Department of Computer Science
San Marcos, TX 78666

Report Number TXSTATE-CS-TR-2010-22

AN EFFORT-BASED FRAMEWORK FOR EVALUATING SOFTWARE USABILITY DESIGN

Dan Tamir
Carl J. Mueller
Oleg V. Komogortsev

2010-03-12

AN EFFORT-BASED FRAMEWORK FOR EVALUATING SOFTWARE USABILITY DESIGN

Abstract

One of the major stakeholder complaints is the usability of software applications. Although there is a rich amount of material on good usability design and evaluation practice, software engineers may need an integrated framework facilitating effective quality assessment. A novel element of the framework, presented in this paper, is its effort-based measure of usability providing developers with an informative model to evaluate software quality, validate usability requirements, and identify missing functionality. Another innovative aspect of this framework is its focus on learning in the process of assessing usability measurements and building the evaluation process around Unified Modeling Languages Use Cases. The framework also provides for additional developer feedback through the notion of designer's and expert's effort representing effort necessary to complete a task. In this paper, we present an effort-based usability model in conjunction with a framework for designing and conducting the evaluation. Experimental results provide evidence of the frameworks utility.

Table of Contents

Table of Figures

1 Introduction

The decades that followed the introduction of the concept of interactive user interface along with important computer graphics concepts by Sutherland have produced a proliferation of user interface devices and procedures, as well as an abundance of approaches and methods for evaluating the effectiveness of the interface or its usability (Nielsen 1993; Shneiderman 1998; Sutherland 1963). There are several definitions for software usability. The IEEE standard glossary defines software usability as "the ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component" (IEEE 1990). As the use of advanced interface techniques became a prominent trend and almost a requirement, the importance of the concept of software usability became the cornerstone of user interface design and evaluation. Many hardware and software companies and vendors found themselves either riding the wave of good usability to enormous success or facing a painful failure due to lack of usability. Software usability is one of the most important factors in a user's decision on whether to acquire and use a system as well as a major factor affecting software failures (Chartette 2005; Leveson and Turner 1993).

The research reported in this paper presents a novel framework for enhancing the capability of the HCI community to advance the state of the art of usability research and practice. Under this framework, a set of users executes a set of identical independent tasks, which emerge from a single scenario. While the tasks share a scenario, they differ in important parameters so that the user cannot "just" memorize a sequence of interaction activities. Throughout the interaction process, certain user activities such as eye movement, time on task, keyboard, and mouse activities are logged. The accumulated data is reduced and several metrics that relate to user effort as well as time on task are extracted. Next, the averages of the per task measurement are compared to an expected learning curve which represents the user's mastery as progress through the set of tasks. We show that the average learning curve can improve a multitude of current usability testing and evaluation techniques. First, we show that the *learnability* of software systems can be accurately assessed and the point of the user's mastery can be identified. Furthermore, the learning curve can be used to compare the *operability* and *understandability* of different systems or different groups of users using the same system. We explain how software designers and developers can employ the novel concepts of *designer's* and *expert's effort* in the process of user interface design and evaluation, and demonstrate the use of the effort based metrics to provide interface designers and developers with a methodology to evaluate their designs as they are completed. In summary this paper, presents a framework permitting software engineers to utilize time and effort based measures to validate software usability and compare the usability of similar systems.

The remainder of the paper is organized in the following way. Section 2 provides a review of related work. Section 3 presents a time and effort based usability model. Section 4 presents a usability testing framework and demonstrates the use of the metrics presented in Section 3. Section 5 presents experimental results demonstrating the validly of the usability model and testing framework. Section 6 presents conclusions and future research.

2 Review of Related Work

The Human Computer Interface (HCI) community is involved with all the aspects of usability including definition, standardization, assessment, and measurement along with research and development of new concepts (Andre et al. 2001; Bevan 1999; Blandford et al. 2008; Caulton 2001; Dennerlein and Johnson 2006; Dumas and Redish 1999; Folmer et al. 2003; Hornbaeck

2006; Howarth et al. 2009; John and Kieras 1996; Moore and Fitz 1993; Nielsen 1993; Rubin and Chisnell 2008; Seffah et al. 2006a; Tamir et al. 2008; Tullis and Albert 2008). Research shows that despite the fact that software developers are fully aware of the importance of usability they tend to neglect this aspect during the software development lifecycle (Vukelja et al. 2007). This may be due to confusion about the way to design a usable interface and perform usability tests or lack of precise usability "tools" available to software developers.

Two international standards are addressing the issue of usability, the ISO 9241-11 and the ISO/IEC 9126-1:2001 (ISO 1998; ISO 2001). The ISO 9241-11 standard views software usability from the prospective of the user and provides three main metrics: *satisfaction*, *efficiency*, and *effectiveness*. These characteristics are important because users can relate to them, and assign measurable values to their experience with the software in terms of these attributes. In addition to these three attributes, the ISO 9241-11 standard provides the evaluator with a set of optional usability characteristics. The ISO/IEC 9126-11 standard views software quality from three perspectives; the perspective of the end user, a test engineer, and a developer (ISO 2001). An end user's perspective of software quality includes *effectiveness*, *productivity*, *satisfaction*, and *safety*. For a test engineer, software quality is composed of externally visible quality characteristics describing *functionality, reliability, efficiency,* and *usability*. Internal quality or the developer's perspective of quality in the ISO/IEC 9126 includes *maintainability* and *portability*. Evaluating software usability using the ISO/IEC 9126 standard requires evaluating both the quality in use and the external quality. The ISO/IEC 9126 defines the following sub-characteristics of usability: *Operability*, *learnability*, *understandability*, *attractiveness*, and *compliance*. Operability is the capability of a user to use the software to accomplish a specific goal. Learnability is the ease with which a user learns to use the software.

