Beyond Usability: An Alternative Usability Evaluation Method, PUT-Q2

Troy Donald Abel
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd
Part of the Art and Design Commons

Recommended Citation
http://lib.dr.iastate.edu/etd/11359

This Dissertation is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Beyond usability: an alternative usability evaluation method, PUT-Q2

by

Troy Donald Abel

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTORATE OF PHILOSOPHY

Major: Human Computer Interaction

Program of Study Committee:
Roger Baer, Major Professor
Ann Thompson
H. S. Venkatagiri
Steve Herrnstadt
Debra Satterfield

Iowa State University
Ames, Iowa

Copyright © Troy Donald Abel, 2010. All rights reserved.
# TABLE OF CONTENTS

LIST OF FIGURES iv  
LIST OF TABLES v  
ACKNOWLEDGEMENTS vi  
ABSTRACT vii  

1. INTRODUCTION 1  
1.1 Contextual Background 1  
1.2 Statement of Problems and Investigation Questions 5  
1.3 Approach and Justification 6  

2. REVIEW OF LITERATURE 7  
2.1 Human Variables 7  
2.2 Foundations of Human Computer Interaction 30  
2.3 Psychology of HCI 33  
2.4 Usability evaluation methods (UEM’s) 43  
2.5 Qualitative data analysis in the social sciences 49  

3. METHODOLOGY 62  
3.1 Overview 62  
3.2 Sampling 63  
3.3 Data collection 63  
3.4 Analysis Model 67  

4. PILOT STUDY ANALYSIS 71  
4.1 Qualitative & Quantitative Analysis 71  
4.2 Data Visualizations 87  

5. CONCLUSIONS, SHORTCOMINGS, AND FUTURE RESEARCH 95  
5.1 General thoughts 100  
5.2 PUT-Q2: A new usability testing tool 103  
5.3 Conclusions & Future Implications 105  
5.4 Future research 105  

APPENDIX A: FACEREADER 107
LIST OF FIGURES

Figure 1: LeDoux model 11
Figure 2: Plutchik’s emotional wheel 14
Figure 3: Plutchik’s primary & secondary emotions 15
Figure 4: Author model of Cupchik’s emotional process 19
Figure 5: Csikszentmihalyi’s model of Flow 21
Figure 6: Card and Moran’s Human Computer Model 36
Figure 7: 5-second stimuli including AOI’s 74
Figure 8: Grid dwell times 75
Figure 9: Scan path 76
Figure 10: Atlas.ti software screenshot 86
Figure 11: Atlas.ti software screenshot 87
Figure 12: PUT-Q2 Subject Matrix (Side 1) 90
Figure 13: PUT-Q2 Subject Matrix (Side 2) 91
Figure 14: Sample Population Matrix (Side 1) 92
Figure 15: Sample Population Matrix (Side 2) 93
Figure 16: Sample size resulting in 5 subjects 97
Figure 17: Sample size resulting in 37 subjects 98
Figure 18: Population vs. n=5 Random sample power test (Task G) 99
Figure 19: Population vs. n=5 Random sample power test (Task H) 100
Figure 20: Sample PUT-Q2 matrix #1 103
Figure 21: Sample PUT-Q2 matrix #2 104
Figure 22: Progression of overall research agenda 106
**LIST OF TABLES**

Table 1: Data collected by tool .................................................. 64

Table 2: Task descriptions and total errors by task ....................... 77

Table 3: Tasks by cardinal sort order ............................................ 78

Table 4: Mean task times as correlated to cardinal sort order .......... 79

Table 5: Descriptive statistics and test subject task times, sorted by task 84
ACKNOWLEDGEMENTS

Recognize your potential and strive to achieve it— for anything is possible, even a PhD.

My journey towards a doctorate would not have been achieved without the support of several individuals.

I must recognize the most influential educator in my life, Professor Roger Baer. With his tutelage I was able to realize my life-long goal of attaining a terminal degree in higher education. Furthermore he has played a pivotal role in my desire to become an educator, researcher and innovator. Moreover, Professor Baer has taught me what it means to be a true mentor.

Without the unconditional love and support from my family, my achievements would not have been realized and I only hope to have made them proud. My mother was always ready to motivate, while my father followed through with words of support. My brothers and sister-in-law provided copious amounts of encouragement, while my niece Nina provided comic relief.
ABSTRACT

Usability can be thought of as a measure or degree to which a system satisfies the needs of the human. Usability is a quality inherent to any given system, which assists in determining the efficiency, effectiveness and satisfaction levels of those involved in the interaction. Everyday we are bombarded with interactions and experiences that shape our thoughts, values, and judgments as well as test our limits of interaction with technology. These interactions have progressed at such an intense pace that humans have become practically slaves to technological innovation. Humans are forced to conform with needs of technology, rather then technology conforming to human needs. This fact must be rectified and becomes the primary focus of this thesis.

Current models in usability evaluation methods (UEMs) analyze the quantitative data collected during testing. These statistical studies provide insight into limited aspects of usability, and most overlook human dimensions, including perception and affective responses; thus leaving a glaring pitfall in the overall analysis of system usability. By analyzing a new qualitative channel of data, this research attempts to explain these human-dimensional factors. Up to this point no evaluation model has been largely accepted which attempts to fuse both qualitative and quantitative data.

This research proposes an alternative UEM, incorporating both qualitative and quantitative data, called the Perception and Usability Testing combining Qualitative and Quantitative data, or PUT-Q2. This new usability evaluation method presents complex qualitative and quantitative data in graphical visualizations and matrices that assist the usability expert in uncovering additional correlations and usability issues with their system.
1. INTRODUCTION

1.1 Contextual background

“Usability is about human behavior. It recognizes that humans are lazy, get emotional, are not interested in putting a lot of effort into, say, getting a credit card and generally prefer things that are easy to do vs. those that are hard to do.”

David McQuillen
“Taking Usability Offline” Darwin Magazine, June 2003

Usability can be thought of as a measure or degree to which a system satisfies the needs of the human. Usability is a quality inherent to any given system, which assists in determining the efficiency, effectiveness and satisfaction levels of those involved in the interaction. Everyday we are bombarded with interactions and experiences that shape our thoughts, values, and judgments as well as test our limits of interaction with technology. For example, once time-consuming research tasks can be accomplished in minutes on the Internet. Moreover, with the click of a mouse button, the contents of a book can be delivered straight to your door or even downloaded to your hand-held device without the need for wires or connections. We beam contact information from one Blackberry to another and send short text messages via our mobile phones much like the instant messages on the PC of the past.

This interaction with technology has progressed at such an intense pace that humans have become practically slaves to technological innovation. Humans are forced to conform with needs of technology rather than technology conforming to human needs. This fact must be rectified and becomes the primary focus of this thesis.
The advent of the industrial age brought with it a continuous stream of products utilizing technology. Throughout the industrial age (and now) technology has infiltrated our lives, moving from simply being a tool of production to integral parts of everyday life. (Thackara, 2006) Along with this “infiltration” comes the need to study these technological interactions, amongst humans and technology, in hopes of producing more positive experiences. The relationship between human and technology represents an area of explosive research with new disciplines being shaped such as experience design, and human computer interaction. The latter an emerging field concerns itself with the specific experience of human and technology.

Human Computer Interaction is a fledgling field just beginning to form a solid foundation of research and knowledge. Like many disciplines HCI draws from areas such as psychology, visual communication design, statistics, sociology, and anthropology. When these areas diverge they represent a unique opportunity for researchers to expand the knowledge base of HCI and enrich the discipline.

Several cannons of HCI have been declared in the literature. Some include: ubiquitous computing, user centered design and usability evaluation. Of these cannons, usability evaluation methods (hereinto referred to as UEM), also commonly referred to as simply “usability testing,” (hereinto referred to as UT) emerges as one of the most prolific cannons currently having published research in HCI today.

We have overlooked the needs of the human, instead concentrating on the newest technology and how it could help streamline our everyday lives. This misdirection has

* Please note: Usability Evaluation Methods (Methodologies) UEM’s, and Usability Testing (UT) may be used interchange Debbie throughout this publication.
caused a shift away from human-centered design, to technology-centered design. Humans have been trained to conform and adapt to the ever-changing face of technology. (Thackara, 2006) We must rally to shift the design focus from technology for its own sake to technology designed to serve human experience. How users interpret and participate with the experience of technology cannot be separated from how humans decipher that experiential action itself. (McCarthy & Wright, 2007)

For it is this resulting experience that has the potential to be shaped and modified by the designer to provide the catalyst for positive social change and interaction among the targeted audience members and the segment of society they represent. Designers can’t control the experiences individuals will have with the system; however, “designers can, and do, work with concern for the quality of people’s experiences relating to individual products and to systems of things. In this way, designers have the ability to influence positively both the beauty of these interactions and…affect people’s perceptions of the company offering them.” (McDonagh et al., 2003) Additionally, “Technology should bring more to our lives then the improved performance of tasks: it should add richness and enjoyment. A good way to bring fun and enjoyment to our lives is to trust in the skill of artists. Fortunately, there are many around.” (Norman, 2005) The artifacts of everyday life must be considered a supporting tool in the symbiotic relationship demonstrated in the system of human-computer interaction, and not the primary. (Thackara, 2006) Designers are beginning to realize that emotions play an important role in interaction with a system. “We cognitive scientists now understand that emotion is a necessary part
of life affecting how you feel, how you behave, and how you think…Emotions reflect our personal experiences, associations, and memories.” (Norman, 2005)

In an attempt to evaluate these interactions, or experiences, current practitioners’ in HCI utilize usability evaluation methods. These UEMs rely heavily on quantitative statistical analysis to predict the success or failure of a system interaction. However, these tests fail to represent the human aspects of HCI.
1.2 Statement of Problems & Pilot Investigative Questions

Problems

1. Current usability testing methods (UEMs) do not currently incorporate and evaluate the human dimension (interaction/reaction) beyond quantitative data in an attempt to truly uncover the interactive experience.

2. Traditionally UEMs are conducted on a relatively small sample size (n=5) based on published research that denounced the need for a larger sample sizes. Statistically speaking, this sample size appears to be too small.

Investigation Questions

1. Can a qualitative dimension of human interaction be incorporated into traditional UEM’s to assist in uncovering otherwise overlooked aspects of usability testing?

2. By sampling a large-scale usability study can the current UEM standard sample size (n=5) be affirmed as an acceptable number?

3. Does a correlation exist between perceived difficulty and actually difficulty of usability task completed by test subjects?
1.3 Approach and Justification

As stated, the current UEMs concern themselves with statistical measures which include such measures as: time on task, number of errors, number of mouse clicks, and number of times a subject asks for assistance. While these numbers are important and reveal shortcomings of a site's design, they lack the richness of narrative that can be revealed by utilizing a true mixed method approach combining both quantitative and qualitative research.

This investigation is primarily an exploratory study researching the use of both quantitative and qualitative data in large-scale usability testing in an effort at producing a new conceptual model and framework for usability testing methods. Secondly, this research aims to affirm or reject the currently accepted sample size (n=5) of traditional UEMs.

This research is important because to date, no published UEM incorporates both quantitative and qualitative data when analyzing the overall usability of an e-commerce website. By constructing a new usability testing methodology we hope to uncover innovative insight into perceived-anticipated usability as well as fuse qualitative data with common descriptive statistics to reveal a new overall usability score.
2. REVIEW OF THE LITERATURE

The interdisciplinary and fledgling nature of Human Computer Interaction requires a diverse review of literature. Five important areas have been identified for inclusion in this dissertation. Human Variables, including emotion and cognition, play important roles in user satisfaction and overall affective impact of system interaction. Foundations in HCI have been set forth in the literature and assist in forming a solid base on which to present further research. A realm of psychology concentrating on the factors of human computer interaction has emerged and plays a pivotal role in understand the interaction between human and technology. Usability evaluation methods provide varying strategies to analyze the success of a system interaction. Finally, qualitative data assists in providing insight into the human experience during an interaction with technology.

2.1 Human Variables

Human Emotions

Emotions are fundamental in enriching any system interaction. (Brave & Nass, 2003) In the past, systems were developed aesthetically and without regard or response to the emotional influence they possessed. (Papanek, 1985) Furthermore, system designers concluded that interactions with technology, particularly computers, wereunemotional and sterile. (Brave & Nass, 2003) However recent design philosophers, psychologists, neuroscientists, and scholars have suggested that emotion plays an integral role in our interactions with technology including computers and the interfaces designed to interact via this medium.
As designers of interfaces and interactive systems, we must recognize and centralize the emotional-volitional nature of any system. (McCarthy & Wright, 2007) Designers must understand they do not “design emotions,” but rather they design for the optimum experience that results from personal interaction with the objects experienced in everyday life. (Sanders, 2001) Designers lay the foundation and create the scaffolding that their target audience will use to create their personal emotional experiences. (Sanders, 2001) Emotion is no longer regarded as the occasional outburst directed at the computer screen or the frustration observed when attempting to decipher a cryptic error message. It is now understood that a wide range of emotions play seminal roles in practically every goal-oriented activity. (Brave & Nass, 2003) Furthermore, emotion is often aligned with behavioral design. According to Donald Norman, behavioral design is about look and feel -- the total experience of using a product. Norman goes on to say, “Reflection is about one’s thoughts afterwards, how it makes one feel, the image it portrays, and the message it extends to others regarding individual experience.” (Norman, 2005) “Many psychologists now argue that it is impossible for a person to have a thought or perform an action without engaging, at least unconsciously, his or her emotional system.” (Brave & Nass, 2003) Moreover, people are like a phenomenologist -- they take their inner feelings much more seriously than the forces of outside occurrences exhorted over them. (Csikszentmihalyi, 1998)

The influence emotion plays on any one individual can vary greatly. People bring previous experiences, values, and prejudices to each interaction they face. “Emotions are in some respect the most subjective elements of consciousness, since it is only the person
himself or herself who can tell whether he or she truly experiences love, shame, gratitude, or happiness. Yet an emotion is also the most objective content of the mind, because the ‘gut feeling’ we experience when we are in love, or ashamed, or scared, or happy, is generally more real to us than what we observe in the world outside.” (Csikszentmihalyi, 1998) No matter what we are experiencing, or doing, how we value the experience is even more important. (Csikszentmihalyi, 1998)

A discussion revolving around emotion would not be complete without highlighting the semantic differences between emotion and feelings. According to Anthony Damasio, in his book, Looking for Spinoza: Joy, Sorrow, and the Feeling Brain, emotions precede feelings. “Feelings [are] mostly shadows of the external manner of emotions,” says Damasio. (Damasio, 2003) Damasio goes on to explain that a distinct difference exists between emotion and feeling. “It is true that the common usage of the word emotion tends to encompass the notion of feeling. But in our attempt to understand the complex chain of events that begins with emotion and ends up in feeling, we can be helped by a principled separation between the part of the process that is made public and the part that remains private… the former part emotion and the latter part feeling,” states Damasio. (Damasio, 2003) Interestingly enough, he provides an explanation regarding the ordering of emotions and feelings. According to Damasio, emotions withstood the test of evolution, “we have emotions first and feelings after because evolution came up with emotions first and feelings later. Emotions are built from simple reactions that easily promote the survival of an organism and thus could easily prevail in evolution.” (Damasio, 2003) We ultimately concern ourselves with the former- emotions, those
which are made public, “Emotions are actions that mostly occur in the public, visible to others as they occur in the face, in the voice, in specific behaviors. Feelings, on the other hand, are always hidden, like all mental images necessarily are, unseen to anyone other than their rightful owner, the most private property of the organism in whose brain they occur.” (Damasio, 2003)

**How are Emotions Created and Processed?**

Up to this point, emotions have been defined and viewed as a set of user reactions triggered when the individual interacts with a technological system. However, the source of all emotion and feeling is inside the complex human organ known as the brain. Damasio states, “Emotions play out in the theater of the body. Feelings play out in the theater of the mind.” (Damasio, 2003) Furthermore, we must understand the psycho-physiological systems that determine how emotion emerges from interactions with systems such as websites, before we can design, interpret, or attempt to solicit any behavior from our user. The literature provides a plethora of definitions for ‘emotions,’ however, according to Brave and Nass, “generally agreed-on aspects of emotion stand out: (1) emotion is a reaction to events deemed relevant to the needs, goals, or concerns of an individual; and (2) emotion encompasses physiological, affective, behavioral, and cognitive components.” (Brave & Nass, 2003) One model we can refer to in an attempt to understand the emotional system is based on a simplified view of LeDoux’s work in neuropsychology (Figure 1) (Brave & Nass, 2003; LeDoux, 2003)
Figure 1: A simplified model of LeDoux’s work as presented by Brave and Nass (2003).

The model diagrams an interaction involving three key regions of the brain: the thalamus, the limbic system, and the cortex. Sensory input enters the system via the thalamus and is then routed either to the cortex, for higher-level processing, or to the limbic system. These two pathways, the thalamic-limbic and the corticolimbic represent the brains traffic response system. (LeDoux, 2003) LeDoux refers to the limbic system as the “seat of emotion,” an area that constantly evaluates the needs and goals correlation of the incoming information. If input is determined to be relevant, the limbic system routes signals to the body and the cortex, while biasing the cognitive processes and attention. (Brave & Nass, 2003; LeDoux, 2003)

The thalamic-limbic passageway connecting the limbic system and thalamus is responsible for what Damasio refers to as primary emotions: innate aversions and attractions as well as fear. (Damasio, 2005) Designers should take note of these as they are triggered by sudden changes including: strong size contrasts, alarming error
messages, flashing text, and other startling objects of visual stimuli. (Brave & Nass, 2003)

The corticolimbic passageway is responsible for processing higher-level cognitive and emotional operations such as pride, satisfaction, depression, and frustration. An example would be deciding whether or not to delete a file, or recognizing the Apple logo. These emotions and actions require more cognitive processing (Brave & Nass, 2003)

Furthermore, emotions have a direct impact on the interaction between humans and technology. The cortex can trigger not only responses to external stimuli, but also to internally generated stimuli.

Robert Plutchik, in his book, *Emotion: A Psychoevolutionary Synthesis*, defines eight basic emotions and their additive derivate emotions. The basic emotions are of fear, surprise, sadness, disgust, anger, anticipation, joy and acceptance. This basic set is reiterated and slightly differentiated by Damiso when he states, “The primary (or basic) emotions are easier to define because there is an established tradition of lumping certain prominent emotions in this group. The frequent listing includes fear, anger, disgust, surprise, sadness and happiness, the emotions that first come to mind whenever the term, ‘emotions,’ is invoked.” (Plutchik, 1980; Damasio, 2003) Furthermore these emotions cut across lines of gender and culture and have been noted in human and non-human species alike. To validate the completeness of Plutchik’s emotions, we look to Anthony Ortony who compiled a list of emotions from a vast pool of research and published his findings in the seminal article, “What’s basic about basic emotion?” which appeared in the July 1990 issue of *Psychological Review*. In his article, Ortony gathered and compiled the
various emotions detailed by a variety of psychologists, theorists and philosophers in an attempt to answer his research that questioned the validity of a common set of basic emotions. Ortony states, “A widespread assumption in theories of emotion is that there exists a small set of basic emotions. From a biological perspective, this idea is manifested in the belief that there might be neurophysiological and anatomical substrates corresponding to the basic emotions.” (Ortony & Turner, 1990)

For the sake of discussion we will use Plutchik’s model as it provides an opportunity to analyze and create blended emotions, as well as primary emotions. By utilizing the circular diagram of basic emotions we see they are located in very precise locations and relationships to each other. Plutchik defends this positioning when he states, “The eventual decision for the optimal ordering [in any model] will also depend upon the kind of internal consistency and research implications provided by one grouping rather than another. Based on his research, Plutchik sequences the order of emotions on the circle: joy, acceptance, fear, surprise, sadness, disgust, anger, and anticipation.” (Plutchik, 1980, p. 156) “The center of the circle is used to represent the idea of conflict resulting from the mixtures of two or more emotions.” (Plutchik, 1980) (Figure 2)
Plutchik uses a circular diagram to represent the relationships between primary and derived emotions. He places the primary emotions at regular intervals on the wheel. Than he uses the primary, secondary and tertiary dyads of color mixing theory to describe several derived emotions. (Fig. 3) For example, by mixing adjacent emotions or primary dyads he concludes joy + acceptance = friendliness and fear + surprise = alarm.
Plutchik’s model also includes secondary dyads, mixing emotions, once removed, joy + fear = guilt and sadness + anger = sullenness, and finally tertiary dyads, mixing of emotions, twice removed joy + surprise = delight and anticipation + fear = anxiety. (Plutchik, 1980; Damasio, 2005)


![Figure 3: Primary and Secondary dyads based on R. Plutchik’s Emotional Wheel](image)

Plutchik describes his rationale in structuring the order of emotions as he did. In essence, Plutchik assembled a list of synonyms from the *Roget’s Thesaurus* for each of the varied dimensional emotional words. He then presented these to a group of college students to score them. Statistics were generated and the list of intensity determined the eight resulting emotional headers. He further goes on to detail his analytical and empirical studies that followed the creation of the emotional circle. These additional studies
validate not only the completeness of emotions, but also the authority of personality traits formed by the mixture of dyad emotions.

