be found in Appendix B and Appendix C, respectively.

Factor Analysis Matrix

The factor analysis matrix is separated into three sections. The first section is the results of the pre-study registration questions and questionnaire. This was used to calculate the total variance of the auto-completion experience and iTap experience factors (Table 3). The auto-completion experience factor accounted for 57.8% of the variance ($s^2 = 3.43$). The iTap experience factor accounted for 23.5% of the variance ($s^2 = 1.41$). A factor analysis (Table 4) was calculated on each of the two components. For auto-completion experience, this analysis showed that the self-assessed skill level with automatic text completion technology on cellular phones and on any kind of device were the two coefficients with the highest factor loadings, 0.85 and 0.81, respectively. With iTap experience, the frequency of iTap use with a cellular phone had a coefficient effect

of 0.73 and the reported skill level had a coefficient of 0.72. The Cronbach α value for overall internal reliability of the pre-study questionnaire was 0.82.

The second section analyzes the post-study system usability questionnaire for the system usability factor. A variance calculation (Table 5) was performed which showed a total variance of 62.13% ($s^2 = 4.97$). A factor matrix (Table 6) was created. It showed a high factor correlation for the questions on being able to enter a word quickly, simple to use the interface, and comfort in using the interface: each was .83. Satisfaction with the ease of using the interface and efficiently being able to enter a word had a factor loading coefficient of 0.82. The Cronbach α value for overall internal reliability of the self-reported usability questionnaire was 0.91.

The third section consisted of the results of the post-study task questions, which concerned declarative knowledge how the interface works. This section analyzed the knowledge and cellular phone knowledge factors and found a total variance of $s^2 = 1.88$ and $s^2 = 1.17$, respectively (Table 7). These results were then used to create a factor analysis on the two components (Table 8). For iTap knowledge, the question on the number of clicks to type *cab* had the highest factor loading at 0.58, and the questions on selecting *w* before pressing the next letter and the number of clicks required to type *aaa* had a factor loading of 0.57. For cellular phone knowledge, the question on the number to press to get the letter *n* had a factor loading of 0.67. The Cronbach α value for the iTap performance factor was 0.49.

Table 3:
Total Variance Explained by Pre-Study Questionnaire

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	3.43	57.18	57.18
2	1.41	23.50	80.68

Table 4:
Factor Matrix of Self-reported Pre-Study Questionnaire

		Factor	
		1	2
Completion	Skill level of automatic text completion technology on a cellular phone	.85	-.30
Experience	Skill level of automatic text completion technology on any kind of device	.81	-.30
	Frequency of automatic text completion on a cellular phone	.78	-.24
	Frequency of automatic text completion on any kind of device	.75	-.35
iTap	Frequency of iTap on a cellular phone	.65	.73
Experience	Skill level of iTap on a cellular phone	.66	.72

Table 5:
Total Variance Explained by Self-reported System Usability Questionnaire

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	4.97	62.13	62.13

Table 6:
Factor Matrix of Self-reported System Usability Questionnaire

		Factor
		1
System	I am able to enter a word quickly using this interface.	.83
Usability	It was simple to use this interface.	.83
	I feel comfortable using this interface.	.83
	I am satisfied with how easy it is to use this interface.	.82
	I am able to enter a word efficiently using this interface.	.82
	I believe I became productive quickly using this interface.	.77
	I can effectively enter a word using this interface.	.74
	It was easy to learn to use this interface.	.65

Table 7:
Total Variance Explained by Performance Questionnaire

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	1.88	23.54	23.54
2	1.17	14.59	38.13