Understandability is the ability of a user to understand the capabilities of the software. Attractiveness relates to the appeal of the software to a user. Finally, Compliance measures how well the software adheres to standards and regulations relating to usability. Recommended approaches for measuring each of the sub-characteristics are provided (ISO 2001; ISO 2003). These recommended measurements rely heavily on psychometric and cognitive evaluation techniques.

Many experts view ease of learning or time to learn as an important characteristic of software usability, and it is also essential to the IEEE definition of usability. (McCall et al. 1977; Nielsen 1993; Seffah et al. 2006b; Shneiderman 1998; Winter et al. 2008). Nielsen's adds memorability as a major characteristic of software usability. Shneiderman and Nielsen classify computer users as novice and experts. Shniederman extends this definition to include an occasional user. These categorizations emphasize learning as a primary characteristic of usability. A novice is in the process of learning to use the software. An occasional user is continually relearning to use the software. An expert has already mastered the software and learns very little each time they use the software. This leads to the conclusion that it is desirable to design the evaluation of usability in a way that enables establishing the learnability of the software.

Software engineers generally employ a number of different methods to evaluate software quality characteristics, such as correctness proofs, inspections, and testing (Beizer 1990; Fagan 1976; Gries 1987; Kit 1995; Myers 1979; Pressman 2010). Usability evaluation has a much richer assortment of methodologies available to the evaluator (Andre et al. 2001; Blandford et al. 2008; Dumas and Redish 1999; Fitzpatrick 1999; Hornbaeck 2006; Howarth et al. 2009; Mills et al. 1986; Nielsen 1993; Nielsen 2005; Rubin and Chisnell 2008; Skov and Stage 2005; Tullis and Albert 2008). The paper "Strategies for Evaluating Software Usability" illustrates the large assortment of evaluation methods (Fitzpatrick 1999). The list of generic methods cited by Fitzpatrick, includes: Observation, Questionnaire, Interview, Empirical Methods, User Groups, Cognitive Walkthroughs, Heuristic Methods, Review Methods, and Modeling Methods.

Each of these evaluation methods has strengths and weaknesses. Reviews, Cognitive Walkthroughs, and Heuristic Methods are good approaches for evaluating user interface and providing quality feedback to designers early in the development process. Inspections, however, are limited to investigating interface issues and do not evaluate the way that the hardware and software interact. Modeling methodologies, such as Goals, Operators, Methods, and Selection rules (GOMS), provide excellent techniques for evaluating user interface designs before they are implemented (John and Kieras 1996). Observation, Questionnaire, Interviews, and User Groups offer excellent feedback about the user perception of the usability of the software. The Empirical approach provides valuable data about the components' interaction with users and employs the same concepts as those used in the validation stage of development to evaluate other quality characteristics exhibited by the software.

Observation, Questionnaires, Interviews, User Groups and Empirical or Execution based testing methodologies employ users' experience with the software as the basis of their evaluation. From the prospective of a software developer, all of these methodologies are problematic. Acquiring accurate data from the user-based methods requires having all of the hardware and software for the facility under evaluation. This type of summative evaluation in a traditional or a waterfall development process is limited to the validation or final phase of a development cycle (Hilbert and Redmiles 2000; IEEE 1990; Kit 1995; Royce 1970).

Most software development processes divide quality evaluation methods into Verification and Validation (V&V) activities. Classifying an evaluation methodology as a Verification or Validation task depends on the type of information produced by the activity. If an evaluation methodology produces data about the product under development, it is a verification activity (IEEE 1990). If an evaluation methodology produces data about the way a product addresses the users needs, the methodology is classified as a validation activity and serves as a formative evaluation (Hilbert and Redmiles 2000; IEEE 1990). A rigorous V&V plan employs multiple evaluation methods in order to provide the evaluator with the maximum amount of data for their evaluation, and serves as a normative evaluation of the product (IEEE 2002; IEEE 2004; Kit 1995).

Most of the literature on empirical usability evaluation or usability testing comes from the Human Computer Interaction (HCI) community (Dumas and Redish 1999; Nielsen 1993; Rubin and Chisnell 2008; Tullis and Albert 2008). One of the most comprehensive discussions on usability evaluation occurs in Handbook of Usability Testing (Rubin and Chisnell 2008). The approach described in this text provides a good combination of inspections and empirical evaluation or testing activities. Even though the text provides no differentiation between inspections and tests, the authors provide a robust definition permitting most software engineers to identify the activity. For example, the authors propose an exploratory test conducted early in the cycle to provide feedback between the requirements and design phase. Software testing terminology defines a test in terms of the execution of software (IEEE 1990; Myers 1979). Using a classical development methodology, such as the waterfall model, there is nothing to execute during this phase of the development. Even when prototyping concepts are employed, evaluations at this point in time are characterized as an inspection (Pressman 2010).

Most cognitive evaluation methods however, require expertise in the cognitive science and are not readily available to the developers. Moreover, currently cognitive evaluation does not exploit contemporary and advanced technology to its fullest. For example, time on task, which is an important factor of usability, is often measured using a manual stop-watch operated by an HCI expert during an observation session. Current technology can provide the HCI experts with a "perfect stop-watch" that measures to a sub second accuracy the timing events related to the interface and correlate them with spatial and temporal user and system activities. Moreover, it can supply the evaluator with correlation of time on task and interface events in different areas of the screen, correlation between timing and eye movements of the user which can give indication on widget placement, as well as an indication that at a specific time the user rapidly moved the mouse back and forth from one widget to another. In addition, the concepts introduced in the following section can assist in the development and validation of software usability, and provide an enhanced issue and defect resolution techniques.

3 Effort-Based Usability Metrics

As stated in the previous section, the ISO/IEC 9126-1 lists 5 main sub characteristics of usability: operability, learnability, understandability, attractiveness, and compliance. The model described in this paper concentrates on learnability, operability, and understandability, where we list learnability first since it is a cornerstone of the new approach and is one of the most frequently cited software usability characteristics.