The distance between two emotions dictates the level of ‘blending’ and ‘mixing’ that could occur, thus diluting the primary emotions. (Damasio, 2003) The more distant the emotions, the less likely they are to mix. Some emotional theorists, including Antonio Damasio, believe that this mixing of adjacent and different dyadic emotions occurs only in humans, and produces the higher ordered derived emotions. (Damasio, 2005) The “higher level,” set of derived emotions is usually, “thought of as a cognitive operation.” (Damasio, 2005) The lower levels, or the basic eight emotions in our working model, are universally shared with lower species.

What influences emotion, can emotions be shaped?

Finite differences regarding semantics are a constant element of contention among scientists involved with human emotional research; however, all agree that emotions are a necessary part of the human cognitive system. According to Donald Norman, “Emotions are inseparable from and a necessary part of cognition.” (Norman, 2005) Norman continues, “Emotion is the conscious experience of affect, complete with attribution of its cause and identification of its object.” (Norman, 2005)

Individuals outside the realm of professional psychological research also echo this thesis. For example, Clotaire Rapaille in his book, Culture Code, discusses the relationship between experiences, emotion, and the culture in which these experiences occur. Rapaille describes several of, what he calls “cultural codes” along with their
associated “imprints.” The code is an attempt to unlock the conventions for personal interpretation. “An imprint and its code are like a lock and its combination. If you have all the right numbers in the right sequence, you can open the lock. Doing so over a vast array of imprints has profound implications. It brings us to the answer to one of our most fundamental questions: “Why do we act the way we do?” (Rapaille, 2007) The term “imprint,” was first introduced by Konrad Lorenz as described to us by Rapaille, “The combination of the experience and its accompanying emotion creates something known widely as an imprint. Once an imprint occurs, it strongly conditions our thought processes and shapes our future actions. Each imprint helps make us more of who we are. The combination of imprints defines us.” These imprints, Rapaille states, influence us on an unconscious level. (Rapaille, 2007)

In addition to Rapaille, Gerald Cupchik, a noted professor of psychology at the University of Toronto, contributes some interesting thoughts on the shaping and interaction of emotions and experiences. Cupchik studies the psychology of emotion and aesthetics. He argues that consciousness acts as the division between the interaction of physical and social worlds. He suggests that the various theories of emotion be subdivided into two contrasting groups. “Consciousness serves as a sentient boundary between stimulation from the external physical or social worlds and the internal bodily world. Emotions are a part of consciousness and reflect the complex interaction of mind and body.” Cupchik goes on to say, “While a unified theory of emotion remains elusive, the main theories can be divided into complimentary ‘action’ and ‘experience’ oriented groups. The action-oriented approaches to emotion associated with centralism,
behaviorism, and cognitivism focus on the adaptive and purposive mind. In contrast, experience oriented theories relate to peripheralism, psychodynamics, and phenomenology/existentialism, [that] encompass bodily reactions to social meanings.” (Cupchik, 2003) Furthermore, Cupchik explains his thoughts on the importance of emotions by saying, “One can argue that in action mode feelings are the shadows of cognition. When the pattern of ideas is coherent, then there is a feeling of calm or pleasure. When the ideas do not fit together harmoniously, there is the experience of tension.” (Cupchik, 2003) This notion of action groups illuminates the belief that experiences can be shaped. “Given that feelings reflect the state of cognition, so to speak, artists and designers can use their feelings as an index of the state of their projects.” (Cupchik, 2003)

Cupchik continues his discussion describing the contrasting realms of thought and emotional response to situations and experiences in terms of two processes. “Bottom-up processes are more characteristic of the experience mode in which the body is the focus and the mind serves as background or context. From this perspective, cognition serves as a context for emotions… feedback from bodily states and muscular memories lend coherence to the overall experience…” Cupchik believes that bottom-up and top-down processing relates to one another in much the same way that the designer thinks of the figure–ground relationship. “Top-down processes are typically of the action mode whereby the mind is the central figure dominating the body as ground.” In other words, “cognitions govern feelings.” (Cupchik, 2003)
With this in mind, Cupchik says, “In the sum the interaction of mind and body can be characterized in complementary processes depending on whether the figure/ground relation is mind over body or body over mind. When the mind is dominant, then the body functions in terms of [emotions], whereas when the body is dominant, it awakens the mind’s eye with memories and symbolically meaningful experiences.” (Cupchik, 2003)

For example, imagine how a smell can trigger memories or how a person can involuntarily drift into inattentiveness. These are prime examples of the body acting as the figure over the mind functioning as the ground.

The model in Figure 4 illustrates the reciprocal nature of Cupchik’s top-down and bottom-up processing theory. As the model suggests, both processes contribute to the overall rich user experience. Ultimately, the sum of all human emotional responses

![Figure 4: Author’s model based on Cupchik’s description of emotional processing](image-url)
constitutes what we know as experiences. John McCarthy states in his book, *Technology as Experience*, “By making the emotional-volitional nature of the act central to our account of experience, we focus on felt life as the concerns, fears, confusion, ambivalence, interests, desires, and expectations that permeate our sense making.” (McCarthy & Wright, 2007)

**How are emotions optimized (Flow)?**

According to Mihaly Csikszentmihalyi, experience can be described as a journey over the course of a given amount of time. (Csikszentmihalyi, 1998) Moreover, “To live means to experience—through doing, feeling, and thinking. Experience takes place in time, so time is the ultimate scarce resource we have.” (Csikszentmihalyi, 1998) For this reason, time becomes an essential factor in the interactive system, as well as the concept of optimized emotions, which Csikszentmihalyi calls ”flow.“

The relationship between emotions and human experience relates directly to the attention span of human interactions, which is consistent with Csikszentmihalyi’s concept of time as an essential factor. “Emotions refer to the internal states of consciousness. Negative emotions like sadness, fear, anxiety, or boredom produce ‘psychic entropy’ in the mind, that is, a state in which we cannot use attention effectively to deal with external tasks…” By extrapolation we can conclude the antithesis also be true. (Csikszentmihalyi, 1998)

Csikszentmihalyi explains that these experiences shape us forever, “Over the years, the content of experience will determine the quality of life. Therefore, one of the
most essential decisions any of us can make is about how one’s time is allocated or invested.” (Csikszentmihalyi, 1998)

Flow has been described as the point when the optimal levels of challenges (obstacles) and skills (personal) are met. “Flow tends to occur when a person’s skills are fully involved in overcoming a challenge that is just about manageable. Optimal experiences usually involve a fine balance between one’s ability to act, and the available opportunities to action.” (Csikszentmihalyi, 1998) As skill level rises challenges must
also rise otherwise the user will become bored or apathetic. The optimum level of flow occurs when the skills are high enough to balance the challenges presented. (Figure 5)

Furthermore, “human beings feel best in flow when they are fully involved in meeting a challenge, solving a problem, discovering something new. Most activities that produce flow also have clear goals, clear rules, immediate feedback, a set of external demands that focuses our attention and makes demands on our skills.” (Csikszentmihalyi, 1998) Donald Norman agrees: “In the flow state, you become so engrossed and captured by the activity being performed that it is as if you and the activity were one. You are in a trance where the world disappears from consciousness. Time stops. … Flow is a motivating, captivating, addictive state.” (Norman, 2005)

**Human behavior**

Philosophers such as Carl Jung, B.F. Skinner, and Abraham Maslow acknowledge that human behavior is driven by motivation, emotion, and affective factors. These understandings reiterate the importance that products and services must cater to the human aspect of the interaction. This human aspect must be recognized along with their associated human behaviors. “Companies have turned to design to differentiate their offerings through human-centered innovation and to create stronger emotional connections with their customers.” (McDonagh et al., 2003)

One example of a brand that has used principles of human behavior to shape its direction in the consumer market is the technology and communication company Apple, Incorporated. Apple’s iPod was the first of its kind. This innovation also came with a
hefty price tag one that consumers were apprehensive to accept. In order to be successful, Apple realized they needed to hone in on the entire experience of purchasing the iPod. Apple realized that consumers were much more emotionally driven to purchase a $350 music player on impulse when their entire emotional experience consisted of positive human behavior. From the design of the store to the packaging, Apple designed the ultimate consumer buying experience and made the iPod a highly successful personal entertainment and information system. Other corporations such as Braun and Philips have recognized that design adds a strategic advantage to their businesses. (McDonagh et al., 2003)

One of the most important aspects of human behavior is the complex processing required for the human brain to acquire, store, and process new information, as well as novel experiences. In addition, humans utilize selective perception; what is seen means exactly what the viewer wants it to mean. John Morgan, author of the popular visual communication textbook, *See What I Mean: An Introduction to Visual Communication* says it best, “We direct our attention mainly to those messages we know we will like: they suit our tastes, confirm our prejudices, or excite our indignation in ways which appeal to our self-esteem. However, selection does not end there. When we are presented with a complex message, we are likely to notice particularly those parts of it which confirm our previous attitudes.” (Morgan & Welton, 1992) Essentially humans thrive on pattern association. They become accustomed to seeing something presented in a certain manner and format, and become comfortable and proficient in their usage.
Designers must also understand that human behavior does not occur in an emotional or experiential vacuum. “Precious experience, prejudice, and the brain’s preference for simple explanations all color the interpretations which we give to the impressions of our senses. In order to help the audience focus on those parts of a message which are intended to be important, some attempt must be made to anticipate these selective processes; however, at times these processes can be anticipated so that we sketch key points and leave the receiver’s imagination to close the gaps.” (Morgan & Welton, 1992)

**What motivates human behavior?**

Human motivation can be discussed and interpreted in terms of behavior. Rapaille states, “…Henri Laborit, … drew a clear connection between learning and emotion, showing that without the latter the former was impossible. The stronger the emotion, the more clearly an experience is learned.” (Rapaille, 2007) Norman describes this occurrence well, “…everything that you do has both a cognitive and an affective component– cognitive to assign meaning, affective to assign value. You cannot escape affect: it is always there. More important, the affective state, whether positive or negative affect, changes how we think.” (Norman, 2005) Humans strive to achieve a positive affect in every situation in which they find themselves. They seek out pleasure or denounce and escape from pain. (Skinner, 1981) Motivation propels human behavior forward to achieve positive affect and personal satisfaction. It is for this reason we must ensure the interaction between human and technology be tailored to meet the specific
emotional needs of users. In our case the interface take on the activity theory role of mediator.

Donald Norman describes an affordance as, “the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used.” (Norman, 2005) In our case the ”thing” becomes the mediating tool for which the interaction takes place through, the interface. This affordance affects the motivational level of the user. In our current framework the affordance is the website interface which in turn affects the users’ level of motivation in the interaction. The designers’ job than includes eliminating or diminishing the intrusion of barriers between the human and the affordance level. Norman describes this task, “The design challenge is to keep the virtues while removing the barriers: make it easier to store, send, share. Make it easier to find just the desired pictures years after they have been taken and put into storage.” (Norman, 2005)

**Human perception & the “experience”**

Each of the major behavioral theoretical approaches develops a view on human motivation. Operant learning suggests consequence is the primary component determinant in motivation. In contrast classical conditioning contends that biological response to external stimuli produces the necessary motivation to guide behavior. B.F. Skinner has given us yet another explanation for behavior. He believes that behavior reflects the natural instinctual response of all beings to gravitate towards the positive or
pleasurable stimuli or experience; we are therefore motivated to move toward positive
stimuli and away from negative ones. (Skinner, 1981)

Another interesting cognitive approach to human motivation is called expectancy
theory; devised and explored by Victor Vroom, a theorist interested in studying employee
motivation in the workplace. Expectancy theory proposes an interesting relationship
between motivation, expectancy, instrumentality, and value. In detailing these
components Vroom describes expectancy as the perceived probability of success,
instrumentality as the connection of success and reward, and value as the value of
obtaining the goal. When combined they create the following equation: Motivation =
Expectancy * Instrumentality * Value. (Vroom, 1967) The use of the multiplier as the
mathematical operator indicates that all three of the variables (expectancy,
instrumentality and value) must rank high in order to produce a high value of motivation.
If the individual does not believe they can achieve the goal, doesn’t connect with the
result or goal, or see the personal value in completing the goal, than any low value will
cause motivation to fall.

Perception plays an important role in the processing of information and take place
in fractions of seconds. Furthermore they have direct affect regarding the experience of
human interactions. A side note appearing in the March 2006 issue of Communications of
the ACM, “First impressions not only count, some lasting impressions are made in the
first 50 milliseconds of viewing. So concludes a study by researchers at Carleton
University in Ottawa who found the brain can make flash judgments almost as fast as the
eye can take in information…the study shows the snap decisions Internet users make about the quality of a Web page have a lasting impact …”

One a final note, Moran and Card stress the importance of perception during the system interaction, and specifically caution designers to be aware of “perceptual causalities” of perception. Human perception happens at such an extreme rate designers must be aware of specific “perceptual phenomena derive[d] from the fact that similar visual stimuli that occur within one Perceptual Processor cycle tend to fuse into a single coherent percept.” They describe suggest this example, “… the rate at which frames of a moving picture need to be changed to create the illusion of motion.” While similar stimuli are processed in rapid succession, “one way for two distinct stimuli to fuse is for the first event to appear to cause the other.”

Cognitive Thin Slicing: The power of a glance

In Rambler Dr. Samuel Johnson, an essayist, poet, biographer, lexicographer, and critic of English literature says, “Few have strength of reason to overrule the perceptions of sense, and yet fewer have curiosity or benevolence to struggle long against the first impression: he who therefore fails to please in his salutation and address is at once rejected, and never obtains an opportunity of showing his latest excellences or essential qualities.” (Johnson & Bate, 1968) The concept of first impressions has produced several maxims throughout the ages, one of the most famous being, “You never have a second chance to make a first impression.” The New York Times columnist, Malcolm Gladwell, in his book, Blink, has applied this concept to cognitive human behavior and popular
culture, a concept borrowed from the neurosciences that Gladwell describes as thin-slicing. Thin slicing is a type of thought that happens in the blink of an eye. This rapid cognition, in essence, produces “Narrow slivers of experience.” (Gladwell, 2007) These slivers than influence and optimize human experiences. The concept of thin-slicing is something innate in human beings. We are constantly subconsciously evaluating people, situations and experiences.

According to Gladwell, “Thin-slicing refers to the ability of our unconscious to find patterns in situations and behavior based on very narrow slices of experience.” (Gladwell, 2007) Humans make associations between situations and experiences by comparing experiences already processed and using pattern matching to correlate past experience and new situations. “We [humans] make connections much more quickly between pairs of ideas that are already related in our minds than we do between pairs of ideas that are unfamiliar to us.” (Gladwell, 2007) Rapid cognition is a phenomenon which occurs in humans. “Thin-slicing is not an exotic gift. It is a central part of what it means to be human. Thin-slice occurs whenever we meet a new person or have to make sense of something quickly or encounter a novel situation. We thin-slice because we have to, and we come to rely on that ability because there are lots of hidden fists out there, lots of situations where careful attention to the details of a very thin slice, even for no more than a second or two, can tell us an awful lot.” (Gladwell, 2007) The human brain toggles between the conscious and unconscious modes of thinking constantly, analyzing and making decisions for us. (Gladwell, 2007) This constant engagement of our conscious and subconscious brain provides a compulsory thought process of complex situations at
an automated level. “Thin-slicing is part of what makes the unconscious so dazzling. … when our unconscious engages in thin-slicing, what we are doing is an automated, accelerated unconscious version of… [processing] complex situations.” (Gladwell, 2007)

However, this power of glance has the power to promote presumed prejudices, false assumptions and incorrect pattern matching. “Part of what it means to take thin-slicing and first impressions seriously is accepting the fact that sometimes we can know more about someone or something in the blink of an eye than we can after months of study. But we also have to acknowledge and understand those circumstances when rapid cognition leads us astray.” (Gladwell, 2007)

Donald Norman also provides thoughts on first impressions, “The overall impact of a product comes through reflection [the reflective level]– in retrospective memory and reassessment.” (Norman, 2005) He continues by stating, “We pay more attention to, and remember, messages that we like. If we are faced with a message we dislike, or which fails to confirm our prejudices, we tend to ignore those parts which make us uncomfortable.” Norman provides the necessary theory to support Gladwell’s idea and provides us with the knowledge to apply the notion of first impressions to website design. Says Norman, “For example if we dislike or mistrust the source, our interpretation of the message is likely to be hostile. In all this first impressions are vital. If we begin with a false idea about the purpose of a communication such an initial error is unlikely to be corrected, and mistakes may snowball.” (Norman, 2005)
2.2 Foundations of Human Computer Interaction

Vannevar Bush wrote in 1945, “Consider a future device for individual use which is a sort of mechanized private file and library. … He can add marginal notes and comments, taking advantage of … photography, and it could even be arranged so that he can do this by a stylus…” (Vannevar, 1945) Bush had provided a glimpse into the future; he envisioned a world where information was available at the touch of a button, or flip of a switch. Bush predicted the advent of the modern microcomputer and personal computer. Baecker and Buxton concluded, “Bush’s vision was remarkable. He not only foresaw the application of the computer to information storage and retrieval…but he also correctly anticipated the multimedia nature of computer use in the future.” (Baecker & Buxton, 1990) His visions came just two years after John Atanasoff unveiled the “Atanasoff-Berry Computer” to the world. The system designed and build by Atanasoff was a simple one function binary machine and after much debate and legal jockeying, Atanasoff was legally recognized as the inventor of the first “electronic digital computer.”

While Bush was extremely ahead of his time, conversations regarding the relationship between human and machine continued well into the 1950’s. Thinkers began to, “see the potential of the computer as a facilitator of aspects of human creativity and problem solving.” (Baecker & Buxton, 1990) From these small kernels, larger ideas on the use of computers began to take shape. Computers evolved, and so too did their applicable use in everyday lives. This expansion of computer system usage eventually necessitated a science, which dealt directly with the design, assessment, and resulting experience, hence HCI emerges into the common day vernacular.
Origins of Human Computer Interaction

Before Human Computer Interaction was even known as a discipline (HCI), scientists were busy working on the next iteration of computer systems. These scientists laid the framework and fundamentals of HCI. For example, Ivan Sutherland led pioneering work at the MIT Lincoln Laboratory in the 1960’s. Sutherland worked on a unique system that dealt with pictorial and iconographic representations on a monitor. (Sutherland, 1964) The ideas he generated have lead to the modern day icon seen in user interfaces.

While Sutherland’s work did not have a direct link to HCI, he was concerned with the interface through which his system was to be activated and manipulated. For example, Sutherland concerned himself with, “the concept of the internal hierarchic structure of a computer-represented picture [graphic] and its definition in terms of sub pictures.” The foundations of HCI study and research initiated in the early 1960’s, and congealed in the mid-80’s when individuals such as Donald Norman, Jakob Neilson and Cathleen Wharton began publishing best-practice research.

Cannons of HCI

HCI has exploded onto the academic and professional arenas and therefore has produced a prodigious amount of literature. Several seminal articles and books are “must reads,” for the student of Human Computer Interaction. This list includes Buxton, Nielsen, Norman, Tufte, Cooper, Dumas, Jacko with Sears, and Dix. These authors produced
pioneering research, important to the study of Human Computer Interaction. The most applicable of those discussed here.