Table 8:
Factor Matrix of Performance Questionnaire

		Factor	
		1	2
iTap	How many mouse clicks are required to type <i>cab</i> and submit it?	.58	-.36
Performance	If the next letter is <i>w</i> and you press the button <i>9 wxyz</i> , do you need to select <i>w</i> before entering the next letter?	.57	-.04
	How many mouse clicks are required to type <i>aaa</i> and submit it?	.57	-.56
	If the next letter is <i>b</i> and you press the button <i>2 abc</i> , do you need to select <i>b</i> before entering the next letter?	.55	.30
	What requires more mouse clicks? Typing <i>add</i> or <i>bad</i> ?	.49	-.04
	If you are typing <i>phone</i> and you have already typed <i>phm</i> , only one choice appears: <i>pho</i> , do you still need to select <i>pho</i> before entering the next letter?	.43	.19
Cellular		.30	.67
Phone	What number do you have to press to enter the letter <i>n</i> ?		
Performance	Is the delete button labeled <i><</i> ?	.30	.37

Correlation of Study Questionnaire by Clicks and Completion Time

After the results from the pre-study and post-study questionnaires were grouped into each of the factors, they were then correlated with the average number of clicks and the average amount of time it took the participants to complete the spelling of each of the words. With regards to clicks, system usability showed the highest statistical significance with all of the responses being significant. Participants cited they were able to effectively enter a word ($r = -.26$) and they felt comfortable using this interface ($r = -.23$). Similar

results were found with respect to the average time to complete each spelling. Completion time was correlated with ease of learning ($r = -.18$) and productivity using the interface ($r = -.18$). Results were found to be statistically significant for all of the question responses except for the ability to efficiently enter a word using the interface. This was not found to be statistically significant. In general, participants who made fewer clicks and were faster at spelling each of the words were more satisfied with the usability of the system. iTap knowledge had similar findings for clicks on average where all of the results were found to be statistically significant except for the question on typing *phm* and selecting *pho* first, which was not found to be statistically significant. In contrast, none of the findings for time on average were found to be statistically significant. Participants reporting a high skill level with automatic text completion on any kind of device ($r = .12$) and on cellular phones ($r = .11$) had a statistical significance for clicks on average. For time on average, a high skill level on any device ($r = -.13$) and a high skill level on cellular phones ($r = -.16$) were found significant. Demographically, it took older participants fewer clicks to complete the spelling of the word ($r = -.13, p < .05$) but it also took them more time ($r = .15, p < .01$). Table 9 contains the complete results from this correlation.

Table 9:

Correlation of Questions in Auto-Completion Experience, iTap Experience, Subjective Usability, Cellular Phone Knowledge, iTap Knowledge, Clicks on Average, and Time on Average.

	Clicks	Time
Gender		
Male	.05	-.06
Age	-.13 *	.15 **
Auto-Completion Experience	.11 *	-.11 *
Skill level for any device	.12 *	-.13 *
Frequency for any device	.09	-.06
Skill level for cellular phone	.11 *	-.16 **
Frequency for on cellular phone	.07	-.07
iTap Experience	-.01	-.06
Skill level for iTap	.01	-.09
Frequency for iTap	-.02	-.04
Subjective Usability	-.24 **	-.18 **
Satisfied with how easy it is to use this interface	-.14 **	-.16 **
Simple to use this interface	-.22 **	-.12 *
Effectively enter a word using this interface	-.26 **	-.15 **
Quickly enter a word using this interface	-.13 **	-.12 *
Efficiently enter a word using this interface	-.19 **	-.08
Feel comfortable using this interface	-.23 **	-.14 *
Easy to learn to use this interface	-.13 *	-.18 **
Quickly Productive to use this interface	-.21 **	-.18 **
Cellular Phone Knowledge	.00	.04
What number do you have to press to enter the letter <i>n</i> ?	.00	.02
Is the delete button labeled <?	.00	.04
	**	
iTap Knowledge	-.27	-.10
If the next letter is <i>w</i> , do you need to select <i>w</i> first?	-.22 **	-.09
If the next letter is <i>b</i> , do you need to select <i>b</i> first?	-.11 *	-.01
If you have already typed <i>phm</i> , do you still need to select <i>pho</i> first?	-.03	-.03
What requires more clicks? Typing <i>add</i> or <i>bad</i> ?	-.14 **	-.10
How many mouse clicks are required to type <i>aaa</i> and submit it?	-.19 **	-.03
How many mouse clicks are required to type <i>cab</i> and submit it?	-.15 **	-.04