It is possible to construct a model of usability using the sub-characteristics of operability, learnability, and understandability. With the elimination of the subjective characteristics such as attractiveness, compliance, and satisfaction, the model may lack a level of precision but should provide a basis for both summative and formative evaluation. In this section, we present the theoretical model of effort based usability and elaborate on the learnability, operability, and understandability aspects.

3.1 Effort-Based Usability Model

One of the hypotheses that govern this research is that effort, which can be measured by several objective measures, closely (and inversely) correlates with usability. For this model, E denotes all the effort required to complete a task with computer software, as defined by the following vectors:

$$E = \begin{pmatrix} E_{mental} \\ E_{physical} \end{pmatrix}$$

$$E_{mental} = \begin{pmatrix} E_{eye\_mental} \\ E_{other\_mental} \end{pmatrix}$$

$$E_{physical} = \begin{pmatrix} E_{manual\_physical} \\ E_{eye\_physical} \\ E_{other\_physical} \end{pmatrix}$$

Where:

E_{eye_mental} the amount of mental effort to complete the task measured by eye related metrics.

E_{other_mental} the amount of mental effort measured by other metrics.

$E_{physical}$ the amount of physical effort to complete the task.

$E_{manual\_physical}$ the amount of manual effort to complete the task. Manual effort includes, but is not limited to, the movement of fingers, hands, arms, etc.

$E_{eye\_physical}$ the amount of physical effort measured by eye movement related metrics.

$E_{other\_physical}$ the amount of physical effort measured by other metrics.

The accurate estimation of the total effort (E) requires a complete knowledge of the mental and physical state of a human being. This is not possible with the current technology; therefore, approximations techniques are used to estimate total effort (E). Logging keystroke and mouse activity approximates the manual effort ($E_{manual\_physical}$) expended by a subject. An eye-tracking device allows logging eye position data to estimate the amount of mental effort (E_{eye_mental}) and physical effort ($E_{eye\_physical}$) in terms of eye movement metrics. Terms such as E_{other_mental} and $E_{other\_physical}$ are estimation factors that might be contributing to the effort required for task completion that cannot be measured accurately by current technology.

3.1.1 Mental Effort

Salomon defines the notion of Mental effort as the "number of non-automatic elaborations applied to a unit of material" (Solomon 1983). He employs the concept in motivational and cognitive aspects of information processing. A related term is cognitive load, which refers to the ability to process new information under time constraints. The accurate assessment of information processing and cognitive load is hard due to the fact that it involves attention, perception, recognition, memory, learning, etc. Nevertheless, it can be partially estimated by eye movement metrics measured by an eye tracking device (Andreassi 1995), (Ikehara and Crosby 2005). Modern eye tracking devices are similar to web cameras, without any parts affixed to the subject's body (Duchowski 2007). Eye trackers provide useful data even in the absence of overt behavior. With this device, it is possible to record eye position and identify several eye movement types. The main types of the eye movements are (Duchowski 2007):

Fixation – eye movement that keeps an eye gaze stable with respect to a stationary target providing visual pictures with high acuity,

Saccade – very rapid eye movement from one fixation point to another,

Pursuit – stabilizes the retina with respect to a moving object of interest.

In the absence of dynamically moving targets, the Human Visual System usually does not exhibit pursuits. Therefore, parameters related to smooth pursuit are not discussed in this paper. In addition to basic eye movement types, eye tracking systems can provide biometric data such as pupil diameter. Since pupil diameter might be sensitive to light conditions and changes of brightness level of the screen, it is not included as a mental effort metric in this research.

Many researchers consider the following metrics as a measure of the cognitive load and mental effort:

Average fixation duration: measured in milliseconds (Crosby et al. 2009; Jacob and Karn 2003).

Average pupil diameter: Eye tracking systems enable measuring biometric data such as pupil diameter. (Fitts et al. 1950), (Kahneman 1973; Marshall 2002).

Number of fixations: Due to non-optimal representation, overall fixations relate to less efficient searching (Goldberg and Kotval 1999). Increased effort is associated with high amounts of fixations.

Average saccade amplitude: Large saccade amplitude, measured in degrees, can be indicative of meaningful cues (Goldberg et al. 2002; Goldberg and Kotval 1999). To a certain extent, large average saccade amplitude represents low mental effort due to the notion that saccades of large amplitudes indicate easier instruction of meaningful cues (Fuhrmann et al. 2009).

Number of saccades: High number of saccades indicates extensive searching, therefore less efficient time allocation to task completion(Goldberg and Kotval 1999). Increased effort is associated with high saccade levels.

Generally, the user expands some type of effort at any given time. This effort might be related to eye muscle movements, brain activity, and manual activity. Hence, $E(t)$, the effort expanded by the user at time t, is a continuous time function (i.e., analog function). For the mental effort, it is assumed that:

$$E_{eye\_mental}(t) = \int_{t_0}^{t} \left(f_1(t) + f_2(t) + f_3(t)\right)dt$$

Where: $f_1(t)$ is a function of the fixation duration and number of fixations; $f_2(t)$ is a function of the pupil diameter, and $f_3(t)$ is a function of the saccade-amplitude and number of saccades.

Note that $E(t)$ is a monotonically increasing function. The definition of effort uses continuous time functions. In practice, given the discrete nature of computer interaction, these measures are quantized by converting integrals to sums. Occasionally, eye-tracking devices produce data that is below a reliability threshold. Periods where the data is not reliable are excluded from integration.

Direct evaluation of the actual functions $E(t), f_1(t), f_2(t)$, and $f_3(t)$ is complicated and is beyond the scope of this research. Nevertheless, a research to estimate the expanded effort through a model of the eye muscles is currently ongoing. In the current research, we represent $E_{eye\_mental}(t)$ as a vector consisting of two elements:

$$E_{eye\_mental}(t) = \begin{Bmatrix} \text{Average fixation duration} \\ \text{Number of fixations} \end{Bmatrix}$$

Where the average fixation duration and the number of fixations are calculated over a period $[t_0, t]$. since saccades are highly correlated to physical effort, they are listed under $E_{eye\_physical}$

3.1.2 Physical Effort

The main components of the manual physical effort expanded by the user relate to mouse and keyboard activities. Hence, in the case of interactive computer tasks, it may be possible to calculate effort by counting mouse clicks, keyboard clicks, Mickeys, etc. The term Mickey denotes the number of pixels (at the mouse resolution) traversed by the user while moving the mouse from a point (x_0, y_0) to a point (x_1, y_1).