In their study of publications in HCI, Chen and Panjwani uncovered what they deem the “Core of HCI Research.” Chen and Panjwani scoured 10 journals, which published in the HCI realm and generated a list of the 60 most cited authors. Once this was established the extended their search to ALL journals (regardless of discipline) and articles for these 60 authors. Once complete the extended these results into co-author publications as well and visualized the results to find common and overlapping themes in the literature. Once complete they asserted, “The core of HCI, at the centre of the network, is featured by three overwhelmingly prominent articles: Cards et al. [1983], Newell & Simon [1972], and Nielsen [1993]. The citations of these three were so high that we did not need to attempt to specify the nature of such references. We conjecture instead that the three masterpieces formed the cornerstones of HCI. By examining other emerging trends, we expect to improve our understanding of the roots of HCI.” (Chen et al., 2005)

Once the cornerstones were defined, Chen and Panjwani continued to define 7 cannons of HCI: 1) Knowledge representation and problem solving methods, 2) the worldwide web, 3) ubiquitous computing and context-aware computing, 4) usability evaluation, 5) user centered design, 6) perceptual control, 7) enterprise resource planning. Of these #6) perceptual control is the smallest area published in. “This is a relatively small area compared with the others…Taylor 1988 appears to be a main connection
between this area and the core HCI; the article was about layered protocols for computer-human dialogue.” (Chen et al., 2005)

More recently Carrol (2001) identified 4 roots of HCI and suggested fusing these roots together can drive the evolution of HCI. These for are: 1) prototyping and iterative development from software engineering, 2) software psychology and human factors of computing systems, 3) user interface software from computer graphics, and finally 4) models, theories, and frameworks from cognitive science. (Carroll, 2003)

In addition to the “core/root” dimensions of HCI, we must acknowledge that HCI shares several common factors of the human interaction experience with other fields of research including product experience design. For example, Desmet presents a general framework in his book Framework of Product Experience. Along with Hekkert, they suggest a “framework for product experience that applies to all affective responses that can be experiences in human-product interaction.” By simply substituting ‘computer,’ for ‘product,’ we can then continue on with the framework. “Three distinct components…of product experience are…[1] aesthetic experience, [2] experience of meaning, and [3] emotional experience.” (Desmet & Hekkert, 2007)

2.3 Psychology of HCI and the Human Computer Model

Desmet continues discussing each of his components, one of which states the unique role cognition plays in the HCI. “…At the level of meaning [2] cognition comes into play. Through cognitive processes, like interpretation, memory retrieval, and associations, we are able to recognize metaphors, assign personality, or other expressive characteristics,
and assess the personal or symbolic significance of products.” From this discussion we begin to realize the importance the cognitive process plays in HCI. For this reason we now turn our attention to the psychology of HCI.

When searching for literature pertinent to the relationships linking psychology and human computer interaction, *The Psychology of Human-Computer Interaction* by Stuart K. Card and Thomas P. Moran is billed as the resource for practitioners and researchers. Their book has been cited by 4325 research documents as of May 9, 2010. The majority of material relating to the psychology of HCI draws from this pinnacle publication, and therefore becomes the original source for the area. For this reason, we defer to the original text for the greater part of this review, and supplement Card and Moran when necessary with current ideas surrounding the topic at hand.

Card and Moran take a “scientific psychology” approach to this particular area saying, “…the subject of this book, is how humans interact with computers. A scientific psychology should help us in arranging this interface so it is easy, efficient, error-free and even enjoyable.” This approach they conclude is, “that the knowledge of human cognitive behavior is sufficiently advanced to enable it’s applications in computer science and other practical domains,” namely HCI.

In addition to cognitive behavioral psychology, several prongs of understanding in experimental psychology lend themselves to HCI, As sourced by Card and Moran, “…perception, performance, memory, learning, problem solving and psycholinguistics,” play pivotal roles in the interaction between human and computer. They continue to
describe how to apply psychology to HCI in terms of analyzing certain variables, “Such a psychology [of HCI] must be based on task analysis, calculation, and approximation.”

They continue on a rather lengthy discussion and conclude, “When psychology is applied in the context of a specific task, much of the activity hardly seems like psychology at all, but rather like an analysis of the task itself. The reason for this is clear: humans behave in a goal-oriented way. [Humans] …attempt to adapt to the task environment to attain their goals. Once the goals are known or can be assumed, the structure of the task environment provides a large amount of the predictive content of psychology.” For these reasons, task analysis has become the norm for analyzing the success or failure of an interaction, the subject trying to finish a goal (the task) successfully. (Card, 1983)

**The Human Computer Model**

Card and Moran have introduced the hugely applied “Human computer model” of HCI. This model has allowed researchers and practitioner’s the ability to visualize the interaction, and apply the laws of mathematical reasoning. By assigning mathematical formulas to each dimension of the interaction, Card and Moran have distilled interaction to a series of formulas and calculations. (See Figure 6)
The model segregates the interaction into several distinct areas: the perceptual system, the cognitive system and the motor processor system. Along with these systems are working memory consisting of visual image and auditory storage areas.

**Human-Computer model: Perceptual system**

The area of most concern to this research is the perceptual system, which consists of sensors and buffers, with the most important being the visual image storage area of memory. During any interaction, the perceptual system, “carries sensations of the
physical world detected by the body’s sensory systems to internal representations of the mind by means of integrated sensory systems.”

During human-computer interaction the perceptual system, particularly the eye, plays the most important role as it provides the visual representation for processing, “The retina is sensitive to light and records its intensity, wave length, and spatial distribution. Although the eye takes in the visual scene over a wide angle, not quite a full half-hemisphere, detail is obtained only over a narrow region (about 2 degrees across).” As designers we must be sensitive to the field of coverage area the eye can supply to the processing units upon any exposure to stimuli. While limited focus and information can be supplied at any one time, the eye is in constant motion supplying information to the processors though an area, “[c]alled the fovea, the remainder of the retina [which] provides peripheral vision for orientation. The eye is in continual movement in a sequence of saccades, each taking about 30 msec to jump to the new point of regard and dwelling there 60-700 msec for a total duration of Eye-movement = 230 [70-700] msec. This calculation can be viewed as the “perceptual cycle.” (Card, 1983)

While the fovea provides for peripheral information intake, the body subconsciously adjusts for targets outside the limited range, “Whenever the target is more than…30 degrees away from the fovea, head movements occur to reduce the angular distance. ..Eye movements and head movements-operate as an integrated system, largely automatically, to provide a continual representation of the visual scene of Interest to the [viewer].” Fitts’ law, which attempts to predict the amount of time it takes for a specific
target area to be activated during a human-computer interaction, is heavily based on the assumptions of Card and Moran. (Card, 1983)

Applying psychology to Design

“Design is where the action is in the human-computer interface. It is luring design that there are enough degrees of freedom to make a difference. An applied psychology brought to bear at some other point is destined to be half crippled in its impact. We suspect that many psychologists would tend to pick evaluation as the main focus for application (though some might have picked training).” This suggests that Card and Moran would like to see more emphasis dedicated to the measurement of the experience in terms of aesthetic quality and emotional impact, rather than pure task analysis and evaluation.

The psychology as applied to interface and design can be best described by Moran and Card as, “The psychology of the human-computer interface is generally individual psychology: the study of a human behaving within a non-human environment (though, interestingly, interacting with another active agent).”

Whilst applying psychological principles to the design of interfaces and systems, Moran and Card also describe a framework for which the designer may work. Of ultimate concern of an applied psychology is performance, the success or failure of the human-computer system. Several variables as identified by Moran and Card present themselves in this applied model, “…which [they] call ‘performance variables.’ The basic performance variables of a human-computer system are concerned with what tasks the
system can do (functionality), how long it takes to acquire the functionality (learning), how long it takes to accomplish tasks (time), how frequently errors occur, and how consequential they are.” These ‘performance variables,’ have been the focus of all standard usability testing methodologies.

The performance variables of a human-computer system have been determined by overarching structural variables, which Moran and Card define in their Rationality Principle. This principle can be summarized in the formula:

\[
\text{Task + User + Computer} \rightarrow \text{System Performance}.
\]

From this formula of the Rationality Principle, designers can extricate an intrinsic system structural performance prediction, which can be used during the design phases. This had been termed the “performance model,” by Moran and Card. “The designer’s job is to specify a human-computer system satisfying the requirements…[however], the performance aspects of a system are not derivable from a descriptive specification,” for this reason we turn to the performance model and illustrate that model as a function of several ‘performance variables,”

\[
\text{Model(Task, User, Computer) } \rightarrow \text{ Performance Prediction}
\]

**Design functions as related to human computer interaction**

The Merriam-Webster Dictionary gives the following broad definition of “Design”

“1: to create, fashion, execute, or construct according to plan. 2 a: to conceive and plan out in the mind, b: to have as a purpose, c: to devise for
A more utilitarian definition is contributed by Victor Papanek: “The mode of action by which a design fulfills its purpose is its function.” (Papanek, 1985) While Rickard Buchanan writes, “Design is the human power to conceive, plan, and realize products that serve human beings in the accomplishment of any individual or collective purpose.” (Buchanan, 2001) Buchanan’s definition explains the meaning of design in terms of the problems it solves and the novel systems design creates.

The most successfully designed systems now provide new or unique experiences and depend less on aesthetics. This transition has been based, in part, on audience research, and task analysis. “Design must become an innovative, highly creative, cross disciplinary tool … It must be more research oriented, and we must stop defiling the earth itself with poorly designed objects and structures.” (Papanek, 1985) Designers must become more aware of the consequences of their designed systems, and the associated emotions they may evoke.

Donald Norman sub-divides design into three “thought” categories, or levels, which are subsidiaries of the construct Norman calls “Behavioral Design.” The three thought levels are visceral, behavioral, and reflective. The design requirements for each level differ in their necessities. Norman explains that all three levels appear in real experiences, and rarely are seen independent of one another. Describing the visceral level, he writes, “The visceral level is pre-consciousness, pre-thought. This is where appearance matters and first impressions are formed. Visceral design is about the initial
impact of a product, about its appearance, touch and feel. The next level is the behavioral level, “The behavioral level is about use and about experience with a product.” The finally level is reflective thought. “It is only at the reflective level that the consciousness and the highest levels of feeling, emotions and cognition reside. It is only here that the full impact of both thought and emotions are experienced.” (Norman, 2005)

The visceral and behavioral thought levels describe only the here and now -- instantaneous reactions to the immediate moment. In contrast thought at the reflective level extends into the future and makes comparisons with the past. “Reflective design, therefore, is about long-term relationships, about the feelings of satisfaction produced by owning, displaying, and using a product. A person’s self-identity is located within the reflective level, and here is where the interaction between the product and our identity is important as demonstrated in pride (or shame) of ownership or use. Customer interaction and service matter at this level.” (Norman, 2005) As noted earlier, humans strive to group and define experiences though patterns. Norman echoes this sentiment, “The visceral level is incapable of reasoning: of comparing a situation with past history. It works by what cognitive scientists call pattern matching.” (Norman, 2005)

Allowing reaction to design to be fragmented into three levels of thought allows the design to address each individual requirement independent of the others. Visceral design concentrates on the appearance of the system while behavioral design focuses on the personal pleasure and effectiveness of the system’s use. Finally reflective design directs attention to the self-image and personal satisfaction the system creates which the
user stores in long-term memory. (Norman, 2005) Appropriate design strategies at the visceral and behavioral levels should lead to a positive reflective stage.

Humans are prewired to receive emotional signals from our environment, signals that are automatically processed and interpreted at the visceral level. “Visceral design is what nature does.” (Norman, 2005) Norman includes natural occurring examples such as symmetry, round smooth surfaces and objects that are pleasant to touch. Continuing in this discussion of symmetry Norman adds: “Human preference for faces and bodies that are symmetrical presumable reflects selection of the fittest.” (Norman, 2005) Attentiveness to visceral design is important because, “Visceral design is all about immediate emotional impact.” (Norman, 2005) Furthermore, when we perceive something as “pretty” that judgment comes directly from the visceral level.

The behavioral level concerns itself with usage. “Behavioral design is all about use. Appearance doesn’t really matter. Rationale doesn’t matter. Performance does. This is the aspect of design that practitioners in the usability community focus upon.” (Norman, 2005) By focusing solely on this area, usability practitioners loose sight of the visceral and reflective components resulting in sub-par interactions. They have neglected to address the intended outcome of the interaction. “The first step in good behavioral design is to understand just how people will use a product.” (Norman, 2005) Norman’s reflective level is all about how people feel after using the product. “Their real value can be in fulfilling people’s emotional needs, and one of the most important needs of all is to establish one’s self-image and one’s place in the world.”

“Attractiveness is a visceral-level phenomenon– the response is entirely to the surface of
an object. Beauty comes from the reflective level. Beauty looks below the surface. Beauty comes from the conscious reflection and experience. It is influenced by knowledge, learning, and culture. Objects that are unattractive on the surface can give pleasure. Discordant music, for example, can be beautiful. Ugly art can be beautiful.” (Norman, 2005)

When relating design to experience design, or human-centered design, we will do well to be aware of Donald Norman’s definition, “User experience encompasses all aspects of the end-user’s interaction with the company, its services, and its products. … True user experience goes far beyond giving customers what they say they want or providing checklist features. In order to achieve high-quality user experience in a company's offerings there must be a seamless merging of the services of multiple disciplines: including engineering, marketing, graphical and industrial design, and interface design.” (Norman, 2005)

2.4 Usability Evaluation Methods (UEMs)
Gray and Salzman define usability evaluation methods as methods, “…used to evaluate the interaction of the human with the computer for the purpose of identifying aspects of this interaction that can be improved to increase usability.” (Gray & Salzman, 1998) When devising the standards of measure to determine the success or failure of a system, not only have researchers turned to authors such as Moran and Card, Nielsen and Powers, they turned to the International Standards Organization. ISO 9241-11: Ergonomic requirements for office work with visual display terminals, Part 11: guidance on usability.
This document attempts to set a “standard,” of usability for “office work with visual display terminals.” These standards have allowed researchers and practitioners to justify the use of quantitative data and complex mathematical functions and equations to define the standard usability evaluation method. (Frøkjær et al., 2000)

These standards essentially declare, “Usability comprises the aspects of effectiveness, efficiency, and satisfaction.” (Frøkjær et al., 2000) With satisfaction playing on key measurable role in the usability of a system, the ISO standard defines satisfaction as, “…the users’ comfort with positive attitudes towards the use of the system.” Frokjaer et al. suggest, “Users’ satisfaction can be measured by attitude rating scales such as SUMI… the software usability measurement inventory.” (Frøkjær et al., 2000)

As stated above, traditionally usability evaluation methods concern themselves with the quantitative measures of system interaction while overlooking the user satisfaction variable. Frokjaer et al. research examined the correlations amongst effectiveness, efficiency and satisfaction in several UEM’s. They found, “user satisfaction is not simply correlated with performance measures such as task completion time and grade.” Moreover they conclude that traditional UEM’s do not capture the true factors of usability of any tested system, “This is a problem for the HCI community, since more then half of the last three years of CHI-studies concerning compiled tasks do not measure all aspects of usability.” (Frøkjær et al., 2000)

Even with flawed methodologies Gray and Salzman state, “…doing something is almost always better then doing nothing.” They caution HCI practitioners on making
claims and choices based on the results of such tests, “… choices based on misleading or erroneous claims can be detrimental- compromising the quality and integrity of the evaluation.” They argue that the current set of usability measures does not possess the commitment necessary to accurately predict usability, “[UEM’s] require a substantial commitment of time and resources. Necessarily, all such experiments are limited in scope, and these limits must be explicitly acknowledged.” By defining the limit and score the research realizes that, “Although something is being measures, is far from obvious that these measures really reflect sensitivity to usability.” (Gray & Salzman, 1998)

As important as usability is to the central tenet of Human Computer Interaction, Gray and Salzman state it best, “usability is a core construct in human-computer interaction (HCI). Methods to evaluate the usability of various software packages [interfaces, etc] have been of intense interest to HCI researchers and practitioners alike. Various UEM’s have been created and promoted, while the appeal of some UEM’s rest on common sense and the persuasiveness of proponents of that UEM.” They suggest the current trend of testing is flawed and question the efficacy of current models of testing. (Gray & Salzman, 1998)

**Reviews of traditional usability evaluation methods**

Typically UEM’s can be classified in one of two categories: analytic or empirical.

Analytic UEM’s include techniques such as heuristic evaluation, cognitive walkthroughs, guidelines, GOMS, etc. While empirical methodologies, “include a wide range of
methods and procedures that are often referred to simply as user testing.” (Gray & Salzman, 1998)

Gray and Salzman conducted an exhaustive analysis of several popular usability testing methodologies. From this research, we begin to see the flaws present in current UEM’s. One must first realize the intrinsic difference between experiments and usability testing. Current models of usability testing borrow their analysis methodologies from statistical inferences. “Statistically conclusion validity is more of an issue for those who would conduct experiments than for those involved in user testing. Because user testing assumes so many of the trappings of [the] experimental method it may seem reasonable to accept standards that are appropriate for user testing as appropriate[d] [from] experimental research.” (emphasis added). (Gray & Salzman, 1998)

The number of participants required to make statistical assertions is one glaring distinction between experiments and current usability testing. Where N=number of participants, several experiments (N>15) on the use of usability studies have concluded that for general user testing, only a few participants are needed to identify problems (N=7-10). Furthermore even fewer are needed to identify severe issues (N=3-5). (Virzi, 1992; Nielsen, 1994)

Several concerns are immediately raised when N<15 and a distinction must be made between standards appropriate for, “the usability lab and those appropriate for research.” (Gray & Salzman, 1998) Researchers using N>15, must be reminded that the assertions and claims came from, “experiments that used large numbers of participants (Virzi performed 3 studies with 12, 20 and 20 participants, Nielsen and Moluch used 34
and Nielsen used three groups of 31, 19, and 14).” We must ask, why then do current usability standards allow for N<15?

The usability standards currently being applied in the discipline are inherently flawed. Several have design issues, “The design of many of the experiments is such that neither the data they produce nor the conclusions drawn from the data are reliable or valid.” [gray] Small flaws with the testing methodologies combined with low sample sizes (N<10), necessitate the acknowledgement of skewed results, “it is necessary that researchers and practitioners understand how small features of an experimental design can cast large shadows over the results and conclusions that can be drawn.” (Gray & Salzman, 1998)

Those UEM’s that currently rely on Nielsens’ suggestion of N=(3 to 5), suffer from low statistical power, a lack of random heterogeneity of participants and over comparisons with flawed assertions. With low statistical power the simple descriptive statistics (averages, percents, tallies) currently utilized in usability testing, fail to provide any statistical certainty of a testable hypothesis, therefore unable to predict the effect of interaction. (Gray & Salzman, 1998)

Salsman and Gray point to Nielsen’s study as one particular example of flawed statistical conclusions. This particular study concentrated on the usability specialist vs. standard user and their response to certain interface designer. Nielsen’s study utilized a larger sample size, which mitigates the “wildcard effect,” however he failed to neither publish nor report any of his statistical test results (ie T-test, standard deviations, correlations, etc.) Without the publication of test results, Nielsen went so far as to deduce,
“usability specialists with expertise in the specific kind of interface being evaluated [double experts] did much better than regular usability specialists without such expertise [single experts], especially with regard to certain usability problems that were unique to that kind of interface.” A reader is lead to believe he relied on a comparison of group means without publishing those results.