** Correlation is significant at the .01 level (2-tailed)

* Correlation is significant at the .05 level (2-tailed)

Correlation of Factors

The correlation among all of the factors was calculated. Those factors were also correlated against the clicks on average and time on average. Those results show that participants with fewer clicks and a faster word completion time have greater satisfaction with the usability of the system. These results were statistically significant. Participants who tested higher on their declarative knowledge of the iTap interface (iTap knowledge) completed words with fewer clicks ($r = -.27, p < .01$), and participants who had more clicks also took more time ($r = .23, p < .01$). Participants who reported having experience with automatic text completion on any kind of device, showed greater experience with iTap knowledge ($r = .37, p < .01$) and greater experience with cellular phone knowledge ($r = .19, p < .01$). Those participants also had more clicks on average ($r = .11, p < .05$) but completed the spellings in less time ($r = -.11, p < .05$). Table 10 displays all of the results of the correlation matrix.

Table 10:
Correlation Matrix of Auto-completion Experience, iTap Experience, System Usability, Cellular Phone Knowledge, iTap Knowledge, Clicks on Average, and Time on Average.

	Auto- Completion Experience	iTap Experience	System Usability	Cellular Phone Knowledge	iTap Knowledge	Clicks on Average	Time on Average
Auto-Completion Experience	-						
iTap Experience	.37**	-					
System Usability	.05	.07	-				
Cellular Phone Knowledge	.19**	.09	-.03	-			
iTap Knowledge	-.05	.01	.01	.12*	-		
Clicks on Average	.11*	-.01	-.24**	.00	-.27**	-	
Time on Average	-.11*	-.06	-.18**	.04	-.10	.23**	-

** Correlation is significant at the .01 level (2-tailed)

* Correlation is significant at the .05 level (2-tailed)

Correlation of Measures by Each Factor

Initially, the data collected is first used to calculate the NE Ratio for clicks. The NE Ratio is calculated by taking the novice performance of a task and dividing it by the expert performance of the same task (Urokohara et al., 2000):

$$NERatio = \frac{Novice(TN)}{Expert(TE)}$$

Novice task performance is calculated by adding the total number of clicks for the spelling of each word. Expert performance is calculated by using CogTool to count the minimum number of clicks required to spell each word correctly. The calculation for the NE ratio performed in this study is not part of the original NE ratio study. This was applied to all words except for the first word *phone*. The results generated from this word were inconsistent with the remaining results. This is believed to be caused by the participants not understanding how to use the interface. The NE ratio for time was calculated in a similar way. The participant's time to spell each word was calculated and used as the novice variable. The minimum amount of time to complete the spelling of each word was calculated using the fastest time from each word with the least number of clicks. The results of the NE ratio for clicks showed significant variation in system usability ($r = -.24$) and iTap knowledge ($r = -.27$). This indicates that participants who performed better and were more satisfied with the usability of the interface used fewer clicks in spelling each word. Both values were statistically significant. System usability was also statistically significant ($r = -.18, p < .01$) and auto-completion experience ($r = -$

.11, $p < .05$), indicating similar results for the amount of time required to spell each word. Complete results can be found in Table 11.

The calculation of the log measure shows that people who were satisfied with the usability of the system made fewer clicks ($r = -.22$, $p < .01$) and required less time to spell each word ($r = -.15$, $p < .01$). Participants who performed well in the iTap knowledge factor also had fewer clicks ($r = -.27$, $p < .01$).

The median measure was calculated to find the number that evenly separates the number of clicks per word and the amount of time taken per word. This number was then correlated with each of the factors. For clicks, only iTap knowledge showed a significant correlation ($r = -.26$). Time showed a significant correlation for auto-completion experience ($r = -.13$) and system usability ($r = -.14$).