In similarity to the definition of $E_{eye\_mental}(t)$, $E_{manual\_physical}(t)$ is defined to be:

$$E_{manual\_physical}(t) = \int_{t_0}^{t} \left(f_4(t) + f_5(t) + f_6(t) + f_7(t)\right) dt$$

Where: $f_4(t), f_5(t)$ and $f_6(t)$ are (respectively) functions of the number of mickeys, mouse clicks, and keystrokes by a subject during the time interval $[t_0, t]$; and $f_7(t)$ is a function that serves as a penalty factor that measures the number of times the user switched from mouse to keyboard or vice versa during the interval.

As in the case of mental effort, direct evolution of the functions $f_4(t), f_5(t), f_6(t)$, and $f_7(t)$ is beyond the scope of this research. In the current research we represent $E_{manual\_physical}(t)$ as a vector consisting of the elements:

$$E_{manual\_physical}(t) = \begin{Bmatrix} \text{mickeys count} \\ \text{mouse clicks count} \\ \text{kestrokes count} \\ \text{transitions count} \end{Bmatrix}$$

Where the counts are calculated over a period $[t_0, t]$.

Ideally, the effort expended by the Human Visual System (HVS) to complete a task is represented by the amount of energy spent by the HVS during the task. The energy expended is dependent on the amount of eye movements exhibited by the HVS, the total eye path traversed and the amount of force exerted by each individual extraocular muscle force during each eye rotation. In similarity to the discussion on E_{eye_mental}, the following parameters are considered.

Average saccade amplitude (see definition above).

Number of saccades (see definition above).

Total eye path traversed: This metric, measured in degrees, presents the total distance traversed by the eyes between consecutive fixation points during a task. This metric takes into account the number of fixations, number of saccades and exhibited saccades' amplitudes. The length of the path traversed by the eye is proportional to the effort expended by the HVS.

Extraocular muscle force: The amount of energy, measured in grams per degrees per second, required for the operation of extraocular muscles relates to the amount of force that each muscle applies to the eye globe during fixations and saccades. Based on the Oculomotor Plant Mechanical Model, it is possible to extract individual extraocular muscle force values from recorded eye position points (Komogortsev and Khan 2009).

The total eye physical effort is approximated by:

$$E_{eye\_physical}(t) = \int_{t_0}^{t} \left(f_8(t) + f_9(t) + f_{10}(t)\right) dt$$

Where: $f_8(t)$ is a function of the saccades amplitude and number of saccades; $f_9(t)$ is a function of eye path traversed, and $f_{10}(t)$ is a function of the total amount of force exerted by the extraocular muscles. The integration includes only periods with reliable data.

In this research we represent $E_{eye\_physical}(t)$ as a vector consisting of the elements:

$$E_{eye\_physical}(t) = \begin{Bmatrix} \text{average saccade amplitude} \\ \text{saccade count} \\ \text{eye path traveresed} \end{Bmatrix}$$

Where the counts are calculated over the period $[t_0, t]$.

3.2 Learnability-Based Usability Model

The methodology proposed in this paper is centered on concepts that relate to learning and learnability. The idea is to evaluate several aspects of usability as the user completes a set of tasks originating from a single scenario. Typically, as subjects master an application, the time to complete tasks with the same scenario becomes shorter (Ebbinghaus 1885; Hax and Majluf 1982; Wright 1936). To illustrate, consider the following example. Assume that a set of n subjects selected at random complete a set of k tasks. Further, assume that the subjects are computer literate but unfamiliar with the application under evaluation. The objective of each task is to make travel reservations, and each task requires about the same effort. The set of k tasks have the same scenario with different data and different constraints. When plotting the

Time-On-Task (TOT) averages (T_{avg}) for these subjects, a curve with a strong fit to either a power law or exponential decay curve is said to reflect learning or represents a learning curve (Hax and Majluf 1982; Ritter and Schooler 2001; Tullis and Albert 2008; Wright 1936).

Selection of a model depends on how subjects learn. If a human's performance improves based on a fixed percentage, then the exponential decay curve is appropriate (Ritter and Schooler 2001). Using an exponential function assumes a uniform learning rate where learning everything about the software is possible. On the other hand, if a human's performance improves on an ever decreasing rate, then the power law is the appropriate choice. In this research, the power law is used because mastering computer software is measured as an ever decreasing percent of what is possible to learn. Furthermore, experience gained from our experiments supports the use of the power law. In this context, learning is modeled by time on task and/or by effort per task according to the following equation:

$$Time\ (effort) = E_{base} + \alpha + B \times N^{-\beta} \tag{1}$$

Where: E_{base}, the baseline effort, is a constant that represents the minimal time or effort required for completing the task, and $\alpha \geq 0$ is a constant relating understandability. Due to understandability issues, some users may never reach the minimum time or effort required to complete a task and their performance converges to a value higher than E_{base}. *B*, referred to as the learning range, represents the approximate time or effort for completion of the first task. *N* is the number of trials. β is a positive number representing the rate of subject learning. Note that after a large number of tasks, the time or effort approaches the constant $E_{base} + \alpha$.

To further elaborate, E_{base}, the minimal time or effort required for completing the task can be associated with the expert or designer effort. That is, the level of performance of one or more experts, such as the application's designer or a person accepted as an expert on a specific Commercial-Of-The-Shelf (COTS) application. E_{exp}, is referred to as the expert effort. In many cases, an expert may not achieve a baseline (E_{base}) level of performance.