Salzman and Gray conclude their research with an investigation of (Desurvire et al., 1992) Desurvire, Konszeila and Atwood’s 1992 study, one that attempted to reaffirm Nielsen’s previous claim that 3-5 participants (N=3-5) was an acceptable sample size to reflect usability issues. Desurview et.al., conducted user-testing and made several claims regarding sample size. Salsman and Gray, “indicate the authors Desurvire, Konszeila and Atwood’s were overly zealous in interpreting almost every difference between two numbers as real (we count 57 such claims.” (Gray & Salzman, 1998)

With all currently implemented usability methods of evaluation containing some flaw, one realizes that creating a flawless UEM is unattainable. The researcher can only strive to be as descriptive in their interpretation and predictions as possible. The researcher must also acknowledge the human factors in the interaction setting aside the “effect of validity,” or, “measuring usability” of the data. The studies Salsman and Gray reviewed all, “emphasized the problem-count approach to usability with the goal (implicit or explicit) of providing focused feedback to software designers on specific problems that if fixed would increase usability. Unfortunately, by ignoring threats to the effect construct, the message that these studies convey is an erroneous one… UEM’s that name the most potential problems [are] the most effective. If practitioners are to use such quick
and easy measures with confidence, then links between interface features and performance outcomes must be carefully forged.” (emphasis added). (Gray & Salzman, 1998)

2.5 Qualitative data analysis in social sciences (human sciences)

Creswell defines qualitative research best when he states, “Qualitative research is often, but not always, exploratory in nature this… exploratory research generates information about unknown aspects of a phenomenon.” (Creswell, 1997) Qualitative methodologies of analysis concern themselves with the analysis of human experience, with some authors defining qualitative statistics as those residing in the realm of human science research. Van Manen and Dilthey characterizes this matter, “with one word:

Geisteswissenschaften, the human world with it’s characteristics of Geist: human mind, thought, consciousness, values, feelings, emotions, actions, and purposes. All of which manifest themselves in a lived human experience.” Dilthey further continues his description of the human science with, “Nature tries to taxonomize everything; human science concerns itself with explicating meaning from human experiences and phenomena as well as understanding.” (Van Manen, 1990)

Concentrating on the lived human experience

These “lived human experiences,” are analyzed in an attempt to understand, at a deeper level, the rationales for human behavior. Constructivists believe, “learners construct their own reality or at least interpret it based upon their perceptions of experiences, so an
individual’s knowledge is a function of one’s prior experiences, mental structures, and beliefs that are used to interpret objects and events.” Moreover, that “what someone knows is grounded in perception of the physical and social experiences which are comprehended by the mind.” (Jonassen, 1992) These cognitive reflections of perception influence the future behavioral patterns of humans and when uncovered, can provide a wealth of knowledge for the designer of any interactive system. Morse (2007) asserts qualitative research is much more about “what is” rather than “how much.” (Morse, 2007)

Dithely provides a suggestion on what constitutes a “lived experience,” as, “…in its most basic form lived experience involves our immediate, pre-reflective consciousness of life: a reflexive or self-given awareness which is, as awareness, unaware of itself.” (Dilthey et al., 1996) He continues and adds that, “A lived experience does not confront me as something perceived or represented; it is not given to me, but the reality of lived experience is there-for-me because I have a reflexive awareness of it, because I possess it immediately as belonging to me in some sense. Only in thought does it become objective.” (Dilthey et al., 1996) Finally, Dilthey concludes that all experiences consist of a beginning and ending, and these two points define the phenomenon, and the aim of phenomenology is to take these lived experiences and translate them into a textual expression of their respective essence. (Dilthey et al., 1996)

VanMann draws from Dithely when he states, “Phenomenological research is the study of lived experiences.” Phenomenology asks, “What is this or that kind of experience like?” It [phenomenology] is difference from almost every other science in
that it attempts to gain insightful descriptions of the way we experience the world pre-reflectively, without taxonomizing, classifying, or abstracting it [the experience].” In conclusion, “Phenomenology is, in a broad sense, a philosophy or *theory of the unique*; it is interested in what is essentially not replaceable. We need to be reminded that in our desire to find out what is effective systematic intervention (from an experimental research point of view), we tend to forget that the change we aim for may have a different significance for different persons.” (Van Manen, 1990)

With qualitative research and inquiry incorporating several aspects of phenomenology, we begin to see the influence philosophy has had on the methodologies involved in the analysis of qualitative data.

**Choosing among traditions**

Qualitative research has not always been seen as an acceptable means of justification during research, especially when the relationship to the discipline of philosophy is involved. It has only been recently, in the past 20 to 30 years, that qualitative research has gained widespread acceptant. (Teddlie & Tashakkori, 2008) Denzin and Lincoln state, “over the past two decades, a quiet methodological revolution has been taking place in the social sciences … the extent to which the ‘qualitative revolution’ has over-taken the social sciences and related professional fields has been nothing short of amazing.” Researchers are beginning to see the richness qualitative inquiry can provide to their quantitative results. (Denzin & Lincoln, 1998)

Several “traditions of qualitative inquiry,” have been suggested by a plethora of authors over the past 30 years, the majority of which conclude that qualitative data is
exploratory in its purpose, contrary to the nature of quantitative data whose purpose in research is, “often confirmatory in nature and driven by theory and the current state of knowledge about the phenomenon under study.” (Creswell, 1997; Denzin & Lincoln, 1998; Teddlie & Tashakkori, 2008) Padgett went so far as to state, “…qualitative data [exists] on a continuum based on the degree of abstraction and processing involved.” Teddlie suggests, “There are several traditions associated with qualitative research (Creswell, 1998, Patton, 2002) including grounded theory, critical theory, phenomenology, biography, and case study.” (Teddlie & Tashakkori, 2008) (Patton, 1987; Creswell, 1997) This exploration and abstraction of data is unique to qualitative inquire, and is taxonomized best by Creswell.

**The Case study and Phenomenology traditions**

For purposes of this study, Creswell’s five traditions of qualitative inquire were examined. As previously demonstrated, various authors have presented several traditions; however, Creswell has created the most accessible list: biography, phenomenology, grounded theory, ethnography, and case study. Creswell’s five traditions have become foundational in the field of qualitative research. Several excellent descriptions of each tradition exist; however, only two are specific to this research and therefore will be discussed at length.

**Case Study**

According to Creswell, “Whereas some consider ‘the case’ an object of study (Staje, 1995) and others consider it a methodology (eg Merriam, 1998) a case study is an
exploration of a “bounded system” or a case (or multiple cases) over time through
detailed, in-depth data collection involving multiple sources of information rich in
content.” (Creswell, 1997)

Padgett defines case studies as heuristic devices that involve the assemblage and
summary of data so that it may be viewed holistically. (Padgett, 1998) While Teddlie
defines case study research as, “… research involved in developing an in-depth analysis
of a single case or of multiple cases.” Teddlie further explains the origins of the case
study as emerging, “…from several fields, such as political science, evaluation research,
business, law, and so forth. Data collection for case study research typically involves a
variety of sources that may include quantitative data relevant to the case or cases.”
(Teddlie & Tashakkori, 2008) In essence, the case study can be viewed as the tradition of
qualitative inquiry charged with the investigation of individual (or group) events.

**Phenomenology**

With many variegated definitions of phenomenology it is difficult to ascertain an exact
definition; however, Teddlie defines is most succinctly, “phenomenology is a research
orientation stressing researchers' subjective experiences, social perceptions, and ‘naïve’
analysis of events and phenomena.” (Teddlie & Tashakkori, 2008) While Creswell
defines phenomenology as a method which, “describes the meaning of lived experiences
for several individuals about a concept or...phenomenon…This involves exploration of
the ‘structures of consciousness in human experiences’” (Creswell, 1997) Furthermore,
“A phenomenological study describes the meaning of the lived experiences for several
individuals about a concept or the phenomenon. Thenomenologists explore the structures of consciousness in human experiences.” (Polkinghorne, 1989a) (Moustakas, 1994)

With a working definition of phenomenology, we can now turn our attention to preferential approaches of phenomenology analysis. Take Creswell’s preferred approach, psychological, as it relates to a phenomenological study. He quotes Moustakes and recounts from the Duquesne Study: “[the central tenets of psychological approach is] to determine what an experience means for the persons who have had the experience and are able to provide a comprehensive descriptions of it. From the individual descriptions, general or universal meaning are derived, in other words, the essence of structures of the experience.” (Moustakas, 1994; Creswell, 1997)

Moustakes (1994) approach stresses the importance of context in qualitative data analysis. He set forth five dimensions of his preferred approach he termed transcendental phenomenology:

1) “Transcendental phenomenology is concerned with wholeness, which involves examining entities from many perspectives until the 'essence' of the phenomenon is revealed.

2) It seeks meaning from 'appearances' and arrives at 'essences' through 'intuition and reflection on conscious acts of experience.'

3) It is committed to 'descriptions of experiences,' not analyses or explanations

4) The investigator has a 'personal interest' in the entity under investigation; therefore, the process is necessarily 'autobiographical.'

5) The primary evidence of scientific investigation is the investigator's 'thinking, intuiting, reflecting, and judging.”
All researchers agree, one must be cautious when analyzing qualitative data and must strive to maintain the integrity of data collected. Several methodologies exist to assist the researcher. (Moustakas, 1994; Creswell, 1997; Teddlie & Tashakkori, 2008)

2.5 Qualitative Analysis and validity

Regardless of tradition, modern day qualitative research can be categorized into two schools of thought: findings that are discovered, and findings that are social constructions; Creswell further expands on Padget when he states, “almost all qualitative data analysis can be divided into two types: categorical or contextualizing strategies.” Creswell suggests that certain, “Categorical strategies break down narrative data into smaller units and then rearrange those units to produce categories that facilitate a better understanding of the research question.” He then continues to describe contextualizing as “… strategies [that] interpret narrative data in the context of a coherent whole 'text' that includes interconnections among the narrative elements.” (Creswell, 1997) Regardless of taxonomy or categorical nomenclature, “Underlying all qualitative research lies a common foundation: pattern recognition and thematic development. (Boyatzis, 1998)

Before any attempt to analyze the data is made, all researchers agree that qualitative research should attempt to saturate the data (continue to acquire data until redundancies are found). Once saturated, the data is then categorized, with each category representing a unit of information, i.e. event, trigger, instance, error, etc. (Strauss & Corbin, 1990) Several variations of pattern recognition and thematic development, each
with their own set of standards, rules, suggestions and methodologies have emerged. Therefore we must acknowledge the lack of universal standard and Creswell states, “undoubtedly, no consensus exists for the analysis of the forms of qualitative data.” However all agree that prior to any codification the researcher must be self-conscious and aware of their personal feelings and emotions in an attempt to “bracket” out these prejudices. (Creswell, 1997) These methods will be discussed under the header of qualitative validity.

Methodologies suggest a basic outline of qualitative data analysis will assist in inspection of analysis: start with overview of data. Read through while marking transcripts with initial thoughts and potential categories. This stage allows the research to “get a feel” for the data. Next, some suggest the research take their findings and notations back to the subject for verification. Next the research scrutinizes the words used, metaphors incorporated, or the metaphors the tester suggested. Finally, codes/categories/themes, are developed and utilized to distill and codifying the data into usable dimensions to describe the phenomena of lived experience. (Bogdan & Biklen, 1992)

With regard to coding, Teddlie suggests three general types of qualitative analysis regardless of tradition used. He states, “there is a search for themes, which are the dominant features or characteristics of a phenomenon under study, across all types of qualitative data analysis. Most qualitative analytic techniques involve generating emergent themes that evolve from the study of specific pieces of information that the investigator has collected… Although called a variety of different names, thematic
analysis has been used in virtually all human sciences.” (Teddlie & Tashakkori, 2008)

Teddlie suggests three general types of qualitative data analysis:

1) Categorical strategies: break down narrative data and rearrange those data to produce categories that facilitate comparisons, thus leading to a better understanding of the research questions.

2) Contextualizing (holistic) strategies.

3) Qualitative data displays are visually presentations of the themes that emerge from the qualitative data analysis. Displays may be used to summarize information from either categorical or contextualizing strategies or as a separate data analysis scheme.

**Coding / analysis: Categorical strategies**

Padgget states, “Coding and thematic development are the most commonly used analytic procedures in qualitative research.” (Padgett, 1998) For example, Lincoln and Guba refined a technique known as Constant comparative analysis from the initial work by Glaser and Strauss. This technique allows, “allows analysts to compare different pieces of data, refine or tighten up categories, and move on to higher conceptual levels. (Glaser, 1968; Lincoln, 1991) Taylor and Bogdan assert that the qualitative researcher using the constant comparative method, “simultaneously codes and analyzes data in order to develop concepts. By continually comparing specific incidents in the data, the researcher refines these concepts, identifies their properties, explores their relationships to one another, and integrates them into a coherent theory.” (Taylor & Bogdan, 1984)  

Glaser and Strauss extend this definition of constant comparative analysis by providing four stages of analysis:
1) “comparing incidents applicable to each category- each “incident” is compared to a category to which it might (or might not) belong.

2) Integrating categories and their properties- comparing 'incidents' to tentative versions of rules that will describe the category

3) Delimiting the theory – reducing the original larger list of categories to a parsimonious set of more inclusive, saturated categories.

4) Writing the theory”

In addition to Glaser and Strauss’ constant comparative analysis model, “adapted the work of Barney Glaser to describe several useful ways to approach qualitative data.” He states, “these represent a set of options that a research may draw on–especially useful for graduate students and others new to qualitative methods.”

These nine areas are:

1) Process (phases, transitions, sequences)
2) degree of intensity
3) typologies
4) strategies (tactics, techniques, mechanisms)
5) interactions (mutual effects, interdependence)
6) identity (self-concept, self-reflection)
7) turning points (critical junctures, points of no return)
8) cultural and social norms
9) consensus (conformity versus conflict)”

(Bohm, 2004)
Coffey & Atkinsson provide yet another insight when they state, “coding breaks the data apart in analytically relevant ways in order to lead toward further questions about the data.” (Coffey & Atkinson, 1996)

Majority of researchers stay well within the boundary of interpretation and theorizing as to avoid producing a fully developed “theory.” For example Flick (2004) claims that theory development places an “unrealistic burden for many studies, especially graduate theses and dissertations.” Furthermore he asserts this should not preclude theoretical thinking. (Flick et al., 2004) Padgeet states “tremendous variety exists in how they [coding methodologies] are executed.”

“Similar to interpretive biographies, phenomenologist view verification and standards as largely related to the researcher’s interpretation. To illustrate different conceptions of verification in psychological approaches to phenomenology, neither empirical nor transcendental phenomenologist place substantial emphasis on verification beyond the perspective of the researcher.” (Creswell, 1997)

Dukes (1984) defines several procedures for “verification.” The first involves viewing the data through the lens of the researcher and an outside reviewer. Second, the reaction of the outside review matches his/her experience and result of the researcher. Third, The research asks if the patterns and phenomenon logically fit together; a process Dukes dubs “rational analysis of spontaneous recognition.” Finally the strength of the results depend on if the research can subsume the results against another set of data. Creswell suggests that Polkinghorne (1989) comes closest in his reasoning, “when he [Polkinghorne] discusses whether the findings are “valid.” To Creswell, validity refers to
the notion that an idea is well grounded and well supported. He asks, ‘Does the general structural description provide an accurate portrait of the common features and structural connections that are manifest in the examples collected?’ (Creswell, 1997) (Polkinghorne, 1989b)

Based on Polkinghorne’s assertions Creswell provides five questions that researchers might ask themselves while assessing the validity of qualitative conclusions:

1) Did the interviewer influence the contents of the subjects’ descriptions in such a way that the descriptions do not truly reflect the subjects’ actual experience?

2) Is the transcription accurate, and does it convey the meaning of the oral presentation in the interview?

3) In the analysis of the transcriptions, where there conclusions other than those offered by the researcher that could have been derived? Has the researcher identified these alternatives?

4) Is it possible to go from the general structural description to the transcriptions and to account for the specific contents and connections in the original examples of the experience?

5) Is the structural description situation specific, or does it hold in general for the experience in other situations. (Moustakas, 1994)

Finally Teddlie and Tashakkori sums up the process of validation of qualitative data when he states, “Inference quality is a term that has been proposed to incorporate the terms internal validity and trustworthiness (Teddlie & Tashakkori, 2008) Inference quality refers to the standard for evaluating the quality of conclusions that are made on the basis of both the quantitative and qualitative findings. Inference transferability is an umbrella term that has been proposed to incorporate the terms external validity
(quantitative) and transferability (qualitative). (Teddlie & Tashakkori, 2008) Inference transferability is the degree to which the conclusions from an MM study may be applied to other settings, people, time periods, contexts, and so on.”

Conclusion

The preceding review of literature covers a variety of diverse disciplines in an attempt to present varying aspects of Human Computer Interaction directly related to the research focus of this dissertation. Special attention should be paid to the psychology of HCI as well as the review of existing usability methodologies. In an attempt to provide inclusion of an additional qualitative channel of data relating to the human experience, this review focused on universally accepted concepts of HCI psychology as well as a complete review of current industry standard usability tests.
3. METHODOLOGY

3.1 Overview

When designing any new methodology or conceptual framework difficulties will present themselves, and this pilot study was no exception. This study proposed the introduction of several new data streams working together simultaneously while testing.

This methodology contains five unique prongs: survey instruments, measurement of perception and information absorption via eye tracking; the measurement of emotional reactions via facial reading software; collection of traditional UEM quantitative statistics (time on task, number of errors, etc.); and finally, collection of qualitative transcripts of each unique testing subject. These prongs were collected during the course of a 20-25 minute session with each participant completing a prescribed interaction with a popular e-commerce website.

The creation of testing methodology consisted of an initial focus group comprising five test subjects in an attempt at streamlining and refining the methodology in hopes of identifying potential downfalls. These five participants helped reveal flaws in the methodology, difficulties with hardware, assisted in finalizing the testing script, and overall ensured the success of the pilot study to follow. During this phase several items of concern were discovered. Typical hardware issues arose due to the nature of video, audio and gaze data all to be captured simultaneously and on the same capture computer. Secondly, acquiring near perfect video of the subject for post-processing by FaceReader (the emotional facial processing software) proved extremely difficult due to the stringent requirements of lighting and camera position.
After modifications to the testing methodology and analysis were complete, the large-scale pilot test commenced according to the refined analysis model based on the findings of the pilot study. Below are descriptions of the initial and modified analysis models.

3.2 Sampling

As this study questions the validity of traditional usability studies utilizing a low sample size, usually n<=5, a larger sample size was necessary. An initial hope for n=30 was set, with a final sample size of n=37. Random sampling was achieved through random emails and word-of-mouth.

3.3 Data collection

The resulting five-step model of testing is outlined in Table XX. Each step collects pertinent information. First, the pre-test user survey captures demographic information as well as internet usage and familiarity with the testing site (in this case, BestBuy.com). Second, the subjects’ eye and gaze data are calibrated before being shown the BestBuy.com homepage for five seconds. (Five seconds was determined through a trial and error process with members of the focus group and furthermore based initially in part, by the research of Card Moran and their Human-Computer Model, particularly their perception cycle of 50ms per stimuli). Once the initial exposure was complete, the participant continued to step three and was asked to examine a set of task cards and rank and were asked to put the simplest tasks on their left and continuing to sort until the most
difficult task was placed on their right, based on their initial exposure to the homepage. This step allows insight into a new dimension we shall call perceived usability. By allowing the test subject to sort their tasks according to what they “perceived” to be the simplest to complete to most difficult to complete. This unique alternative to standard task assignment order has the potential to reveal unique dependencies between cardinal order of perceived usability and traditional UEM statistics.

Once sorted, the tasks were collected and stapled to finalize the subjects chosen task order. The participant was then briefed on how to interact with the testing machine and asked to complete as many tasks as possible in the ten minutes of testing time.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Data Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMI, eye tracking</td>
<td>Eye gaze data: Dwell times, gaze paths Stimuli Photos (each webpage) Trial text log file (time on task)</td>
</tr>
<tr>
<td>Morae</td>
<td>Full face video Screen and mouse recording Full audio captured Interaction measures (clicks, scroll, etc)</td>
</tr>
<tr>
<td>FaceReader</td>
<td>Full face video to analyze</td>
</tr>
<tr>
<td>Sorting</td>
<td>Cardinal sort order based on initial exposure / perceived usability</td>
</tr>
<tr>
<td>Spreadsheets</td>
<td>Pre and Post test user data</td>
</tr>
</tbody>
</table>

Table 1: Data collected by tool
During the fourth state, participants attempt to complete as many tasks as possible, in the order they indicated, while hardware and software collected several data streams: eye tracking data, full facial video, full audio, full screen, and computer interaction including number of clicks, etc.