A reciprocal measure was also calculated for the number of clicks and the amount of time taken for each word. Being the reciprocal, it showed similar results for clicks in system usability and iTap knowledge as the other measures, $r = .20$ and $r = .28$, respectively. These were both shown to be statistically significant. It also showed similar results in time with the auto-completion experience and system usability factors, $r = .15$, $p < .01$ and $r = .14$, $p < .05$, respectively.

The D measurement is calculated to assess the equal-weight average of the minimum and maximum amount of clicks and the completion time. This calculation is expressed as follows:

$$D = \frac{\sum M}{\sum SD}$$

These results show a high correlation between the number of clicks and the system usability factor ($r = .18, p < .01$). It also points to a correlation between the auto-completion experience and iTap knowledge factors, $r = -.12$ and $.11$, respectively ($p < .05$). There was less correlation with the amount of time with system usability having $r = .13$. None of the other factors showed a statistical significance towards the number of clicks or the amount of time.

The final measure, Cohen's d , is used to show the variance between the factors and the number of clicks and the amount of time. The effect size indicates a strong correlation between the number of clicks and the iTap knowledge factor, $r = .52$, and system usability factor, $r = .37$. Auto-completion experience was negatively correlated with the Cohen's d for clicks, $r = -.16$. While it appears that more experienced users had better knowledge of the system, the data shows that participants who had more auto-completion experience may use more clicks to accomplish the task. Each of these factors was statistically significant. In relation to time, each of the factors showed a significant correlation ($p < .01$) except for cellular phone knowledge. System usability had an effect size of $r = .25$, iTap knowledge was $r = .23$, auto-completion experience was $r = .19$, and iTap experience showed an effect size of $r = .17$.

Table 11:

Correlation of Auto-Completion Experience, iTap Experience, Usability, Cellular Phone Knowledge, iTap Knowledge, NE Ratio, Log, Median, Reciprocal, *D*, and Cohen's *d*

	Clicks					
	NE Ratio	Log	Median	Reciprocal	<i>D</i>	Cohen's <i>d</i>
Auto-Completion Experience	.11	.10	.05	-.10	-.12*	-.16**
iTap Experience	-.01	-.02	-.07	.02	-.10	.06
System Usability	-.24**	-.22**	-.10	.20**	.18**	.37**
Cellular Phone Knowledge	.01	.02	-.03	-.02	.02	-.05
iTap Knowledge	-.27**	-.27**	-.26**	.28**	.11*	.52**
** $p < .01$ (two-tailed)						
* $p < .05$ (two-tailed)						
	Time					
	NE Ratio	Log	Median	Reciprocal	<i>D</i>	Cohen's <i>d</i>
Auto-Completion Experience	-.11*	-.14*	-.13*	.15**	-.05	.19**
iTap Experience	-.06	-.06	-.09	.07	-.02	.17**
System Usability	-.18**	-.15**	-.14*	.14*	.13*	.25**
Cellular Phone Knowledge	.04	.03	.03	-.04	-.02	.00
iTap Knowledge	-.10	-.08	-.06	.06	.10	.23**
** $p < .01$ (two-tailed)						
* $p < .05$ (two-tailed)						

Results Summary

The Correlation among self-reported system usability, the number of clicks, and time were stronger than any of the other factors. The participants that were more satisfied with the usability of the system generally performed better with fewer clicks and less time. Demographically, older participants had fewer clicks but required more time to finish the spelling than the younger participants. When comparing each of the factors against the other, the participants who were more satisfied with the interface had fewer clicks and were quicker at spelling each of the words. Using the correlational table on

each factor, there was a significantly greater correlation with the Cohen's d measures than with the NE ratio. In analyzing Cohen's d , it had a stronger correlation for comparing clicks and time with each of the factors than any of the other measures.