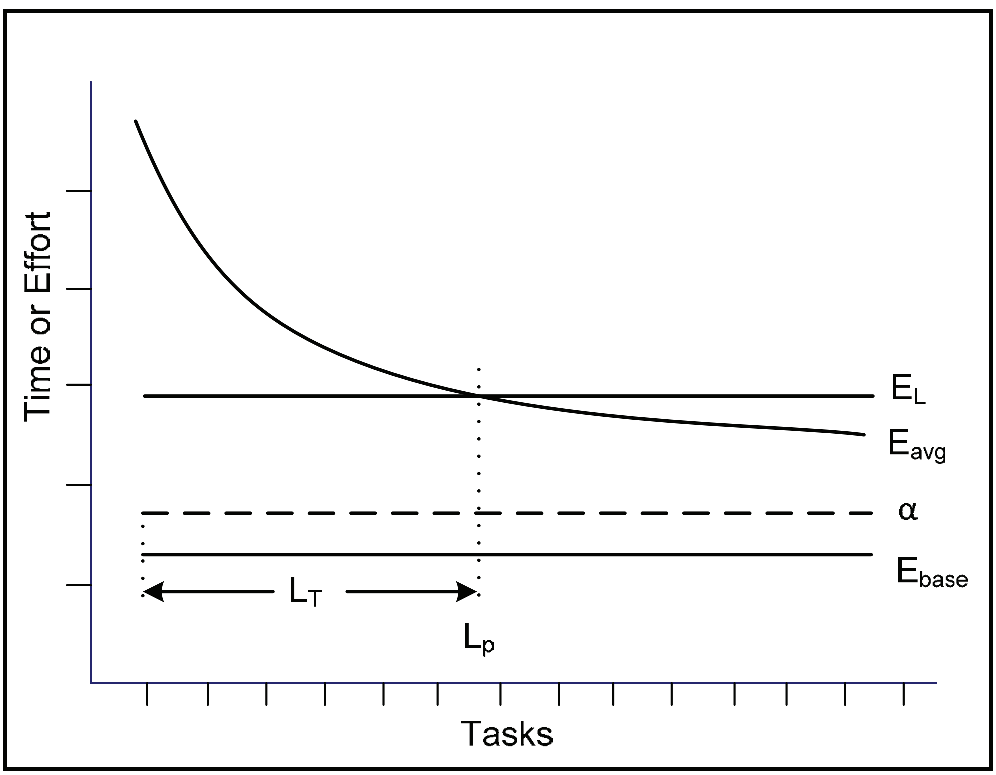

Figure 1 Performance-Based Model

Figure 1 illustrates a usability model based on average effort to complete a set of tasks with a common scenario. Assuming that learning to an acceptable level of performance occurs during the execution of the first few tasks, the task where the subject's effort reaches this acceptable level of performance is the learning point (L_P). Summing the average task duration to the left of the learning point (L_P) indicates how much time (L_T) the average subject requires to reach an acceptable level of performance. Data to the right of the learning point (L_P) describes the amount of effort required to accomplish a task by a trained user. Learnability, Operability and Understandability are the sub-characteristics that put the subject's effort into a context and are described in the next sections.

3.3 Learnability

Learnability, the ease with which a user learns to use the software, is possibly the most critical characteristic of software usability; and it may account for many of the user complaints about software usability. All computer based tasks require some learning. Current human interface design practice doesn't always address this concept directly but usually addresses it indirectly by indentifying levels of user expertise (Nielsen 1993; Shneiderman 1998).

It is possible to measure learnability by plotting either the average Time-On-Task (TOT) or the average Effort-On-Task (EOT), that is, the average effort (E_{avg}) expended by a group of subjects for a task, and then fitting the subjects' average performance for each task in the set to a Power Law curve (Ebbinghaus 1885; Hax and Majluf 1982; Ritter and Schooler 2001; Tullis and Albert 2008; Wright 1936). When there is a tight fit to the Power Law curve, then it is possible to say that learning is occurring. By using both the goodness of fit (R^2) and the learning rate (β), it is possible to establish measurable requirements for the software and to compare the learnability of different applications. If the effort or time a subject expends on a series of tasks has a strong fit ($R^2 \geq .7$) to a Power Law curve, then it is possible to assert that learning is observed in the evaluation. Since humans always learn, a plot of time or effort that does not produce a good fit indicates that there is another problem masking the learning. Using the learning rate, a test engineer can estimate the number of tasks necessary to evaluate the software.

Another learnability feature that can be inferred from the learning model is the learning point (L_p), which indicates that the average subject has reached an acceptable level of performance (E_L). This first requires establishing a satisfactory level of learning (r). It is possible to set the level of learning (r) as a percent of the constant E_{base}, from Equation 1. Where E_{base} represents the minimal time (effort) required for completing the task. It is then possible to calculate the acceptable effort (E_L) in the following way:

$$E_L = \frac{X}{100} \times E_{base}$$

Where: X is a number between 0 and 100 representing a percentage of mastery. When X=100, the subjects have completely mastered the set of tasks. A more realistic value for X might be 80 denoting that the subjects have reached a level of 80% of the optimal mastery. The learning point $(L_p)$is defined to be the first task where $E_{avg} \leq E_L$. Another method for establishing the learning point (E_L) is based on requirements provide by the stakeholders.