At the conclusion of ten minutes, the subjects arrived at step five and were asked to complete the post-test user follow-up interview. Unlike traditional usability testing methodologies, PUT-Q2 incorporates a interview style of follow, therefore allowing additional qualitative information which can not be captured by using standard Likert scales (which were also incorporated in the post-test survey). Several of the questions in this post-test exchange produced rich dialogue between the subject and tester for later transcription. It was observed that participants would much rather share their thoughts in a relaxed conversational manner, rather than choose their reaction on a Likert scale.

Once the post-test was complete, the participant was thanked for their participation and asked if they had any questions regarding the procedure. This concluded the usability test and initiated data analysis.

3.1.3 Data preparation

Upon completion of the five step testing model, data was recorded and backed up externally to an individual test subject folder (ie. 32-M) to maintain the organization and integrity of data. Backup copies of the data were then analyzed. See Table 1 for a list of data presented by each tool. As illustrated, an enormous amount of data was collected.

Analysis commenced with the viewing of full face video with FaceReader. After completing analysis on the initial five subjects in the focus group it was concluded that
FaceReader was not providing the reliable results necessary to incorporate into this study. FaceReader results were heavily biased in different directions for different participants. This bias is a recognized downfall to the FaceReader software. In an effort to eliminate this bias FaceReader proposes extreme lighting (which some subjects complained about) and the placement of the video recording device in a position relative to eye level parallel to the subject. This is nearly an impossible task when the subject is in a position of interacting with a monitor. For these reasons FaceReader data was dismissed. It has been suggested that FaceReader is still in its infancy, and may be better seen as a tertiary data stream rather than a primary one. (See Appendix A for information and justification of elimination.)

SensoMetric Instruments iRED-X eye tracking machine collected data including gaze paths and information absorption rates. Once analysis began it quickly became apparent that the software produced a stimuli screen capture for each and every web page the participant loaded. The current version of software did not allow for the comparison of images across multiple test participants to match stimuli images and map them to multiple participants. The test concluded with over 1,000 stimuli images of the Bestbuy.com website and the hand sort and compare proved far too difficult, tedious and arduous to complete in any expected amount of time. For this reason all eye data except that which pertains to the initial five-second exposure was eliminated from this pilot study.

Next was collection of usability statistical quantitative data. Tests were analyzed and data for time on task, cardinal order of task, and number of errors was recorded.
Qualitative data was transcribed from each participants sessions from start to finish then “cleaned,” in an effort to eliminate testers data and only allow the respondents data to remain. Both qualitative and quantitative data were then moved into their respective software analysis packages described in the section to follow.

3.4 Analysis model

As discussed above, by incorporating several new data streams to an established usability test proved to be more difficult then anticipated. In essence this study “Bit off more then it could chew,” and therefore needed a more refined focus. The decision was made to concentrate on the initial perception and information absorption rates as well as the inclusion of qualitative data into a traditionally based UEM as related to a large-scale pilot study.

This new phased structure provided a much clearer path for this project to progress and managed to define an even more specific focus for the initial pilot study. During Phase I of the initial Pilot Study, the analysis of qualitative data was brought to the forefront along with the inclusion of perception and it’s relationship to usability. The analysis of qualitative data requires an exorbitant amount of time and therefore Phase I was dedicated solely on the implementation and usage of qualitative data as well as the methodology behind the coding and the method in which to display the data. As an aside to these two goals, an additional investigation question included justifying the universally accepted usability testing sample size of (n=5).
Quantitative Data Analysis Model

The revised model called for analysis of three quantitative data streams. Each stream’s results were calculated across the entire population ‘N,’ as well as for each individual test participant ‘n.’

1) Eye Tracking data: Results from the initial five-second exposure to an e-commerce homepage were translated into heat-maps, heat-indices, information absorption grids, and gaze paths.

2) Cardinal Task Sort: Cardinal order of tasks was recorded.

3) Traditional UEM statistics: Time on task, total task completion time and the success or failure of tasks was recorded.

Once gathered, these streams were fed into JMP and Excel for statistical processing. Traditional descriptive statistics were generated.

Qualitative Data Analysis Model

The revised analysis model called for analysis of one qualitative data stream:

Qualitative data: Transcripts were created verbatim for each participants’ test. These transcripts were then “cleaned”* and prepared for insertion into Atlas.ti for qualitative data analysis as described below.

Once the transcripts were processed they were brought into Atlas.ti for coding and analysis. Working under a framework based in the work of Padgget, Lincoln & Guba, and Taylor, a series of codes was developed.

* “Cleaning,” here refers to the process of removing the investigators questions, inserting time code into the transcripts and otherwise removing extra ASCII characters which might interfere with the analysis by the software.
The data was broken down and coded using the following procedure:

1] Code parts of test.
   a) First Impression
   b) Sorting
   c) Testing
   d) Post Test
2] Refine parts of Testing [c above]
   a) Cardinal order
   b) Task
3] Refine parts of Post Test 9 [d above]
   a) Easiest Task
   b) Most difficult task
   c) Color
   d) graphics
   e) layout
   f) ease of tasks as a whole
   g) open-ended comments
4] All transcripts finally have the subject and interviewer responses coded

A total of twenty-five codes were utilized to code the transcripts. Two codes (subject and interviewer) were used to extract or exclude specific comments while nested inside another code (discussed below). It was found that these twenty-five codes produced results that could be quickly queried and analyzed while also presenting a complete snapshot of the task at hand.

Once complete, these associated reports are generated which associate qualitative data with certain queries for analysis. For example, the researcher can query all subject responses during Task A and compare across all 32 test subjects looking for similarities
and themes. This data can then augment the quantitative data collected. The interviewers' responses can easily be removed during this stage to help distill and view only the subject’s direct responses.

The qualitative data analysis of this research has limited its scope to investigate the subjects’ responses before, during, and after a usability test in hopes of providing a complimentary level of data and understanding to the simple usability test of task analysis. Unlike traditional qualitative coding which usually translates textual content into a series of descriptive and meaningful codes across all subjects, the codes used in this methodology were created first and then applied to the text. The primary usage of the textual data is meant to compliment the quantitative statistics of the usability evaluation.

This initial phase of testing was used to test the implementation of basic qualitative coding. This method of coding allows queries to be generated on a per task, per question, and per section basis so that they may compliment the corresponding quantitative statistics.

No translation of qualitative data to quantitative data took place. (ie, no word counts, word occurrences, word pair analysis, no hit list was created, etc). A “pure” approach to qualitative analysis was taken, allowing the data to express itself and in turn compliment the quantitative data. While allowing the “pure” qualitative data to come forward, the subjectivity is left to the data analyzer and usability analysis expert.
4. DATA ANALYSIS

4.1 Qualitative & Quantitative Analysis

The revised testing and analysis model was applied to a pilot study conducted during the months of June and July in the year 2009 on the campus of Iowa State University. All Iowa State University Internal Review Board requirements were met and the study approved to utilize human test subjects. All relevant paperwork and documents can be found in the appendix. A random sampling was achieved through word-of-mouth, and email or face-to-face solicitations all within Ames, Iowa. A broad cross section of age, education level, professional status and industry were achieved. The total number of participants achieved was n=37. However, five of these individuals became members of the focus group of n=5 (subjects 01-05); therefore, the final number in the pilot large-scale usability test was n=32. Therefore, this pilot study concentrated analysis on the data gathered from test subjects numbered 06-37 garnering n=32 as a sample population size.

Several instruments were used to compile and analyze data in this pilot study. The Red 2000 Eye Tracking system manufactured by SensoMetric, Inc. was utilized to capture eye data and included both hardware and software. The Morae suite of software was used to capture video and audio of the subject as well as the computer screen. Atlas.ti was used to code the qualitative data. Microsoft Excel was used to compile the data sources and compute statistics. Additional statistics were calculated using the Statistical Analysis Package for Usability Testing Expanded v.2.3 produced by Jeff Sauro and MeasuringUsability.com. Adobe Illustrator was used to design information graphics as well as all subject matrices and the final population overview documents.
Upon completion of the PUT-Q2 research pilot study the data was combined and analyzed through a series of stages. First, all data was reviewed for consistency, reliability and anomalies. Any deviations were recorded and eliminated from analysis. The remaining raw quantitative data was then imported in an Excel spreadsheet for quantitative analysis.

From the successful audio captures, transcripts were generated verbatim and imported into Atlas.ti and coded using descriptors. Once the transcripts were coded, qualitative reports were generated based on the variables of first impression, sorting, task number and post testing.

This LSU-EM (large scale usability evaluation method) pilot study generated enormous amounts of data and therefore required careful compilation to ensure the integrity of the resulting statistics and testing conclusions. As mentioned previously, due to technical difficulties and inconsistency in data results, some streams were eliminated on a per subject basis. The results of these eliminations appear in Table XXX.

Once all data streams were analyzed, the resulting quantitative and qualitative data were used to create static data visualizations for each subject as well as for the sample population; these documents are deemed the PUT-Q2 Subject & Sample Matrices.

As previously presented in the methodology section, three streams of data are analyzed for the quantitative section of PUT-Q2:
1) **Eye Tracking data**: Results from the initial five-second exposure to an e-commerce homepage were translated into heat-maps, heat-indices, information absorption grids, and gaze paths.

2) **Cardinal Task Sort**: Cardinal order of tasks was recorded.

3) **Traditional UEM statistics**: Time on task, total task completion time and the success or failure of tasks was recorded.

**Eye Tracking data**

Eye tracking data was extracted from the BeGaze software application designed by SMI. BeGaze generates an abundance of reports of which four are particularly useful: heat-maps, heat-indices, information absorption grids, and gaze paths. To begin analysis, certain areas of the BestBuy.com homepage were tagged “areas of interest.” These areas divide the homepage into six distinct areas of interest: brand, navigation 1, navigation 2, navigation 2, carousel, and content. (See Figure 7) The bounds of these areas were defined in BeGaze which then allowed for calculation of the order entry sequence (the order in which the subject viewed each area of interest), dwell time (the cumulative amount of time a subject spent in an area of interest), and the percent of total time dwelled in each area of interest.
Figure 7: 5-second stimuli image including Areas of Interest
In addition to larger key areas of interest, a grid measuring six by eight (6 x 8) was overlaid on the bestbuy.com home page to further detail concentration areas within each area of interest. (See Figure 8)
BeGaze also provided scan path and absorption information for each subject. These two data streams were united and the scan path overlaid on the absorption screen. (See Figure 9).
The results of these visualizations of eye tracking data are then utilized to compile each subjects’ matrix as well as the sample overview.

**Cardinal Task Sort**

The cardinal task sort order is a new variable introduced to the standard task-analysis usability model. By allowing the subject to determine the order to complete tasks informs us of their perceived ease-of-use within the system (BestBuy.com). Additional predictions and conclusions can then be drawn regarding the usability of the site.

<table>
<thead>
<tr>
<th>Task Associations</th>
<th>Task Errors</th>
<th>Task Not Attempted</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) Locate the Sign-in Page</td>
<td>A 0</td>
<td>A 0</td>
</tr>
<tr>
<td>B) Locate a store</td>
<td>B 5</td>
<td>B 0</td>
</tr>
<tr>
<td>C) Locate contact e-mail page</td>
<td>C 0</td>
<td>C 2</td>
</tr>
<tr>
<td>D) Find information on the Low Price Guarantee</td>
<td>D 1</td>
<td>D 3</td>
</tr>
<tr>
<td>E) Find top rated gifts for boys</td>
<td>E 2</td>
<td>E 2</td>
</tr>
<tr>
<td>F) Find all outlet center TV's</td>
<td>F 2</td>
<td>F 4</td>
</tr>
<tr>
<td>G) Locate gift card terms / conditions</td>
<td>G 0</td>
<td>G 3</td>
</tr>
<tr>
<td>H) Add products to cart</td>
<td>H 2</td>
<td>H 0</td>
</tr>
</tbody>
</table>

*Table 2: Task descriptions and total errors by task*

Overall, eight tasks were created which subjects self-sorted (Cardinal sort order) based on their ‘perceived’ ease of use after a five second exposure to BestBuy.com’s homepage. The term Cardinal Sort order will refer to the test subjects’ self-sorted task position within the overall eight task slot positions. These eight Tasks\(^1\) are listed in Table 2. Prior to any analysis task that was not attempted (TNA) or erred (TE) was eliminated, so not to skew the resulting completed test data for any given Task. This process is the norm

---

\(^1\) When capitalized, ‘Task,’ will refer to the eight alpha-assigned tasks presented to the subject in this LSU. We hope to eliminate confusion between the individual Task and the statistic of the time-on task.
amongst usability tests and helps eliminate outliers as well as provide a true statistic of time-on task. The number of errors is important when analyzing the difficulty of the Task. These results can be seen in Table 3. As we analyze the cardinal sort order, several phenomena emerge.

Table 3: Tasks by cardinal sort order

<table>
<thead>
<tr>
<th>Cardinal Order 1</th>
<th>Cardinal Order 2</th>
<th>Cardinal Order 3</th>
<th>Cardinal Order 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>Count</td>
<td>%</td>
<td>Task</td>
</tr>
<tr>
<td>A</td>
<td>22</td>
<td>59.46%</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>7</td>
<td>18.92%</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>5.41%</td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>2.70%</td>
<td>D</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2.70%</td>
<td>E</td>
</tr>
<tr>
<td>F</td>
<td>2</td>
<td>5.41%</td>
<td>F</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>2.70%</td>
<td>G</td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td>2.70%</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The cardinal task sort order (1-8), can be divided into four distinct quartiles for analysis: upper, upper middle, lower middle, and lower. These four quartiles assist in seeking relationships between the cardinal sort order perceived by the test subjects, and the actual ease of use for each Task provided by the time-on task data. Table 4 illustrates the actual time-on task statistic for each Task, sorted from easiest to most difficult, along with the task sort order percentage for the sample population per cardinal sort order. As illustrated, the percentage of subjects selected Tasks A, C, B, D, E, F, F, and {G,H,D} for cardinal sort orders 1-8 respectively. Task F was the most selected task in positions 6 and 7, while {G,H,D} share the same percentage (21.62%) in cardinal sort order position 8. A correlation between the perceived ease of use and the actual time-on task means (see
section 4.3.3 for details on the procedure to prepare the descriptive statistics) occur in cardinal sort positions 1, 2, 6, and 8. In these positions the percentage of subjects self-selecting the particular Task in that position, echoes the sample population mean of time-on task for the given Task. Each cardinal sort order provides additional information on the perceived ease-of use for each task. The results from analyzing these simple statistics could indicate a correlation between perceived ease of use and actual usability results.

### Table 4: Mean task times as correlated to Cardinal Sort order

<table>
<thead>
<tr>
<th>TASKS MEANS (SMALLEST - LARGEST)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>17.86</td>
<td>37.73</td>
<td></td>
</tr>
</tbody>
</table>

(easiest)

<table>
<thead>
<tr>
<th>SUBJECT TASK SORT BY CARDINAL TASK (CO1-CO8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCENT TTL</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>59.46%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QUARTILE</th>
<th>UPPER</th>
<th>UPPER MIDDLE</th>
<th>LOWER MIDDLE</th>
<th>LOWER</th>
</tr>
</thead>
</table>

Cardinal sort order position 1 is largely dominated by Task A, being chosen by 22 subjects (59.46%), with Task B being chosen by only 7 (18.92%). This indicates nearly 60% of the sample population perceived “signing in,” to the BestBuy.com website would be the simplest of the eight tasks. When we compare this to the actual results (Table 4) we see Task A has a mean of 17.86 seconds. This mean is the lowest of all Tasks thus asserting Task A the simplest of the eight task among the sample population. We can say with a 95% certainty that the time to complete Task A will fall between 15.75 seconds and 20.26 seconds with a standard deviation of 12.79 seconds.
Cardinal sort order position 2 was not dominated by any one single task, but rather shared amongst a sixteen-point spread between Task C (14 subjects, 37.84%) and Task B (8 subjects, 21.62%). Upon examination of the respective means, Task C produced a mean result of 37.73 seconds (55.74 seconds arithmetic) while Task B produced a mean result of 55.40 seconds (62.74 seconds arithmetic). These resulting means indicate Task C was the second simplest task to complete; however we should remember Task C was not attempted for 2 subjects as the ten-minute time limit expired before the task commenced resulting in a sample size for Task C as n=30. This observation is only mentioned as a reminder that data was eliminated for those not attempted as well as erred. Task B was not the third simplest task as one might assume since it shared the second largest selection in the cardinal sort order 2 position. This indicates that subjects sorted Task B into a higher quartile then the actual mean results indicate. Furthermore, Task B had a total of 5 Task errors, which echoes the perceived usability for this Task. We can say with a 95% certainty that the time to complete Task C will fall between 29.99 seconds and 47.47 seconds with a standard deviation of 34.59 seconds.

Cardinal sort order position 3, displays Task B capturing 32.43% of the subjects self-selected sort order for this position. The second closest contenders are Task A and C with each resulting in 18.92% of subjects selecting the respective tasks for this position. Cardinal sort order position 3 is the first position we notice the subject-sorted position (Task B in sort position 3) ranks higher then the resulting actual mean data for
time-on task. This phenomenon, of subjects sorting a task ‘easier,’ then the resulting data indicate, also occurs in cardinal sort order position 7.

Results indicate Task D was actually easier than subjects indicated. We can say with a 95\% certainty that the time to complete Task D will fall between 34.48 seconds and 56.47 seconds with a standard deviation of 42.50 seconds.

**Cardinal sort order position 4** is occupied by Task D with 21.62\% of the subjects' sort choice; however, statistical results indicate Task G was the last Task to appear in the upper middle quartile. Furthermore, the majority of subjects’ sorted Task G into cardinal sort order position 8. In other words, test subjects perceived Task G (locate the gist cards terms and conditions) to be more difficult than it actually was to complete, with a mean completion time of 49.14 seconds, third behind Task A, C and D. We can say with a 95\% certainty that the time to complete Task G will fall between 41.62 seconds and 58.02 seconds with a standard deviation of 25.11 seconds.

**Cardinal sort order 5** occupies the first position in the lower middle quartile and also is the first position of the lower half of the four-quartile spread. Statistical results indicate Task B (Locate a store), with a mean completion time of 55.4 seconds, was more difficult to accomplish than subjects. With a majority subject cardinal sort order for Task B landing in cardinal sort order position 3, and the actual statistical results placing Task B in the cardinal position 5 we realize Task B was more difficult to accomplish than initially thought.

**Cardinal sort order position 6** brings again a correlation between the subject sort order and the statistical mean order with Task F holding this position. A majority
with 21.62% of subjects self-sorting Task F into the cardinal sort order 6 position. Task F finished with a mean time-to-complete task of 56.26 seconds. We can say with 95% certainty that Task F will take between 42.62 seconds and 74.27 seconds to complete with a standard deviation of 53.31 seconds.

**Cardinal sort order position 7** had Task F selected by the majority of subjects with a 29.73% selection rate; however, Task F also occupied the cardinal sort order 6 subject-sort position. This is the first occurrence we see the majority of subjects self-sorting a task into 2 cardinal sort order positions (Cardinal sort order 6 and 7 for Task F). Cardinal sort order 7 is also important because it holds the time-to-completion mean for the seventh most difficult task, Task E with a mean time to complete of 77.57 seconds. We can say with a 95% certainty that Task E can be completed between 60.87 seconds and 98.85 seconds with a standard deviation of 61.58 seconds.

**Cardinal sort order position 8** indicates the final task position, and the most difficult task to complete. Task H occupies this spot with a mean time to complete of 115.23 seconds. A correlation does exist between the subjects self-sorted position 8 and the cardinal sort order position 8 as they both exist in this position; however, Tasks G and D were also subject-sorted into the sort order position 8 by a three-way split of the subject population. We can say with 95% certainty that Task H can be completed between 94.72 seconds and 140.18 seconds with a 59.54 second standard deviation.