CHAPTER 5. DISCUSSION

Interface Design

The purpose of this study is to determine which objective measure best captures the concept of usability by correlating several candidate measures with other usability criteria, such as subjective ratings of usability (system usability) and declarative knowledge about the interface (iTap knowledge). However, for the results to generalize to other interfaces, there must be sufficient variance in the familiarity and difficulty of various aspects of the interface. If all aspects of the interface functioned with equal predictability, the candidate measure determined to be best for this interface might not be best for another interface. To obtain data with sufficient variance to be representative of the interactions and experiences of novice users, a cellular phone interface that incorporated both familiar and counterintuitive features was selected. The interface was based on iTap, a commercial failure. It was designed in such a way that users could relate to what they were seeing on the screen but the results of their actions often ran counter to their expectations. This interface was not intended to be usable or easy to master.

Factor Analysis

There was a high correlation among the questions on the self-reported system usability questionnaire, and these questions loaded on a single factor. System usability showed that participants who were more satisfied using the interface completed words with fewer clicks and in less time. Almost all of the questions related to system usability

correlated with high statistical significance with both the number of clicks and completion time. High ratings of system usability indicated that participants were able to enter a word quickly with this interface, the interface was simple to use, that they felt comfortable using the interface, and that they were satisfied with how easy it was to use. This indicates that the users were able to quickly understand the functional requirements of the interface.

The iTap knowledge factor was also correlated with the number of clicks and completion time with high statistical significance. This factor was generated from the post-study iTap knowledge questionnaire that tested the participant's declarative knowledge concerning the interface. The iTap knowledge factor was not as strongly correlated with the number of clicks and completion time as the system usability factor; however, people who correctly answered the questions showed a statistically significant decrease in the number clicks used to spell each of the words. This did not occur with respect to completion time. These participants seemed to put more thought into selecting their next click, which led to the increase in completion time. The negative correlation between the iTap knowledge factor and number of clicks, which is an indicator of error rate, seem to reiterate the results for the system usability factor. Those participants who were able to answer correctly the iTap knowledge questions also made fewer mistakes.

The participants who rated themselves as having more experience with text completion on cellular phones or on any kind of device required more clicks to complete a word but completed it in less time than those who did not rate themselves as highly. Auto-Completion experience made them faster than others, but it also made them more prone to error. In all likelihood, this is because the interface violated their well-learned

expectations. These users seemed to put less thought into selecting the correct button, perhaps because they assumed that they knew what they were doing. When making mistakes, they used the delete button to select a different choice. Given their higher self-assessed rating, it appears that these users attempted to apply what they already knew about automatic text completion devices to this interface. The data, however, indicate they were not as successful, because of the increase in the number of clicks.

The iTap experience factor and cellular phone knowledge factor did not have a statistically significant correlation with the number of clicks and completion time. These factors mainly focused on the participants' prior experience with iTap functionality and their ability to interact with the cellular phone interface.

Participant Performance by Measure

Testing the performance of various measures against the Novice-Expert ratio Method was broken out by a variety of factors. These factors were generated from the various questionnaires and tests associated with this study. The analysis of these factors ultimately showed that using the Cohen's d measure to assess the usability of an interface by objective performance measurements had higher validity than the original NE ratio as assessed by self-reported system usability and declarative knowledge of the interface (iTap knowledge). Table 11 shows that each statistically significant factor has a stronger correlation with Cohen's d than with the NE Ratio measure or any of the proposed alternative measures. These alternatives were log, median, reciprocal, and D . With respect to clicks, Cohen's d was highly correlated with the system usability factor ($r =$

.37) and iTap knowledge factor ($r = .52$). With respect to completion time, Cohen's d also demonstrated high correlations with the systems usability factor ($r = .25$) and iTap knowledge ($r = .23$).