3.4 Effort-based Operability

Operability is the capability of a user to use the software to accomplish a specific goal. A unique feature of this research is that operability can be derived from the learning model. Consider Equation 1. Several correlated parameters derived from this equation can be used to denote operability. The term E_{base} can be used to denote the ultimate operability of a scenario. It is also possible to define operability in terms of either expert effort $(E_{\exp})$or acceptable effort (E_L), that is, the acceptable performance of a non expert after learning the system. Recall that $E_{base} + \alpha$ is the asymptote of Equation 1 with respect to non-expert users. Hence, one can define the operability of a system as:

$$Op = E_{base} + \alpha$$

3.5 Understandability

One method of evaluating understandability is to compare the average subjects' performance on a specific set of tasks to the baseline performance, such as the designer or a person accepted as an expert. A more precise definition of Understandability with a learning model is the ratio of the operability to the baseline performance, as described in the following equation:

$$U = \frac{E_{base} + \alpha}{E_{base}}.$$

In practice, Understandability can be approximated in the following way. Let L_p denote the learning point as defined in section 3.2, then understandability is approximated by:

$$U = \frac{\frac{1}{n}\sum_{i=L_p+1}^{n}\left(E_{avg}\right)_i}{E_{exp}}$$

Where: n is the number of tasks performed. Expert effort (E_{exp}) is used in the equation as the best approximation of baseline effort (E_{base}).

3.6 Relative Measures of Usability

Sections 3.3 to 3.5 defined absolute measures of usability in terms of learnability, operability, and understandability. Often, it is desired to compare the usability of two or more systems. Consider two systems, system $S_1$and system S_2, where S_1 and S_2 are characterized by the learning curves

$EOT_1 = E_{base1} + \alpha_1 + B_1 \times N^{-\beta_1}$ and $EOT_2 = E_{base2} + \alpha_2 + B_2 \times N^{-\beta_2}$ respectively, then each of the components of the sets: $\{E_{basei}, \alpha_i, B_i, \beta_1, L_{pi}, E_{Li}, O_i, U_i\}$ for $i = 1{,}2$ can shed light on the relative usability of system S_1 compared to system S_2.

4 Utilizing the Effort Based Measures Approach

This section presents a methodology for applying the theoretical concepts developed in section 3, in the process of user interface design and evaluation. In specific, it demonstrates the use of the effort based metrics along with the concept of the baseline effort (E_{base}) to provide interface designers and developers with a methodology to evaluate their designs as they are completed, and a usability testing technique that is applicable to both the verification and validation. This framework provides the designer with system level feedback for a specific design element. It is similar to the feedback on the quality of source provided by a unit test.

There are a few ways to estimate the baseline effort. An expert in using the software can be used as an approximation. In a new development, the best "experts" on using an application are the designers of the use case and its human interface. They can evaluate the minimal effort required for completion of each task embedded in the designed software by using a tool or rubric to estimate an ideal subject's effort or measure the effort. Their estimate of E_{base} is referred to as the designer's effort. Finally, it is possible to establish the expert effort (E_{exp}) analytically.

4.1 Designer's Effort

Often the person who is most knowledgeable about the usability of the software is the interface designer who is the actual expert. In this case, the terms expert effort and designer's effort are interchangeable.

Designer's Effort is a notion providing developers with a tool that can reduce the cost of design reviews and prototypes. It also provides the designer with feedback on the quality of the design from a usability test. One of the main benefits of the designer effort evaluation is that it provides designers with a timely low-cost method of evaluating their design and making trade-off decisions in a manner similar to those used to develop other software components.

To further illustrate, assume that a group of 10 subjects records an average Time-On-Task (TOT) of 420 seconds on a specific task. Asking whether this is a good or bad TOT is meaningless. Nevertheless, if the notion of expert effort (E_{exp}) is added to the question, then there is a basis for comparison. For example, assume that a group of 10 subjects recorded an average TOT of *x* seconds, and an expert user recorded a time of y seconds on the same task. This provides information for sound evaluation of the usability of the application. Having an expectation of the

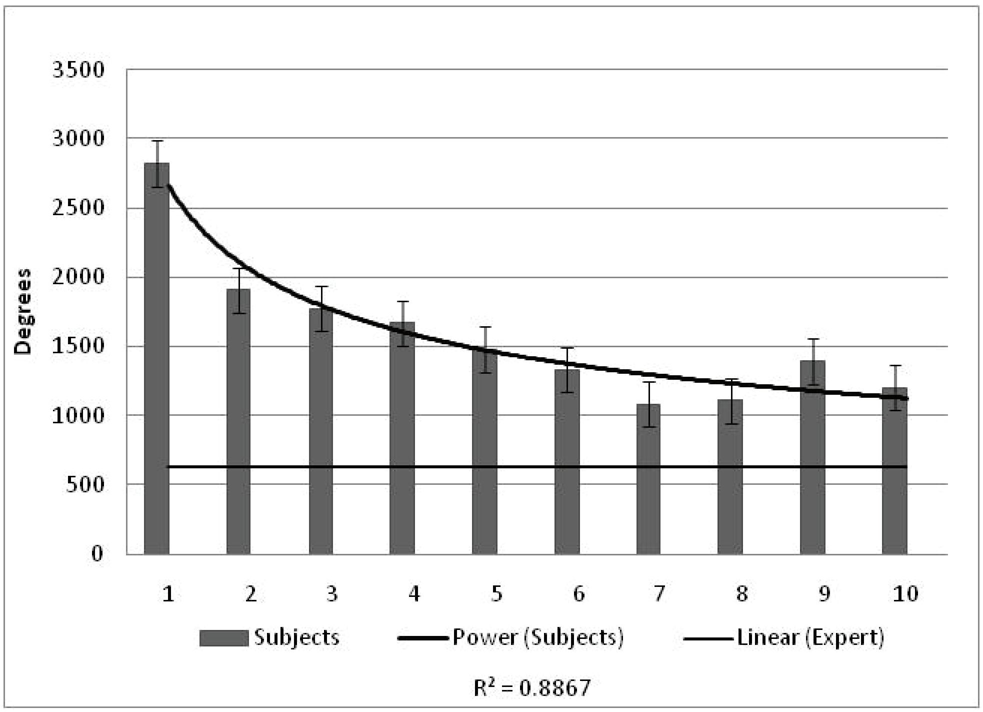

Figure 2. Designer's Effort (eye) for System B

target software function and performance is also one of the fundamental principles of software testing (Myers 1979).