Upon final inspection of the cardinal task sort order amongst all subjects, we find 36 out of a possible 37 factorial (37!) permutations of the subject task sort order, resulting in only 2 subjects placing the eight tasks in the same cardinal sort order positions (when
including the pilot study subjects). For this reason, correlation statistics are not available on the entire sort positions (1-8), as no more then 2 subjects sorted the tasks identically. However, inspection of the data indicates the successful ‘perceived sort order’ for Tasks A, C, F and H into cardinal sort positions 1, 2, 6 and 8 by correlation of the time-on task completion statistics for their respective means. In other words, Tasks A, C, F and H fell in the correct cardinal sort order position according to their mean time-on task statistics. This results in 50% of the total tasks being sorted into the correct cardinal sort order by the sample population (N=32). Since cardinal sort order position 1 and 2 comprise the upper quartile and both correlate between mean time on task and subject sort order, a conclusion could be drawn that the upper quartile can successfully indicate the two simplest tasks of a usability test; however, further research is necessary to compare the means of several LSU’s to determine if this assertion can be proven.

**Traditional UEM statistics**

The descriptive statistics for each task appearing in Table 5, are based on the log transformation of the raw data values. Non-normalized usability data tends to skew to the right, and the results of this study follow this trend. In these cases statisticians, as well as usability experts such as Jackob Nielsen and Jeff Sauro suggest transforming the data by taking the log results (base 10) to better normalize the distribution of data. ((Sauro & Kindlund, 2005))
Table 5: Descriptive statistics and test subject task times, sorted by task

Traditional usability studies usually rely solely on the mean time-on task descriptive statistic to determine the ‘ease of use,’ of any given task when compared to that mean of either a ‘designers best time,’ or another sample mean generated by the test
administrator. As discussed previously, PUT-Q2 attempts to shed new insight into the perceived ease-of use for a given set of tasks by complimenting traditional usability descriptive statistics with new data streams.

**Qualitative PUT-Q2 Data coding**

The revised PUT-Q2 data analysis model was followed to code the transcripts of the pilot study. First a transcription was created of each subject's usability test (see Appendix for complete transcripts). Next, each transcript was read and sectioned according to pre-test, test, and post-test text. After division, the transcripts were imported into Atlas.ti for coding and analysis according to the methodology discussed previously. A screen shot of coding and corresponding codes in Atlas.ti can be seen in Figure 10 and Figure 11.
86

Figure 10: Atlas.ti software during coding
Once the transcripts were completely coded, queries were executed to isolate pres-test, sorting, tasks, and post-test subject’s qualitative data. The resulting reports were exported into text files. Once all processing of quantitative and qualitative data had been completed, the process of assembling data to create the PUT-Q2 Data Package began.

4.2 Data Visualizations

As discussed in the methodology section of this document, a new instrument that could quickly, easily and accurately communicate the results of a large-scale usability study required creation. The result is the PUT-Q2 Data Package. This package includes two unique sections, the subject and population matrices. These matrices are the two core
components in the PUT-Q2 Data package. Their corresponding elements are discussed in detail below.

**PUT-Q2 Subject Matrix**

The PUT-Q2 Subject & Sample Matrices were generated to allow an immediate understanding of each subjects’ testing results as well as the sample results with regard to all data streams. These multi-page documents include information visualizations (information graphics) on the following streams of data: eye-scan paths, information absorption, key areas of eye focus, time on task, cardinal order of tasks, qualitative data correlated with quantitative statistics, sample size geometric means as compared across task completion times, and other complimentary statistical and qualitative data.

The Subject Matrix is a two-page form. The first page contains several information modules (See Figure 12). Module 1 displays the subjects time to complete each task, cardinal sort order, Task mean as compared to the total population, and qualitative selections derived from reports generated from Atlas.ti. Module 2 displays the standard descriptive statistics for the sample population that successfully completed the indicated Task. Module 3 presents the verbal description of each Task, while Module 4 displays the quantitative and qualitative results of the post user test. The final Module displays additional qualitative information selected during the sort and testing phases.

Upon completion of qualitative and quantitative data analysis, the process turned to information visualization. When completed, the PUT-Q2 Subject & Sample Matrices package presents a holistic picture of each subjects’ interaction during the usability test as well as the sample population trends and results. Comparisons can be made by immediate
visualizations between the subject and the population. Additional conclusions can be drawn based on subjective analysis of qualitative data.

The most important aspect of the PUT-Q2 data package is the addition of qualitative data to standard descriptive statistics currently being used in usability testing, in addition to providing a quick-look visual representation of usability study data. For example, a simple eye-tracking analysis of a scan path as related to a five second exposure to the BestBuy.com home page reveals what designers would predict: the eye moves from large to small items as it scans a page, resting longer at larger images to increase the angle of focus, and in turn cognitively process the image (See Figure 13). This phenomenon is used in visual communication design to control the hierarchy of information presented. However, when qualitative data is added to the information graphic a new contextual conclusion can be drawn. This resulting conclusion can better assist and reveal the true usability of a website from a testing subjects point-of-view.
Figure 12: PUT-Q2 Subject Matrix (Side 1)
Figure 13: PUT-Q2 Subject Matrix (Side 2)

PUT-Q2 Sample Population Matrices

The PUT-Q2 Sample Population Matrices were generated to allow an immediate understanding of the overall state of the usability study by presenting all test subjects' testing results as well as the sample results with regard to all data streams.

The Sample Population Matrix is a two-page form. The first page contains several information modules (See Figure 13). Module 1 displays the subjects' time to complete each task, cardinal sort order, Task mean as compared to the total population, and qualitative selections derived from reports generated from Atlas.ti. Module 2 displays the standard descriptive statistics for the sample population that successfully completed the
indicated Task. Module 3 presents the verbal description of each Task, while Module 4 displays the quantitative and qualitative results of the post user test. The final Module displays additional qualitative information selected during the sort and testing phases.

Figure 14: Sample Population Matrix (Side 1)
Upon completion of qualitative and quantitative data analysis, the process turned to information visualization. When completed, the PUT-Q2 Subject & Sample Matrices package presents a holistic picture of each subject’s interaction during the usability test as well as the sample population trends and results. Comparisons can be made by immediate visualizations between the subject and the population. Additional conclusions can be drawn based on subjective analysis of qualitative data.

The most important aspect of the PUT-Q2 data package is the addition of qualitative data to standard descriptive statistics currently being used in usability testing, in addition to providing a quick-look visual representation of usability study data. For example, a simple eye-tracking analysis of a scan path as related to a five second exposure to the BestBuy.com home page reveals what designers would predict: the eye
moves from large to small items as it scans a page, resting longer at larger images to increase the angle of focus, and in turn cognitively process the image. This phenomenon is used in visual communication design to control the hierarchy of information presented. However, when qualitative data is added to the information graphic a new contextual conclusion can be drawn. This resulting conclusion can better assist and reveal the true usability of a website from a testing subjects point-of-view.
5. CONCLUSIONS, SHORTCOMINGS, AND FUTURE RESEARCH

Overview of pilot findings

To review, the two research questions for this research are:

1. Can a qualitative dimension of human interaction be incorporated into traditional UEM’s to assist in uncovering otherwise overlooked aspects of usability testing?

2. By sampling a large-scale usability study can the current UEM standard sample size (n=5) be affirmed as an acceptable number?

Question one deals with the incorporation of qualitative textual data into the traditional usability study, and moreover, if qualitative data can help uncover overlooked aspects of traditional studies. Several will argue the conclusion to this research question will be a subjective one, and some maintain the human factor must be incorporated into usability testing. Regardless of camp, the point remains; one of the simplest ways to incorporate the human factor into traditional usability testing is through qualitative data. This study married quantitative and qualitative data in a new format unique from all other usability studies.

Textual information visualized in tandem with quantitative data only further enhances the overall picture of usability. We see the subjects thoughts before, during and after the usability study. In certain cases, this information may contradict the subjects actual time on task (“I didn’t realize how simple this task would be,” while the time on said task was 25% over the mean). Other invaluable data was collected during the five-second exposure. The study revealed that subjects overwhelmingly agreed the
BestBuy.com home page is cluttered and visually over stimulating. Many subjects completely disregarded the top level of navigation with some revealing they ignore the “top banner,” since it’s ‘usually ads.’ Without this textual insight, task phenomenon may be overlooked.

A secondary emergence dealt with Task H- shopping and the inclusion of the self-sort data stream. The majority of subjects self-sorted shopping into the lower half of the quartile. This should raise concern with the usability study coordinator and beg the question, “Why is the main purpose of our website, shopping, being sorted in the most difficult positions during this study?” Without self-sorting and textual analysis this fact would be overlooked in traditional usability studies.

The conclusion of this researcher is that textual qualitative information adds an additional layer of invaluable data to the traditional usability study and research should continue on new ways to incorporate this data stream.

Research question two sought to affirm or reject, the universally accepted n=5 sample population size for usability testing popularized by Jacob Nielsen as the standard needed to uncover the majority of usability problems with any website. As illustrated in Figure 16, a usability test that wishes to detect a problem which occurs 30% of the time (30 times out of 100), with a detection rate of 85% (will only report correctly 85 times out of 100) the sample size needed is n=5. This is the rationalization Jacob Nielsen uses when he states, “most usability problems can be detected with 5 subjects…” (Nielsen, 1994) These statistics allow a problem to occur during a usability test at a rate practically 1/3, meaning almost 1/3 of the population will experience this problem. As designers and
engineers of technology and interactive experiences, should we be proud of a system that only functions for 2/3’rds of the population? This is the question the usability expert and system developers must ask prior to commencement of usability testing. With a larger sample size, more problems can be detected at a higher confidence interval.

For those system designers and usability experts who wish to test their system at a higher rate, if a sample size of \( n=37 \) is used, a problem that occurs only 5% of the time can be recognized at a confidence level of 85% (See Figure 17). It is the opinion of this researcher that a sample size of \( n=(25-35) \) is acceptable to capture a problem that only occurs to 5% of the population at a confidence level of 85%. With these higher standards, the resulting system will be “more usable,” for a larger population.
In addition to the statistical power test, this research compared two means for each of the eight tasks completed. The population mean was calculated, in addition to a mean of five random subjects. Task G resulted in an overall population mean of 57.59 seconds with n=27; a random sample of n=5 results in a mean of 53.67 seconds. While the two means seem similar, their confidence levels vary at an astonishing rate. (See Figure 18) Statistics on power and confidence inform us that the comparison of Mean A (sample population) to Mean B (n=5 random sample) with a confident level of 95%, the researcher will not capture or detect a problem 79% of the time for this particular task. With n=5, as Nielsen affirms, this result would be very near the 85% acceptable rate of detection.
However, this acceptable rate of discovery with a n=5 sample size result does not apply to all eight task. For example, Task H compares a sample mean of 125.87 seconds to an n=5 sample size mean of 80.40 seconds. The results of this statistical analysis on power and confidence inform us that the comparison of Mean A (sample population) to Mean B (n=5 random sample) with a confident level of 95%, the researcher will not capture or detect a problem almost 98% of the time for this particular task. With n=5, as Nielson affirms, this is not an acceptable rate of detection of <3% (See Figure 18).
As demonstrated through the analysis of two random task subsample means, it is clear that a definitive statement can be made regarding the acceptance or rejection of the n=5 sample size; however, one can assert that without a larger sample size, the rate of detection of problems increases at an exponential rate.

5.1 General Thoughts

Like any exploratory research study, limitations in the methodology must be acknowledged and placed into context across the investigation. This pilot methodology attempted to elicit new insight into traditional usability testing with regard to perceived usability and qualitative data analysis and incorporation in traditional usability evaluation methodologies. However, this pilot cannot erase the biases of culture and gender of the
participants as well as the ability of the equipment to capture eye data. In addition, a study of this magnitude and amount of incoming data streams was certain to have incomplete, corrupt or unusable data. In turn, this lowered the usable number in the general population from n=37 to n=32, but still allowed for accurately implementing the PUT-Q2 through an initial exploration.

As mentioned not all data streams for all subjects were 100% usable. Some may have been due to hardware failures others due to the subject moving during testing, thusly interrupting the tracking of pupils. These difficulties were most helpful and always lead to another modification to the analysis model. The resulting testing model is comparable with popular UEM’s currently used in the discipline of HCI.

Undertaking any enormous task including this research is extremely difficult. Once completed the realization was made that perhaps, any one of the key areas examined could have become a complete dissertation in itself. For this reason a small few areas examined were only afforded top-level analysis, where in the future of PUT-Q2 a much deeper analysis should occur. One example to cite would be the visualization of information in the PUT-Q2 Data Package. To ensure the best possible visualization an investigation would need to be done into alternative forms of graphic information to affirm the best tools are used.

Qualitative coding and the relationship to usability studies

Researches agree qualitative data analysis is innately difficult and extremely arduous. The analysis in this study was no exception. Upon completion of transcription, it was
necessary to devise a methodology that would allow one individual to objectively analyze the data. Ultimately, it was decided that the data would speak for itself, and therefore eliminating any subjective influence in the outcome. To the extent of qualitative analysis, this was merely an introductory phase of study. Eventually codes would be generated that could generalize and attempt to capture the essence of each quotation by the subject, a task that takes an immense amount of time. In the future, PUT-Q2 could, perhaps, undergo a more in-depth analysis of the qualitative transcripts or look to other research in an attempt at developing a new taxonomy of concepts and codes, which could be utilized to further enhance usability evaluation methods.

**Data collection techniques**

This pilot study was conducted on the campus of Iowa State University inside a faculty office. To this extend, the environment was not conducive to collection of traditional usability data. Additionally, unfamiliar hardware and software was introduced into the traditional UEM process, and later discovered much was unusable. Some test subjects felt hesitant during the test since their eye movements were being captured, as they had never been exposed to such machinery prior. In the future new streams of data should be introduced into a usability testing method individually, and not as a group. This allows for greater control over the variables and refinement of each stream of new data.
5.2 PUT-Q2: A new usability testing tool

The inclusion of additional data channels to the standard usability testing methodology required an alternative method of presenting and reviewing all collected data. The decision to translate the numerical data to a graphical visual one allows the usability researcher to quickly scan user test results on individual matrices, as well as review sample population trends. With further inclusion of sample population averages and trends, the comparative nature and presentation of this complex data set allows further correlations to be analyzed.

For example, Figure 19 presents the PUT-Q2 subject matrix for test participant 29-M. The researcher can immediately ascertain the difficulty this participant had with tasks H and E. Not only can the subject matrix reveal problems, but successes as well.
Figure 20, presents the subject matrix for participant 32-F. When compared to the matrix for participant 29-M (Figure 19) the usability researcher can immediately deduce the completion rate for participant 32-F was less than sample average for all tasks except one.

Figure 21: Sample PUT-Q2 matrix #2

This immediate comparative incorporation of data allows for quick and accurate review of usability statistics and successfully incorporates all channels of data collected during the usability study.
5.3 Conclusion & Future Implications

Upon completion of this research project the conclusion can be affirmed that qualitative data can be successfully incorporated into a traditional usability methodology. It can also be stated that a sample size of n=5 cannot be determined without further investigation measuring the comparison of task means for a larger sample size.

In the end more work is needed to refine this proposed usability methodology; however, this research presented an excellent first step into realizing that qualitative data is valuable and should be incorporated in traditional usability evaluation methods. Another lesson this researcher learned is to “shut up and listen.” During the analysis of qualitative transcripts it was discovered that the researcher interrupted what could have been an excellent description of an experience. For this reason some data could have been captured if he would have just listened. Note to self: listen.

5.4 Future Research

As previously discussed in the methodology section of this research document, PUT-Q2 morphed into a phased research project. Once again, the phases are represented in Figure 19.
The second phase of research involves the investigation of information absorption rates and surrounding fields of vision as they relate to initial exposures and interactions with e-commerce websites. The results of Phase II will then augment the PUT-Q2 pilot study and provide clarity of eye tracking data as it related to usability studies and perceptual conclusion based on initial exposures. The phases continue until ultimately the data streams captured during the PUT-Q2 testing are imported automatically into a stand-alone or web driven application that allows for full interaction with the data. One will be able to interact with the information in visual ways drilling down and mining the quantitative and qualitative data through a dynamic experience.
APPENDIX A: FACEREADER

Even with suggested lighting and capture settings, FaceReader by Noldus failed to accurately model and analyze the majority of test subjects. The following two figures are screen captures of FaceReader in use. Notice the ‘Modeling Failed,’ in Figure XXX, and the failure to predict the Personal Characteristics in Figure XXX. Additionally, FaceReader demonstrates heavy bias towards the emotion of anger.
APPENDIX B: APPROVED IRB FORMS

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

DATE: 20 May 2009
TO: Troy Abel
158 Design
CC: Roger Baer
158 Design
FROM: Jan Canny, IRB Administrator
Office of Research Assurances
TITLE: Usability testing of e-commerce homepages
IRB ID: 09-228

Approval Date: 20 May 2009
Date for Continuing Review: 19 May 2010

The Chair of the Institutional Review Board of Iowa State University has reviewed and approved this project. Please refer to the IRB ID number shown above in all correspondence regarding this study.

Your study has been approved according to the dates shown above. To ensure compliance with federal regulations (45 CFR 46 & 21 CFR 56), please be sure to:

- Use the documents with the IRB approval stamp in your research.
- Obtain IRB approval prior to implementing any changes to the study by completing the "Continuing Review and/or Modification" form.
- Immediately inform the IRB of any serious adverse experiences involving risks to subjects or others; and any other unanticipated problems involving risks to subjects or others.
- Stop all research activity if IRB approval lapses, unless continuation is necessary to prevent harm to research participants. Research activity can resume once IRB approval is reestablished.
- Complete a new continuing review form at least three to four weeks prior to the date for continuing review as noted above to provide sufficient time for the IRB to review and approve continuation of the study. We will send a reminder as this date approaches.

Research investigators are expected to comply with the principles of the Belmont Report, and state and federal regulations regarding the involvement of humans in research. These documents are located on the Office of Research Assurances website [www.compliance.iastate.edu] or available by calling (515) 294-4566.

Upon completion of the project, please submit a Project Closure Form to the Office of Research Assurances, 1138 Pearson Hall, to officially close the project.
ISU NEW HUMAN SUBJECTS REVIEW FORM

SECTION I: GENERAL INFORMATION

Principal Investigator (PI): Troy Abel
Degree: AA, AA, BSFA, BA, MFA, PhD student
Department: Art & Design
Phone: 569-850-7478
Fax: 515-294-2725
Signature of Major Professor/Supervising Faculty: Roger Baer

Exempt per 45 CFR 46.1010(C) Date: 12/18/2012

Project Period (Include Start and End Date): [mm/dd/yyyy][05/01/09] to [mm/dd/yyyy][08/10/10]

Key Personnel

List all members and relevant experience of the project personnel. This information is intended to inform the committee of the training and background related to the specific procedures that each person will perform on the project.

<table>
<thead>
<tr>
<th>NAME &amp; DEGREE(S)</th>
<th>SPECIFIC DUTIES ON PROJECT</th>
<th>TRAINING &amp; EXPERIENCE RELATED TO PROCEDURES PERFORMED, DATE OF TRAINING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troy D. Abel</td>
<td>Serve as Principal Investigator. Solicit volunteers, contact participants, schedule and conduct testing and analyze data.</td>
<td>Human Subject Research Certified 09/03/2008</td>
</tr>
<tr>
<td>James R. Ewald</td>
<td>Solicit volunteers, contact participants, schedule and conduct testing.</td>
<td>NIH Office of Extramural Research 10/2008 (#120918)</td>
</tr>
<tr>
<td>Kim Topp</td>
<td>Solicit volunteers, contact participants, schedule and conduct testing.</td>
<td>NIH Office of Extramural Research 09/2009 (#83403)</td>
</tr>
<tr>
<td>Roger Baer</td>
<td>Faculty Sponsor 1/19/03</td>
<td></td>
</tr>
</tbody>
</table>

Research Assurances 03/17/2009
FUNDING INFORMATION

| Internally funded, please provide account number: NA | Funding is pending, please provide OSP Record ID on GoldSheet: NA |
| Externally funded, please provide funding source and account number: NA | Title on GoldSheet if different from above: NA |
| Other: (e.g., funding will be applied for later) NA | Student Project—no funding or funding provided by student: Funding provided by student, as well as grants/loan testing equipment. |

SCIENTIFIC REVIEW

Although the assurance committees are not intended to conduct peer review of research proposals, the federal regulations include language such as “consistent with sound research design,” “rationale for involving animals or humans” and “scientifically valuable research,” which requires that the committees consider in their review the general scientific relevance of a research study. Proposals that do not meet these basic tests are not justifiable and cannot be approved. If an assurance review committee(s) has concerns about the scientific merit of a project and the project was not competitively funded by peer review or was funded by corporate sponsors, the project may be referred to a scientific review committee. The scientific review committee will be ad hoc and will consist of your ISU peers and outside experts as needed. If this situation arises, the PI will be contacted and given the option of agreeing that a consultant may be contacted or withdrawing the proposal from consideration.