Implication of Results

These results indicate that an alternative measure exists that outperforms the Novice-Expert ratio Method in estimating system usability based on novice and expert performance (e.g., number of clicks and completion time). This study determined that by applying Cohen's d , researchers can identify possible usability gaps in interface design with higher validity.

CHAPTER 6. CONCLUSION

Limitations

The cellular phone interface was designed to attempt to provide a relatively unique experience to gather novice data. However, participants' past experiences with similar interfaces and technologies may have affected the results. Additionally, the sampling of participants may contribute to some of these results as well. Most of the participants were between the ages of 18 and 22. Of the 337 total participants, 228 or roughly 68% fell within this age range. As younger people tend to be more technologically savvy than older people, this may have had an effect on the overall results of the study. Other effects with regards to the participants' demographic data were not deemed a factor.

Future Research

This study specifically looked at the possibility of enhancing the Novice-Expert ratio Method through the use of other measures. Repeating the analysis with other interfaces could help to further validate the choice of Cohen's d as an effective usability measure. Additional research could be performed into further pinpointing the exact location of a usability gap. This can be done with factor analysis that is more detailed into the interaction of the system being studied.

An additional issue concerns how to evaluate the performance of users who are unable to complete a task. For example, if 20 novices are assigned to interface A and 20

to interface B, and all the novices assigned to interface A are able to complete a task faster than all those assigned to interface B, except one of the users of interface A who does not complete the task at all, how can the usability of the two interfaces be compared?

Summary

The purpose of this study was to test various scales against the original ratio scale of the Novice-Expert ratio Method. The data collected from the cellular phone interface were analyzed using factor analysis and were correlated against the number of clicks and word completion time. This analysis ultimately showed the Cohen's d measure offered a better objective measure of system usability than the ratio scale and other candidate scales.

REFERENCES

- Card, S.K., Moran, T.P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carroll, J.M. & Rosson, M.B. (1987). Paradox of the active user. In J. M. Carroll (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. Cambridge, Mass.: Bradford Books/MIT Press.
- Desurvire, H., Lawrence, D., & Atwood, M. (1991). Empiricism versus judgement: Comparing user interface evaluation methods on a new telephone-based interface. *SIGCHI Bulletin*, 23(4), 58-59.
- Dillon, A., & Song, M. (1997). An empirical comparison of the usability for novice and expert searchers of a textual and a graphical interface to an art-resource database. *Journal of Digital Information* 1(1).
- Fischer, G. (2001). User modeling in human-computer interaction. *User Modeling and User Adapted Interaction*, 11, 65-86.
- Goonetilleke, R.S., Shih, H.M., On, H.K., Fritsch, J. (2001). Effects of training and representational characteristics in icon design. *International Journal of Human-Computer Studies* 55(5), 741-760.
- Gould J. & Lewis C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM*, 28(3), 300-311.
- Greenwald, A. G., McGhee, D. E., & Schwartz, L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- John, B. E. & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction*, 3(4), 320-351.
- Kjeldskov, J., Skov, M. B., & Stage, J. (2005). Does time heal? A longitudinal study of usability. In Balbo, S. & Bentley, T. (Eds.) *Proceedings of the Australian Computer-Human Interaction Conference 2005 (CHI '05)*, Canberra, Australia, ACM Press, 10 pages.

- Lewis, J. R. (1995) IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7:1, 57-78.
- Luo, L. & John, B. (2005). Predicting task execution time on handheld devices using the keystroke-level model. *Extended Abstracts of CHI 2005* (Portland, OR, April 2-7, 2005). ACM, New York.
- Molich, R. & Nielsen, J. (1990). Heuristic evaluation of user interfaces. *Proceedings of the Computer-Human Interaction Conference 1990* (pp. 249-256).
- Molich, R. & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM*, 33(3), 338-348.
- Nielsen, J. (1992). Finding usability problems through heuristic evaluation. *Proceedings of the Computer-Human Interaction Conference 1992*, 373-380.
- Nielsen, J. (1994). *Usability Engineering*. San Francisco: Morgan Kaufmann.
- Uehling, D.L., & Wolf, K., (1995). User action graphing effort (UsAGE). *Paper Presented at the Conference on Human Factors in Computing Systems*, Denver, Colorado.
- Urokohara, H., Tanaka, K., Furuta, K., Honda, M., & Kurosu, M. (2000). NEM: “Novice expert ratio method” A usability evaluation method to generate a new performance measure. *Paper presented at the CHI '00 extended abstracts on Human factors in computing systems*. The Hague, Netherlands.