As shown in Figure 2, the expert effort (E_{exp}) provides a reference point placing the subject data into a context, making a meaningful evaluation possible. However, comparing the performance of an expert to a group of individuals just becoming familiar with the software is not a valuable comparison. It is only possible to compare the expert's performance to that of the subjects after the subjects have passed the learning point (L_P).

Calculating the effort on a new interface design is not difficult. First, the designer should count the number of keystrokes and the number of mouse button presses. Then measure the distance necessary to move the mouse and count the number of keyboard-to-mouse interchanges. Another less tedious approach is to develop a test harness that displays the interface and a data logging utility. The test harness does not need to be functional beyond the point of operating the interface. A data logging utility, which is the same tool used in the evaluation to collect subject data, can be used by the designer. The ability to calculate effort provides the developer with this feedback mechanism. Extending the notion of effort-based interface evaluation to unit testing provides designers with feedback focused on a specific scenario. Using an eye tracking device provides additional insight into effort required by the interface.

It is also possible to use the notion of Designer's Effort in the evaluation of Commercial-Off-The-Shelf (COTS) software. An evaluation team could use either results from an expert user or an analysis of the user interface to establish the Designer's Effort. A software publisher could provide usability data on the application, but COTS evaluators may find that developing their own data for the application provides an independent review of the software.

Providing an interface designer with a technique for evaluating the ideal efficiency of the interface provides the developer with a method of evaluating designs without calling a meeting or constructing a prototype. Just evaluating manual effort $(E_{manual\_physical})$ would provide a designer with a basis for making tradeoff decisions. For example, one thing that can greatly increase effort is making a switch from the keyboard to the mouse and back. Many designers include "combo box" widgets in a design. There are several different implementations of a "combo box". Generally, they provide a drop down menu to aid the user's selection. Some implementations require the user to make their selection with a mouse button press; other implementations permit the user to type the first character until reaching their selection. Generally, a widget employing a mouse drag and click requires more effort than one that doesn't. Using an effort-based interface evaluation, the designer can see the total effect of their design and, when possible, can select tools or objects to make the design more physically efficient.

In addition to cost effective evaluation of user interfaces, designer's effort provides an approach to establish subject understanding of the application. For example, if after learning the application, the subjects have an understanding equal to the designer's effort, then it is possible to say the subject's knowledge of the application is equal to the designer's knowledge. Normally, subjects expend more effort in completing a set of tasks than an expert, and it is possible to use the difference to express the usability / understandability of an application.

4.2 Designing the Test

There are a number of widely accepted references for designing a usability test (Dumas and Redish 1999; Nielsen 1993; Rubin and Chisnell 2008; Tullis and Albert 2008). These references provide detailed guidelines on almost every aspect of usability testing from the laboratory design to reporting results. Nevertheless, the framework used in this paper and the focus on learning as the vehicle for deriving other usability measures necessitates additional attention in designing tests. Some of the additional considerations relate to the focus of tests on specific parts of the system, the creation of tasks that exploit the new framework, and the number of tasks that have to be conducted in order to evaluate a specific component of the system. Like any other type of test, the process for a usability test consists of preparation, execution, and analysis phases. The following section identifies the test design framework highlighting the elements that are unique to the effort based usability approach.

One of the first steps in constructing a usability test is to establish the usability requirements for the software under evaluation. At a minimum, clients should provide a profile for each user of the application and requirements for the "In Use" Quality characteristics and learnability (ISO 2001). The user profile should include characteristics such as education, experience with user interfaces, skills with a rating of expertise, etc. Describing the systems functionality using Unified Modeling Language (UML) use cases provides a focus for both specifying requirements and evaluating the software (Rumbaugh et al. 1999). It is logical to assume that different tasks require different amount of effort than other tasks; therefore, each use case should have its own set of requirements.

After establishing requirements for each use case, the next step is to design a set of goals or tasks to evaluate a specific use case. The current method for constructing a usability tests concentrates on real world situations and uses them as the basis for designing tasks (Dumas and Redish 1999; Nielsen 1993; Rubin and Chisnell 2008; Tullis and Albert 2008). In light of the experience gained from developing this framework, two more components are required from a test suite:

1. It has to contain tasks that allow the subject to master the use of the system before making measurements of usability.
2. It has to enable a software engineer to diagnose issues.

For this end, an approach that uses a set of test cases or tasks from a scenario based test design technique utilizing a use case diagram (Kaner 2003; Rumbaugh et al. 1999) is adopted. It provides the developer with a known focus, so that issues identified in the test trace to a specific set of interface components. Designing tasks based on use cases also insures complete coverage of the application.

Many human beings learn by repeating a task or action a number of times. If tasks are identical, however, then the subjects can memorize the solution without real learning of the way to solve that class of problems. To address this issue, the developer has to create a series of tasks that are different but based on the same scenario; such a set is referred to as a set of *identical independent tasks*. Figure 3 includes the top level use case diagram for a travel reservation system. For this application, it is possible to randomly create a series of tasks of travel to different destinations under different sets of constraints, such as budget and time constraints, rental car requirements, and accommodation requirements. For example, a few tasks might require a hotel with a pool while others require internet connection. Building a series of tasks from a single scenario and providing complete coverage make it possible to construct multiple identical and independent tasks. Next, a relatively small group of subjects completing a set of identical and independent tasks would allow the developer to measure and thereby observe the learning and performance of the subjects.

With the simple example of the Travel Reservation System illustrated in figure 3, it is feasible to provide 100% coverage of all of the use cases, but in a more complex application this discipline needs to assure coverage of all of the different scenarios. Furthermore, random selection of tasks would present a problem since it is important to select tasks that require about the same completion time (or effort). This would enable observing learning and the improvement of user performance as they master tasks.