☐ Yes ☐ No Has or will this project receive peer review?

If the answer is “yes,” please indicate who did or will conduct the review: Research for dissertation, will produce journal articles, conference papers and presentations, etc., in industry publications and conferences.

If a review was conducted, please indicate the outcome of the review:

NOTE: RESPONSE CELLS WILL EXPAND AS YOU TYPE AND PROVIDE SUFFICIENT SPACE FOR YOUR RESPONSE.

COLLECTION OR RECEIPT OF SAMPLES

Will you be: (Please check all that apply.)

☐ Yes ☐ No Receiving samples from outside of ISU? See examples below.

☐ Yes ☐ No Sending samples outside of ISU? See examples below.

Examples include: genetically modified organisms, body fluids, tissue samples, blood samples, pathogens.

If you will be receiving samples from or sending samples outside of ISU, please identify the name of the outside organization(s) and the identity of the samples you will be sending or receiving outside of ISU:

NA

Please note that some samples may require a USDA Animal Plant Health Inspection Service (APHIS) permit, a USPHS Centers for Disease Control and Prevention (CDC) Import Permit for Etiologic Agents, a Registration for Select Agents, High Consequence Livestock Pathogens and Toxins or Listed Plant Pathogens, or a Material Transfer Agreement (MTA).

☐ Yes ☐ No Does this project involve human research participants?

☐ Yes ☐ No Does this project involve human cell lines or tissue culture (primary OR immortalized), or human blood components, body fluid or tissues?

ASSURANCE

Research Assurances 03/17/2009
• I certify that the information provided in this application is complete and accurate and consistent with any proposal(s) submitted to external funding agencies.
• I agree to provide proper surveillance of this project to ensure that the rights and welfare of the human subject or welfare of animal subjects are protected. I will report any problems to the appropriate assurance review committee(s).
• I agree that I will not begin this project until receipt of official approval from all appropriate committee(s).
• I agree that modifications to the originally approved project will not take place without prior review and approval by the appropriate committee(s), and that all activities will be performed in accordance with all applicable federal, state, local and Iowa State University policies.

CONFLICT OF INTEREST

A conflict of interest can be defined as a set of conditions in which an investigator's or key personnel's judgment regarding a project (including human or animal subject welfare, integrity of the research) may be influenced by a secondary interest (e.g., the proposed project and/or a relationship with the sponsor). ISU's Conflict of Interest Policy requires that investigators and key personnel disclose any significant financial interests or relationships that may present an actual or potential conflict of interest. By signing this form below, you are certifying that all members of the research team, including yourself, have read and understand ISU's Conflict of Interest policy as addressed by the ISU Faculty Handbook (http://www.provost.iastate.edu/faculty) and have made all required disclosures.

☐ Yes ☐ No Do you or any member of your research team have an actual or potential conflict of interest?
☐ Yes ☐ No If yes, have the appropriate disclosure form(s) been completed?

SIGNATURES

Signature of Principal Investigator: Date: 1/28/09

Signature of Department Chair: Date: 1/28/09

Major Professor/Supervising Faculty: Please sign cover page.

PLEASE NOTE: Any changes to an approved protocol must be submitted to the appropriate committee(s) before the changes may be implemented.

Please proceed to SECTION II.
SECTION II: IRB SECTION - STUDY SPECIFIC INFORMATION

STUDY OBJECTIVES

Briefly explain in language understandable to a layperson the specific aim(s) of the study.

This research study is about websites' perceived usability. The purpose of the research is to discover a new methodology for usability testing incorporating qualitative research.

BENEFITS TO SOCIETY AND PARTICIPANTS

Explain in language understandable to a layperson how the information gained in this study will advance knowledge, and/or serve the good of society. Please also describe the direct benefits to research participants; if there are no direct benefits to participants, indicate that. Note: monetary compensation cannot be considered a benefit to participants.

Participation in this research will not benefit them directly. However, the knowledge, information, and data gathered in this research will help future website designers with the design of web systems that are more usable.

PART A: PROJECT INVOLVEMENT

1) □ Yes ☒ No Is this project part of a Training, Center, Program Project Grant?
   Director Name: NA Overall IRB ID: NA

2) ☒ Yes □ No Is the purpose of this project to develop survey instruments?

3) □ Yes ☒ No Does this project involve an investigational new drug (IND)? Number:

4) □ Yes ☒ No Does this project involve an investigational device exemption (IDE)? Number:

5) □ Yes ☒ No Does this project involve existing data or records?

6) □ Yes ☒ No Does this project involve secondary analysis?

7) □ Yes ☒ No Does this project involve pathology or diagnostic specimens?

8) □ Yes ☒ No Does this project require approval from another institution? Please attach letters of approval.

9) □ Yes ☒ No Does this project involve DEXA/CT scans or X-rays?

PART B: MEDICAL HEALTH INFORMATION OR RECORDS

1) □ Yes ☒ No Does your project require the use of a health care provider's records concerning past, present, or future physical, dental, or mental health information about a subject? The Health Insurance Portability and Accountability Act established the conditions under which protected health information may be used or disclosed for research purposes. If your project will involve the use of any past or present clinical information about someone, or if you will add clinical information to someone's treatment record (electronic or paper) during the study, you must complete and submit the Application for Use of Protected Health Information.

PART C: ANTICIPATED ENROLLMENT

Estimated number of participants contacted to reach required enrollment: Phase I: 20 Phase II: 100

<table>
<thead>
<tr>
<th>Number of participants to be enrolled in the study</th>
<th>Phase I: 1-10</th>
<th>Phase II: 25</th>
<th>Males:</th>
<th>Females:</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Minors (Under 18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Pregnant Women/Infants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Cognitively Impaired</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Prisoners</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List estimated percent of the anticipated enrollment that will be minorities if known:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

American Indian: Alaskan Native:  

Research Assurance 03/17/2009
<table>
<thead>
<tr>
<th>Asian or Pacific Islander:</th>
<th>Black or African American:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latino or Hispanic:</td>
<td></td>
</tr>
</tbody>
</table>

**PART D: PARTICIPANT SELECTION**

Please use additional space as necessary to adequately answer each question.

11. Explain the procedures for selecting participants including the inclusion/exclusion criteria and how participants will be contacted or recruited (i.e., Where will the names come from? Will a sample be purchased, will ads, flyers, word of mouth, email list, etc., be used?).

   **Phase I:**
   - Word of mouth, e-mails and personal referrals will be used. Individuals familiar with the current standards of usability testing. Solicited from HCI faculty and students.

   **Phase II:**
   - Word of mouth, e-mails and personal referrals will be used to identify a pool of possible participants. Any individual is eligible as long as they are over eighteen years of age. We are excluding individuals younger than eighteen as they are not normally shoppers at e-commerce websites due to the ability to secure payment (i.e., Credit cards, PayPal, etc.) Once identified, these possible participants will provide an email address if they are interested. From these, a random sample of 30 individuals will be selected to participate in the study from the pool of possible participants.

12. Attach a copy of any recruitment telephone scripts or materials such as ad, flyers, e-mail messages, etc. Recruitment material must include a statement of the voluntary and confidential nature of the research.

**Note:** Please answer each question. If the question does not pertain to this study, please type not applicable (N/A).

**PART E: RESEARCH PLAN**

Include sufficient detail for IRB review of this project independent of the grant, protocol, or other documents.

13. The information needed here is similar to that in the "methods" or "procedures" sections of a research proposal—it should describe the flow of events that will occur during your interactions with subjects. Please describe in detail your plans for collecting data from participants, including all procedures, tasks, or interventions participants will be asked to complete during the research (e.g., random assignment, any conditions or treatment groups into which participants will be divided, mail survey or interview procedures, sensors to be worn, amount of blood drawn, etc.). This information is intended to inform the committee of the procedures used in the study and their potential risk. Please do not respond with "see attached" or "not applicable."

   **Phase I:**
   - Ia: Identify volunteers using approved methods
   - Ib: Set-up testing time
   - Ic: Complete user testing using steps outlined below for Phase I & II
   - Id: Review testing methodology & submit addendums to IRB if necessary before proceeding to Phase II.

   **Phase II:**
   - IiA: Identify volunteers using approved methods
   - IIb: Set-up testing time
   - IIc: Complete user testing using steps outlined below for Phase I & II
   - IId: Compile and review results.
   - IIe: Publish results and findings.

   Testing for Phase I & II:
   1. Brief test subject and review letter of consent.
   2. Brief test subject on test using approved script.
   3. Complete pre-test survey using approved website form.
   4. Complete task cards according to approved script.
   5. Capture audio/video of test subject during task, including video recording of eye movements.
   6. Complete post-test survey using approved website form.
   7. Annotate data and proceed with appropriate Phase.

Research Assurances 03/17/2009

5
14. For studies involving pathology/diagnostic specimens, indicate whether specimens will be collected prospectively and/or already exist “on the shelf” at the time of submission of this review form. If prospective, describe specimen procurement procedures; indicate whether any additional medical information about the subject is being gathered, and whether specimens are linked at any time by code number to the participant’s identity. If this question is not applicable, please type N/A in the response cell.

| NA |

15. For studies involving deception, please justify the deception and indicate the debriefing procedure, including the timing and information to be presented to participants. If this question is not applicable, please type N/A in the response cell.

| NA |

**PART F: CONSENT PROCESS**

16. Describe the consent process for adult participants (those who are age 18 and older). If the consent process does not include documented consent, a waiver of documentation of consent must be requested.

The letter of information / consent will be emailed to the randomly identified participant and they will be given 24 hours to review and respond. If they respond and indicate willingness to participate, a date and time will be established to conduct the testing and it will be assumed that he/she had reviewed the letter of information and is willing to participate. Additionally on the date of testing, the participant will be given a hard-copy of the letter of information for review and prior to the start of testing.

| SIGNATURE |

17. If your study involves minors, please explain how parental consent will be obtained prior to enrollment of the minor(s).

| NA |

18. Please explain how assent will be obtained from minors (younger than 18 years of age), prior to their enrollment. Also, please explain if the assent process will be documented (e.g., a simplified version of the consent form, combined with the parental informed consent document). According to the federal regulations, assent... means a child’s affirmative agreement to participate in research. More failure to object should not, absent affirmative agreement, be construed as assent."

| NA |

**PART G: DATA ANALYSIS**

19. Describe how the data will be analyzed (e.g., statistical methodology, statistical evaluation, statistical measures used to evaluate results).

The data obtained will be analyzed utilizing Excel, SPSS and JMP to help identify significant relationships between data sets and data items. Statistical tests including: ANOVA, T-Tests, Chi-Squared, as well as others will be used to help extract the most descriptive data.

20. If applicable, please indicate the anticipated date that identifiers will be removed from completed survey instruments and/or audio or visual tapes will be erased:

| 08/01/10 Month/Day/Year |

**PART H: RISKS**

Research Assurances 03/17/2009
The concept of risk goes beyond physical risk and includes risks to participants' dignity and self-respect as well as psychological, emotional, legal, social or financial risk.

21. □ Yes  ☒ No  Is the probability of the harm or discomfort anticipated in the proposed research greater than that encountered ordinarily in daily life or during the performance of routine physical or psychological examinations or tests?

22. □ Yes  ☒ No  Is the magnitude of the harm or discomfort greater than that encountered ordinarily in daily life, or during the performance of routine physical or psychological examinations or tests?

23. Describe any risks or discomforts to the participants and how they will be minimized and precautions taken. Do not respond with N/A. If you believe that there will not be risk or discomfort to participants, you must explain why.

NA  See §[42 premium award for expansion of why there are no foreseeable risks. (5)}

24. If this study involves vulnerable populations, including minors, pregnant women, prisoners, the cognitively impaired, or those educationally or economically disadvantaged, what additional protections will be provided to minimize risks?

NA

PART I: COMPENSATION

25. □ Yes  ☒ No  Will participants receive compensation for their participation? If yes, please explain.

Do not make the payment an inducement, only a compensation for expenses and inconvenience. If a person is to receive money or another token of appreciation for their participation, explain when it will be given and any conditions of full or partial payment. (E.g., volunteers will receive $5.00 for each of the five visits in the study or a total of $25.00 if both complete the study. If a participant withdraws from participation, they will receive $5.00 for each of the visits completed.) It is considered undue influence to make completion of the study the basis for compensation.

NA

PART II: CONFIDENTIALITY

26. Describe below the methods that will be used to ensure the confidentiality of data obtained. (For example, who has access to the data, where the data will be stored, security measures for web-based surveys and computer storage, how long data or specimens will be retained, etc.)

There is no identifier in the data gathering instruments and the participant’s identity will be anonymous all throughout the survey. Only the researcher will have access to the data. The data will be entered and kept in a password-protected computer located on the PI’s computer. The survey instruments including questionnaires will be shredded after all the information is entered into the computer for data analysis. All data, including names and video recordings, will be kept confidential. Some video clips may be published but names PART II: REGISTRY PROJECTS will not be included and participants are informed. See §[7343401] To be considered a registry: (1) the individuals must have a common condition or demonstrate common responses to questions; (2) the individuals in the registry might be contacted in the future; and (3) the names/data of the individuals in the registry might be used by investigators other than the one maintaining the registry.

□ Yes  ☒ No  Does this project establish a registry?

If “yes,” please provide the registry name below.

Research Assurances 03/17/2009
Checklist for Attachments

Listed below are the types of documents that should be submitted for IRB review. Please check and attach the documents that are applicable for your study:

☑ A copy of the informed consent document OR ☐ Letter of introduction containing the elements of consent
☐ A copy of the assent form if minors will be enrolled
☐ Letter of approval from cooperating organizations or institutions allowing you to conduct research at their facility
☑ Data-gathering instruments (including surveys)
☐ Recruitment fliers, phone scripts, or any other documents or materials participants will see or hear

The original signed copy of the application form and one set of accompanying materials should be submitted for review. **Federal regulations require that one copy of the grant application or proposal be submitted for comparison with the application for approval.**

FOR IRB USE ONLY:

Initial action by the Institutional Review Board (IRB):

☐ Project approved. Date: [May 30, 2009]
☐ Pending further review. Date: 
☐ Project not approved. Date: 

Follow-up action by the IRB:

[Signature]

[May 30, 2009]

SECTION III: ENVIRONMENTAL HEALTH AND SAFETY INFORMATION

☐ Yes ☑ No Does this project involve human cell or tissue cultures (primary OR immortalized), or human blood components, body fluids or tissues?

PART A: HUMAN CELL LINES

☐ Yes ☑ No Does this project involve human cell or tissue cultures (primary OR immortalized cell lines/strains) that have been documented to be free of bloodborne pathogens? If the answer is "yes," please answer question 1 below and attach copies of the documentation.

1) Please list the specific cell lines/strains to be used, their source and description of use.

<table>
<thead>
<tr>
<th>CELL LINE</th>
<th>SOURCE</th>
<th>DESCRIPTION OF USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Add New Row

2) Please refer to the ISU "Bloodborne Pathogens Manual," which contains the requirements of the OSHA Bloodborne Pathogens Standard. Please list the specific precautions to be followed for this project below (e.g., retractable needles used for blood draws):

NA

Anyone working with human cell lines/strains that have not been documented to be free of bloodborne pathogens is required to have Bloodborne Pathogen Training annually. Current Bloodborne Pathogen Training dates must be

Research Assurances 03/17/2009
Hello! I am working on my PhD dissertation project which involves researching the usability of e-commerce websites. I would like to extend an invitation to you to help me complete my research. This phase of the research involves refinement of my methodology.

As a participant in my research, you will fill out two surveys and complete a total of five tasks while using a large e-commerce website. It will take approximately 15-20 minutes to complete the entire process.

I would like to clarify that your participation is completely voluntary. You may refuse to answer any question that seems too personal, or you may stop the interview/testing at any time. The survey results will be kept strictly confidential. Your actions will be captured via video and audio, and these data streams may be used in publications/presentations to validate the method of testing. We will be asking for your age, occupation, educational background, internet experience, website opinions, etc.

If you are interested in participating, please reply to this e-mail or contact me within the next seven (7) days at:

Troy Abel, Primary Investigator
e-mail: tabel@lastate.edu
phone: 563-650-7478

Roger Baer, Faculty Advisor POS Chair
e-mail: roger@lastate.edu
phone: 515-594-0724

If you are randomly selected to participate in our research, you will receive a follow-up e-mail approximately 48 hours after you express interest.

Thank you for your time and consideration.

Sincerely,
Troy Abel
Hello! I am working on my PhD dissertation project which involves researching the usability of e-commerce websites. I would like to extend an invitation to you to help me complete my research. This phase of the research involves refinement of my methodology.

As a participant in my research, you will fill out two surveys and complete a total of five tasks while using a large e-commerce website. It will take approximately 15-20 minutes to complete the entire process.

I would like to clarify that your participation is completely voluntary. You may refuse to answer any question that seems too personal, or you may stop the interview/testing at any time. The survey results will be kept strictly confidential. Your actions will be captured via video and audio, and these data streams may be used in publication / presentations to validate the method of testing. We will be asking for your age, occupation, educational background, internet experience, website opinions, etc.

If you are interested in participating, please reply to this e-mail or contact me within the next seven (7) days at:

Troy Abel, Primary Investigator
e-mail: tabel@iastate.edu
phone: 563-650-7478

Roger Jarenski, Faculty Advisor/PI Chair
e-mail: rogerj@iastate.edu
phone: 515-294-6724

If you are randomly selected to participate in our research, you will receive a follow up e-mail approximately 48 hours after you express interest.

Thank you for your time and consideration.

Sincerely,

Troy Abel
Thank you for agreeing to participate in my PhD dissertation research! Attached to this e-mail is a letter of information for you to review. Again, you may choose to withdraw from participating at any time without penalty.

In addition to this, please provide a list of times that you will be available for testing. Send this information to me via e-mail.

Troy Abel, Principle Investigator
e-mail: tabeli@iastate.edu
phone: 563-650-7478

Roger Baker, Faculty Advisor POS Chair
e-mail: rebaker@iastate.edu
phone: 515-294-6724

If you have any questions or concerns, feel free to contact me at any time.

Thank you!
INFORMED CONSENT DOCUMENT

Title of Study: E-commerce usability testing methodology enhancement
Investigator: Troy Abel, AA, BA, BSFA, MFA
Faculty Advisor: Roger Baer, BFA, MFA

This is a research study. Please take your time in deciding if you would like to participate. Please feel free to ask questions at any time.

INTRODUCTION
The purpose of this study is to define and test a new website usability testing methodology.

DESCRIPTION OF PROCEDURES
If you agree to participate in this study, your participation will last for fifteen to twenty (15-20) minutes. You will be given a set of tasks to complete. As you interact with the website, your mouse movements, audio/visual information, and eye tracking data will be recorded. You will also be asked demographic information prior to the tests, and a follow-up survey will be completed when testing ends. You may skip any question you do not wish to answer or that makes you feel uncomfortable.

Your video recordings, facial, may be published to validate and describe the new testing methodology; however, your name will never be published in any presentations or publications. Any audio/video recordings not used in publication or presentations will be destroyed by August 8, 2010.

RISKS
There are no foreseeable risks in this study. All audio/video/eye tracking data will be collected transparent to you as a participant. No equipment will be affixed to you in any way.

BENEFITS
This survey and test will not yield any direct benefit to you. However, the knowledge and information gathered in this research will help future research in designing website experiences.

COSTS AND COMPENSATION
You will not be compensated for participating in this study and there is no cost in participating in this survey.