Appendix A: Consent Form

IUPUI and Clarian Study Information Sheet

Testing and Enhancing the Novice-Expert ratio Method

You are invited to participate in the research study described in this document. Please read this document and, if you want to participate, print it for your records. When you have finished reading this document, please click the button "I Consent" to continue to the study or "I Do Not Consent" to exit the study.

Purpose: The purpose of this study is to assist researchers in the testing and potential enhancement of the Novice-Expert ratio Method.

Benefits: The data gathered from this study may contribute to the enhancement of the Novice-Expert ratio Method, which is used to find usability problems in a specified task.

Procedure: You will be using a cell phone-based, text interface to spell 10 words. The time required to do this will be recorded.

Time required: About 15 minutes

Participation: Participation is voluntary. You may refuse to participate at any time. No disadvantage will arise from refusing. Incomplete results are retained.

Age restriction: You must be at least 18 years old to participate.

Compensation: You will not be paid for participating.

Confidentiality: Your personal information will not be identified or shared or used for another purpose. Reported results will not contain information that may be used to identify you.

Risks: While we do not anticipate any risks from participating, you must stop participating and notify the principle investigator if at any time you feel your mental or physical well-being, personal values, or dignity is being harmed.

Dissemination of results: Results may be reported in talks, documents, and publications of the principal investigator, experimenter, and their co-authors.

Questions: If you have any questions or concerns about the study, feel free to contact the principal investigator, Prof. Karl F. MacDorman. If you have any questions about your rights as a research participant, or unresolved problems, complaints, or concerns about a study, contact the IUPUI/Clarian Research Compliance Administration. Contact details are provided on the contact webpage.

By clicking “I Consent” you indicate the following:

I am at least 18 years old; I understand and agree to the following conditions; my questions have been answered satisfactorily; and I have printed a copy of this study information sheet for my records.

I consent. I do not consent.

Contact

Do you have a technical problem with using this website or participating in an experiment?

Email the web developer:

Tim Whalen

Email:

Tel: (317) 371-8203

Do you have any other questions or comments about a study at this website?

Email or phone the principal investigator:

Karl F. MacDorman, Ph.D. (Cambridge)

Associate Professor, Indiana University School of Informatics

Adjunct Professor, Purdue University School of Engineering and Technology

IT 487, 535 West Michigan Street

Indianapolis, Indiana 46202

Email:

Tel: (317) 215-7040

Fax: (206) 350-6089

Do you have any questions about your rights as a research participant, or unresolved problems, complaints, or concerns about a study?

Email or phone the Institutional Review Board:

IUPUI/Clarian Research Compliance Administration

620 Union Drive, UN 618

Indianapolis, IN 46202-5167 USA

Email:

Tel: +1 317 278-3458

Appendix B: Demographic Information and Pre-Screening Questions

IMPORTANT:

This study will take roughly 15 minutes to complete. While you are participating in this study, please select a time and location that will be free from distractions and interruptions. You should also close any other programs that may distract you during the study.

Please tell us a little about yourself:

All fields are required

Gender:

- Female
 Male

Age:

Education:

Country of Birth:

Country of Residence:

1. How many times have you already participated in this study?

- 0 1 2 3 4 5+

2. How would you rate your expertise using automatic text completion technology on any kind of device?

- No Experience Beginner Competent Proficient Expert

3. In the past week how many times have you used automatic text completion on any kind of device (e.g., computer, including web address and search term completion, cell phone, PDA)?

- 0 1 2 3 4 5+

4. How would you rate your expertise using automatic text completion technology on a cellular phone?

No Experience Beginner Competent Proficient Expert

5. In the past year how many times have you used automatic text completion on a cellular phone?

0 1 2 3 4 5+

6. How would you rate your expertise using the iTap interface on a cellular phone?

No Experience Beginner Competent Proficient Expert

7. How many times have you ever used the iTap interface on a cellular phone?

0 1 2 3 4 5+

Appendix C: Post-Study Questionnaire Data

Please rate the usability of the system.

1. What type of internet connection are you currently using?

- ADSL
- Cable
- DSL
- Satellite
- Dial-up

2. Overall, I am satisfied with how easy it is to use this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree
- Neutral
- Slightly Agree
- Moderately Agree
- Strongly Agree

3. It was simple to use this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree
- Neutral
- Slightly Agree
- Moderately Agree
- Strongly Agree

4. I can effectively enter a word using this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree

- Neutral
- Slightly Agree
- Moderately Agree
- Strongly Agree

5. I am able to enter a word quickly using this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree
- Neutral
- Slightly Agree
- Moderately Agree
- Strongly Agree

6. I am able to enter a word efficiently using this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree
- Neutral
- Slightly Agree
- Moderately Agree
- Strongly Agree

7. I feel comfortable using this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree
- Neutral
- Slightly Agree

- Moderately Agree
- Strongly Agree

8. It was easy to learn to use this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree
- Neutral
- Slightly Agree
- Moderately Agree
- Strongly Agree

9. I believe I became productive quickly using this interface.

- Strongly Disagree
- Moderately Disagree
- Slightly Disagree
- Neutral
- Slightly Agree
- Moderately Agree
- Strongly Agree

Please answer the following questions quickly. Do not worry about the accuracy of your answers. Just select the first answer that comes to mind.

1. What number do you have to press to enter the letter **n**?

- 3 4 5 6 7 Don't know

2. If the next letter is **w** and you press the button **9 wxyz**, do you need to select **w** before entering the next letter?

- Yes No Don't know

3. If the next letter is **b** and you press the button **2 abc**, do you need to select **b** before entering the next letter?

- Yes No Don't know

4. If you are typing **phone** and you have already typed **phm**, only one choice appears: **pho**, do you still need to select **pho** before entering the next letter?

- Yes No Don't know

5. What requires more mouse clicks? Typing **add** or **bad**?

- add bad They both take the same number of steps Don't know

6. How many mouse clicks are required to type **aaa** and submit it?

- 3 4 5 6 7 Don't know

7. How many mouse clicks are required to type **cab** and submit it?

- 3 4 5 6 7 Don't know

8. Is the delete button labeled **<**?

- Yes No Don't know

VITA

Timothy Joseph Whalen

[tjwhalen\(@\)iupui.edu](mailto:tjwhalen(@)iupui.edu)

Education

- | | | |
|-------------|------------------------------------------------------------------------------------------|------|
| M.S. | Graduate School of Informatics, Indiana University-
Purdue University at Indianapolis | 2008 |
| B.A. | Liberal Studies, University of Indianapolis | 1997 |

Professional

- 1/2007 - Present, Greenlight BTS, Indianapolis, IN
Programmer
- 4/2001 - 12/2006, Baker & Daniels, Indianapolis, IN
Database Analyst
- 6/2000 - 4/2001, Olive, LLP, Indianapolis, IN
Database Administrator/Developer
- 1/1999 – 6/2000, Management Computer Systems, Inc., Indianapolis, IN
Programmer/Analyst
- 2/1997 - 12/1998, Budget Car Sales, Inc., Indianapolis, IN
MIS Financial Specialist
- 6/1996 – 1/1997, Avantec, Inc., Carmel, IN
Database Specialist