PARTICIPANT RIGHTS
Your participation in this study is completely voluntary and you may refuse to participate or leave the study at any time. If you decide to not participate in the study or
CONFIDENTIALITY
Records which expose a participant’s identity will be kept confidential to the extent permitted by applicable laws and regulations. However, federal government regulatory agencies, auditing departments of Iowa State University, Noldus Corporation (manufacturers of the emotional reading software), and the Institutional Review Board (a committee that reviews and approves human subject research studies) may inspect and/or copy your records for quality assurance and data analysis and training purposes. These records may contain private information which may identify you.

Video recordings and screen captures may be used in publications and presentations; however, at no time will your name be associated with these images. Please be advised that your full-facial image may be published, but your name will never be. However, theoretically your identity could be deduced by identification of your full-facial image in publication or presentation.

There is no identifier in the surveys and only the researchers and those entities listed above will have access to your identity. The data will be entered and kept in a password protected computer located at the PI’s office in the College of Design and any originals shredded.

QUESTIONS OR PROBLEMS
You are encouraged to ask questions at any time during this study.

• For further information about the study contact Troy Abel, 563-650-7478, tabel@iastate.edu, or Roger Baer, 515-294-672, rebaer@iastate.edu.

• If you have any questions about the rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, (515) 294-3115, Office of Research Assurances, Iowa State University, Ames, Iowa 50011.

PARTICIPANT SIGNATURE
Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent prior to your participation in the study.

Participant’s Name (printed)

(Participant’s Signature) (Date)
CONFIDENTIALITY
Records identifying participants will be kept confidential to the extent permitted by applicable laws and regulations and will not be made publicly available. However, federal government regulatory agencies, auditing departments of Iowa State University, Noldus Corporation (manufacturers of the emotional reading software), and the Institutional Review Board (a committee that reviews and approves human subject research studies) may inspect and/or copy your records for quality assurance and data analysis and training purposes. These records may contain private information.

To ensure confidentiality to the extent permitted by law, the following measures will be taken. There is no identifier in the surveys and the participant's identity will remain confidential throughout the testing. Once the researchers and those entities listed above will have access to the data, the data will be entered and kept in a password protected computer located at the PI's office in the College of Design and any originals shredded. If the results are published, your identity will remain confidential.

QUESTIONS OR PROBLEMS
You are encouraged to ask questions at any time during this study.

*For further information about the study contact [investigator name and phone number; for a student project list the name of the major professor or supervising faculty member's name and contact information].

*If you have any questions about your rights of research subjects or research-related injury, please contact the IRB Administrator, (515) 294-4566, IRB@iastate.edu, or Director, (515) 294-3115, Office of Research Assurances, Iowa State University, Ames, Iowa 50011.

PARTICIPANT SIGNATURE
Your signature indicates that you voluntarily agree to participate in this study, that the study has been explained to you, and that you have been given the time to read the document and that your questions have been satisfactorily answered. You will receive a copy of the written informed consent prior to your participation in the study.

Participant's Name (printed)  

(Participant's Signature)  

(Date)
Welcome! Thank you for coming today to help me conduct PhD research on website usability. Please remember that I am researching website usability and not you. If you feel uncomfortable or change your mind about participating you may withdraw from the test at anytime.

Before, during and possibly after there will be short questionnaires for you to fill out. Please remember you may skip any question that makes you uncomfortable in any way.

The test will consist of you completing a set of five (5) - ten (10) tasks via interaction with a website. You will be given 10 minutes to complete as many tasks as possible. You will complete these tasks in any order you wish. During your testing, audio, video, mouse movements and eye movement data will be captured. This will all be transparent to you. I urge you to think out loud during this testing. Your comments will further enhance this study.

If you struggle with a task, don't worry! Just try everything you can and work at a normal pace. Do not feel rushed! You are not expected to complete all 5 - 10 tasks. If you get stuck, I may offer assistance or ask you move on to the next task. You will determine when each task is complete to your satisfaction, and move onto the next task. I will not provide you with confirmation that a task is complete. Again, remember you are not expected to complete all five (5) - ten (10) tasks.

Do you have any questions before we begin? Remember to verbally think out loud as much as possible. I may ask you what you are thinking during your testing.

Please complete this pre-test survey. [Wait for survey to complete]

Thanks! Now here are your task cards. Based on your initial impression of the website you just saw, take these cards (task subject task cards) and put them in order for me from easiest to hardest. [Allow subject to sort.]

Great! Now when the timer onscreen reaches zero you’ll have 10 minutes to complete as many tasks as possible working in the order you just determined.

Do you have any questions?

[Begin timer, testing starts]

[When finished] Again, thanks for helping me with my research. Please feel free to contact me if you have any questions.
PhD Usability Research

Please complete this survey. Remember, you can skip any question you want.

Age Range
- 18-26
- 20-29
- 30-39
- 40-49
- 50-59
- 60-69
- 70+ years

Occupation
- Industry
- Academic
- Student
- Other

Educational Background
- Grade School
- High School
- College
- Bachelor's or Terminal degree
- Other advanced degree

Is English your first language?
- Yes
- No

How often do you use the Internet?
- Almost every day
- One to three times per week
- Once or twice a month
- Less than once a month
- Rarely use it or before
troy d. abel
phd dissertation
research

[Image of a form with questions and options]

Is English your first language?
- Yes
- No

How much do you use the Internet?
- Almost every day
- On average 1-2 hours per week
- Browse or twice a month
- Less than once a month
- I've never used it before

What is your overall comfort level using the Internet?
- Very comfortable
- Somewhat
- Not at all
- I've never used it before

Have you ever used LastLook.com?
- Yes
- No

If "YES" please state your previous interactions with LastLook.com:
1 2 3 4 5
Extremely Dislike \(\square\) Like \(\square\) Neutral \(\square\)\n
If "YES" please describe your likes/dislikes with LastLook.com. (Feel free to just list objectives/descriptive words rather than rates)
Locate a store using this zip code: 50010

Locate the contact e-mail page.

Phase I & II:
Task Cards
Locate the Help page.

Locate the Sign-in page.

Phase I & II:
Task Cards
Locate the sale ad for this week.

Locate the gift card terms and conditions.

Phase I & II
Task Cards
Find all Outlet-Center TV's

Find information on the Low Price Guarantee.

Phase I & II: Task Cards
Add (1) Apple - MacBook Air 13.3" Display and (1) Apple® - MagSafe 60W Power Adapter to your cart.

Empty the contents of your cart.

Find the 'Top Rated' gifts for Boys.
Post Test

Please complete this survey. Remember, you can skip any questions you want.

Thinking of the tasks you completed, which was the easiest?

Why was this the easiest?

Thinking of ALL tasks, which is the most difficult?

Why was this the most difficult of ALL tasks?

How did the websites COLOR affect your testing today?

1 2 3 4 5
<table>
<thead>
<tr>
<th>How did the websites COLOR affect your testing today?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Terrible COLOR: ○ ○ ○ ○ ○ Perfect COLOR</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How did the websites GRAPHICS affect your testing today?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Terrible GRAPHICS: ○ ○ ○ ○ ○ Perfect GRAPHICS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How did the websites LAYOUT affect your testing today?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Terrible LAYOUT: ○ ○ ○ ○ ○ Perfect LAYOUT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How did the websites &quot;OVERALL FEEL&quot; affect your testing today?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>Terrible &quot;OVERALL FEEL&quot;: ○ ○ ○ ○ ○ Perfect &quot;OVERALL FEEL&quot;</td>
</tr>
</tbody>
</table>

---

Overall, how DIFFICULT was it to complete the tasks on this website?  
1 2 3 4 5  
Extremely DIFFICULT: ○ ○ ○ ○ ○ Extremely EASY  

In your own words, please describe the NEGATIVE aspects of this website.

---

Phase I & II:  
User Profile Test  
Administered via online Google Form. Screen-shots provided.
Overall, how DIFFICULT was it to complete the tests on this website?
1 2 3 4 5
Extremely DIFFICULT 〇 〇 〇 〇 Extremely EASY

In your own words, please describe the NEGATIVE aspects of this website.

In your own words, please describe the POSITIVE aspects of this website.

Use the space provided to express any additional concerns / comments about today's test.

Powered by Google Forms
Terms of Service – Additional Terms
APPENDIX C: PUT-Q2 SAMPLE SUBJECT MATRICES

PUT-Q²
PERCEPTION = QUALITATIVE ANALYSIS

09-F
subject matrix

Task Tree

Task Descriptions
1. Love the design page.
2. Store it in the HTW hiring cycle.
3. Locate the correct job page.
4. Find information on the CV page.
5. Find information on the PUT-Q².
6. Find the navigation bar.
7. Find all the other page bars.
8. Locate the different page bars.
9. Add all the different page bars.
10. ADD all the different page bars to the task.

Population Statistics

Other Qualitative Data

PUT-Q2 Subject Data Chart

Subject 09-F

Areas of Interest:

Rating | Sum | Raw

1. Brandon 0 0 0
2. John 0 0 0
3. Sarah 5 5 5
4. Fred 4 4 4
5. Elizabeth 1 1 1
6. Joseph 1 1 1
7. Kevin 1 1 1
8. William 1 1 1

Grid: E-F

1. My impression is too much.
2. I want to know the basic page, there is too much. It seems that the page is too heavy, but it has many many tools.
3. All the problems are related to the navigation bar, and it is too heavy. It is not easy to use at all.
4. It looks heavy and difficult.
5. [None selected] It would be new.
### PUT-Q²

**Perception + Usability Testing / Qualitative + Quantitative Analysis**

#### Task Tree

- **A**: 1.04
- **B**: 0.89
- **C**: 0.75
- **D**: 0.53
- **E**: 0.47
- **F**: 0.33
- **G**: 0.20

#### Task Descriptions
- Locate the user on the page.
- Locate the contact on the page.
- Locate the contact on the page.
- Locate the contact on the page.
- Locate the contact on the page.
- Locate the contact on the page.
- Locate the contact on the page.
- Locate the contact on the page.

#### Post-Test
- Click:
  - [ ]
- Graph:
  - [ ]
- Layout:
  - [ ]
- Overall Feel:
  - [ ]

#### Other Qualitative Data

<table>
<thead>
<tr>
<th>Measure</th>
<th>A 9</th>
<th>B 3</th>
<th>C 2</th>
<th>D 5</th>
<th>E 3</th>
<th>F 9</th>
<th>G 7</th>
<th>H 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender Male/Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>2.29</td>
<td>2.49</td>
<td>2.28</td>
<td>2.83</td>
<td>2.38</td>
<td>2.83</td>
<td>2.38</td>
<td>2.83</td>
</tr>
<tr>
<td>Error Rate (%)</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
</tr>
<tr>
<td>C.R. %</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
<td>20.89</td>
</tr>
<tr>
<td>Cardinal Position</td>
<td>47</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>Position Mean</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
</tbody>
</table>

#### PUT-Q2 Subject Data Chart

**Subject: 19-F**

Key of Answers:
- [ ] I strongly disagree
- [ ] I disagree
- [ ] Neutral
- [ ] I agree
- [ ] I strongly agree

**1. My impressions of the website:**
- [ ] I strongly disagree
- [ ] I disagree
- [ ] Neutral
- [ ] I agree
- [ ] I strongly agree

**2. The layout and design are easy to follow:**
- [ ] I strongly disagree
- [ ] I disagree
- [ ] Neutral
- [ ] I agree
- [ ] I strongly agree

**3. The website is visually appealing:**
- [ ] I strongly disagree
- [ ] I disagree
- [ ] Neutral
- [ ] I agree
- [ ] I strongly agree

**4. The content is presented in a clear and organized manner:**
- [ ] I strongly disagree
- [ ] I disagree
- [ ] Neutral
- [ ] I agree
- [ ] I strongly agree

**5. I would recommend this website to others:**
- [ ] I strongly disagree
- [ ] I disagree
- [ ] Neutral
- [ ] I agree
- [ ] I strongly agree
PUT-Q²
PERCEPTION + USABILITY TESTING / QUALITATIVE + QUANTITATIVE ANALYSIS

// TASK TREE
A 0.15
C 1.28
B 2.07
G 0.53
F 0.37
D 0.16
E 1.14
H 0.25

// POPULATION STATISTICS
<table>
<thead>
<tr>
<th>Measure</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic Mean</td>
<td>17.36</td>
<td>55.40</td>
<td>37.73</td>
<td>66.13</td>
<td>77.57</td>
<td>62.26</td>
<td>61.14</td>
<td>115.23</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>12.79</td>
<td>49.72</td>
<td>34.89</td>
<td>59.85</td>
<td>61.83</td>
<td>53.19</td>
<td>28.67</td>
<td>39.68</td>
</tr>
<tr>
<td>Percentile</td>
<td>40.16</td>
<td>62.26</td>
<td>41.87</td>
<td>64.37</td>
<td>67.47</td>
<td>54.94</td>
<td>43.40</td>
<td>80.67</td>
</tr>
<tr>
<td>Q1-Q3</td>
<td>28.87</td>
<td>54.94</td>
<td>43.40</td>
<td>64.37</td>
<td>67.47</td>
<td>54.94</td>
<td>43.40</td>
<td>80.67</td>
</tr>
<tr>
<td>Percentile</td>
<td>30.62</td>
<td>38.62</td>
<td>36.47</td>
<td>43.40</td>
<td>50.16</td>
<td>54.94</td>
<td>43.40</td>
<td>58.82</td>
</tr>
<tr>
<td>Overall Ease</td>
<td>5.00</td>
<td>7.50</td>
<td>5.00</td>
<td>7.50</td>
<td>5.00</td>
<td>7.50</td>
<td>5.00</td>
<td>7.50</td>
</tr>
</tbody>
</table>

// OTHER QUALITATIVE DATA

PUT-Q² Subject Data Chart

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

Subject: 20-F

// POST-TEST

1. The feature is easy to use.
2. The feature is easy to understand.
3. The feature is easy to learn.
4. The feature is easy to use with my natural language.
5. My experience was good.
6. I am satisfied with the feature.
7. I am likely to use the feature again.
8. I would recommend the feature to others.
9. I would use the feature often.
10. I would continue using the feature.

// SIDE DESCRIPTIONS
A) Locate the icons in the page
B) Locate the text in the page
C) Locate the content in the page
D) Read the information on the page
E) Find the top-level goal for the task
F) Find the main goal for the task
G) Locate the web page in the task
H) Navigate the web page

// POST-TEST

1. The feature is easy to use.
2. The feature is easy to understand.
3. The feature is easy to learn.
4. The feature is easy to use with my natural language.
5. My experience was good.
6. I am satisfied with the feature.
7. I am likely to use the feature again.
8. I would recommend the feature to others.
9. I would use the feature often.
10. I would continue using the feature.

// TASK TREE
A 0.15
C 1.28
B 2.07
G 0.53
F 0.37
D 0.16
E 1.14
H 0.25

// POPULATION STATISTICS
<table>
<thead>
<tr>
<th>Measure</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic Mean</td>
<td>17.36</td>
<td>55.40</td>
<td>37.73</td>
<td>66.13</td>
<td>77.57</td>
<td>62.26</td>
<td>61.14</td>
<td>115.23</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>12.79</td>
<td>49.72</td>
<td>34.89</td>
<td>59.85</td>
<td>61.83</td>
<td>53.19</td>
<td>28.67</td>
<td>39.68</td>
</tr>
<tr>
<td>Percentile</td>
<td>40.16</td>
<td>62.26</td>
<td>41.87</td>
<td>64.37</td>
<td>67.47</td>
<td>54.94</td>
<td>43.40</td>
<td>80.67</td>
</tr>
<tr>
<td>Q1-Q3</td>
<td>28.87</td>
<td>54.94</td>
<td>43.40</td>
<td>64.37</td>
<td>67.47</td>
<td>54.94</td>
<td>43.40</td>
<td>80.67</td>
</tr>
<tr>
<td>Percentile</td>
<td>30.62</td>
<td>38.62</td>
<td>36.47</td>
<td>43.40</td>
<td>50.16</td>
<td>54.94</td>
<td>43.40</td>
<td>58.82</td>
</tr>
<tr>
<td>Overall Ease</td>
<td>5.00</td>
<td>7.50</td>
<td>5.00</td>
<td>7.50</td>
<td>5.00</td>
<td>7.50</td>
<td>5.00</td>
<td>7.50</td>
</tr>
</tbody>
</table>
### PUT-Q² Subject Matrix

**Subject:** 24-M

#### Task Tree

<table>
<thead>
<tr>
<th>Task</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.01</td>
</tr>
<tr>
<td>B</td>
<td>0.32</td>
</tr>
<tr>
<td>C</td>
<td>2.06</td>
</tr>
<tr>
<td>D</td>
<td>0.64</td>
</tr>
<tr>
<td>E</td>
<td>1.61</td>
</tr>
<tr>
<td>F</td>
<td>0.51</td>
</tr>
<tr>
<td>G</td>
<td>2.09</td>
</tr>
<tr>
<td>H</td>
<td>1.52</td>
</tr>
</tbody>
</table>

#### True Descriptions
- AL: enter the page on the left
- BE: click the last option in the left pane
- CF: click the load button
- DH: click the last item on the list
- EF: click on the top row of the table
- FG: click on the left hand side of the screen
- GH: click on the right hand side of the screen
- HI: click on the bottom of the list

#### Population Statistics

<table>
<thead>
<tr>
<th>Measure</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>12.2</td>
<td>7</td>
<td>21.2</td>
<td>11</td>
<td>28.2</td>
<td>6</td>
<td>26.1</td>
<td>0</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.79</td>
<td>0.72</td>
<td>0.69</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>Score Rating</td>
<td>54.10</td>
<td>40.00</td>
<td>54.00</td>
<td>54.00</td>
<td>54.00</td>
<td>54.00</td>
<td>54.00</td>
<td>54.00</td>
</tr>
<tr>
<td>Score</td>
<td>83.32</td>
<td>82.80</td>
<td>83.32</td>
<td>83.32</td>
<td>83.32</td>
<td>83.32</td>
<td>83.32</td>
<td>83.32</td>
</tr>
</tbody>
</table>

#### Other Qualitative Data

- Overall Difficulty: 85%
- Layout: Standard
- Content: 70%
- Overall Feel: 80%
- Expectation: Good

---

### PUT-Q² Subject Data Chart

**Subject:** 24-M

#### Key Concept Descriptions

<table>
<thead>
<tr>
<th>Key Concept</th>
<th>DPA</th>
<th>DBO</th>
<th>DBD</th>
<th>BPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DBO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DBD</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>BPA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Overall Rating: 9/10

- I would recommend this website to others.
- I think the website is engaging and interesting.
- I believe the information is easy to access.
- I think the website is well organized.
- I would recommend this website to others.
### PUT-Q² Perception + Usability Testing / Qualitative + Quantitative Analysis

#### Task 1: Perception

**Task Description:**
- Locate the logo on the homepage.
- Locate the contact page.
- Describe the content and layout.
- Discuss the color scheme.
- Evaluate the overall user experience.
- Note any issues or improvements.

#### Task 2: User Experience

**Task Description:**
- Annotate the user interface with marks indicating areas of improvement.
- Evaluate the design and usability.
- Provide feedback on usability issues.

#### Populations Statistics

<table>
<thead>
<tr>
<th>Category</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
<td>15.86</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
<td>1.20</td>
</tr>
</tbody>
</table>

#### Other Qualitative Data

- **Overall Feedback:**
  - 5: Excellent
  - 4: Good
  - 3: Fair
  - 2: Poor
  - 1: Very Poor

---

#### Subject: 35-M

#### PUT-Q2 Subject Data Chart

- **Ease of Answering:**
  - 6: Very Easy
  - 5: Easy
  - 4: Medium
  - 3: Hard
  - 2: Very Hard

- **Eye Tracking Data:**
  - No Data Available

---

#### Subject Summary

- **Subject Experience:**
  - 5: Excellent
  - 4: Good
  - 3: Fair
  - 2: Poor
  - 1: Very Poor

- **Eye Tracking Use:**
  - Not applicable

- **Self-Reported Feedback:**
  - No issues reported

- **Additional Notes:**
  - None
APPENDIX D: PUT-Q2 POPULATION MATRIX
BIBLIOGRAPHY


Sutherland, IE. “Sketchpad a Man-Machine Graphical Communication System.” SIMULATION 2, no. 5 (1964):