Developing the set of tasks consists of a few steps:

1. Select a use case for evaluation.
2. Convert the input for the use case into a narrative.
3. Identify important events, conditions, or constraints and add their description to the narrative.
4. Test the scenario on all the systems that are under evaluation.
5. Replace the specifics (e.g., a constraint related to budget) of the scenario with blanks or with an option list creating a template.
6. Convert the template into a set of tasks by filling in the blanks with specific and valid data selecting a single occurrence from each option list.
7. Test all of the tasks on all of the systems under evaluation.

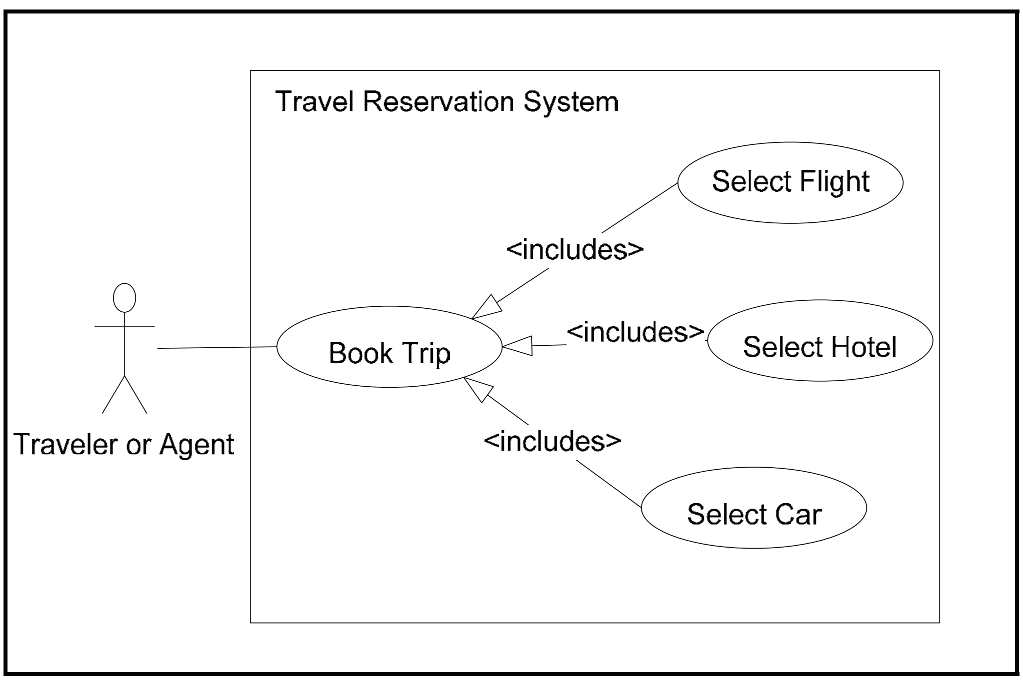

Figure 3. Use Case Diagram

Another set of question relates to the appropriate number of tasks, task length, and number of subjects. For this type of usability test, the literature suggests a number of subjects from six to twenty (Nielsen 1993; Nielsen 2008). Using the experience gained in a large set of field tests, the approach adopted for this research is to use 6 – 10 subjects (according to availability) conduct about 10 identical independent tasks and limit the duration of each task to 6-15 minutes to reduce subject fatigue. Conducting 10 tasks enables accurate identification of the learning point (Komogortsev et al. 2009).

A small pool of 6-10 subjects permits using this approach as part of the construction phase, after the developers have completed their normal testing or as part of an iterative development process. When using this technique as part of the construction phase with scenarios without any unusual conditions, the test provides the designer with feedback about the quality of the use case early in the development process. Conducting a complete usability test is better when the software is at its most stable configuration.

The main requirement for a test facility is a minimal number of distractions. A small office or conference room is adequate. In addition, the designers would use a harness for logging user activity. There is no need for a stopwatch to record the time since the logging harness contains this information in addition to other valuable information, such as manual user activity and potential eye tracking data. While this technique is not intended to replace elaborate facilities to conduct usability tests, it can be used to complement current usability evaluations and reduce the number of elaborate testing, thereby reducing the total usability evaluation cost (Dumas and Redish 1999; Rubin and Chisnell 2008). Another novelty of this framework is the addition of a software module that measures the percent completion of each task by the user. This can be compared to the user perception of task completion.

The largest external expense to implement the tools and techniques discussed in this paper is the cost of acquiring subjects. Compensation for university students is less expensive and might have a number of non-monetary alternatives. A research that is currently ongoing identifies several cases where a student population can serve as a good sample for the actual system users. In other cases, however, temporary agencies can probably supply an adequate number of subjects conforming to the user profile.

The next section elaborates on a set of experiments performed to assess the utility of the new framework.

5 Experimental Results

An experiment using two travel reservation systems (referred to as system *A* and system *B*) was conducted to ascertain the assertions of effort-based usability. For this experiment, each subject completed 10 travel reservation tasks. Ten subjects provided data for System A and 10 for System B. In addition, this experiment also provides a great deal of insight into designing and conducting a usability test.

5.1 Travel Reservation System Comparison

The data acquired for logging actual interaction and eye tracking produced a number of very important results. Trend analysis of physical effort expended by the users corresponds to the expected learning curve. In addition, the data, verified via ANOVA analysis, supports the framework's model. The following sections contain a detailed account of the results.

5.2 Data Reduction and Analysis

An event driven logging program is devised to obtain details of mouse and keystroke activities from the operating system event queue. The program saves each event along with a time stamp into a file. The logged events are: Mickeys, keystrokes, mouse button clicks, mouse wheel rolling, and mouse wheel clicks. In the reported experiments, the program has generated about 60,000 time stamped events per task (about 10 minutes). The eye tracking system produces an extensive log of time stamped events including parameters such as fixation duration and saccade amplitude. In addition, accurate measurement of task completion time is enabled through the eye tracking device.

A data reduction program applied to the events log, counts the total number of events (e.g., Mickeys) per task. A similar program is used for eye activity events. Both programs execute the entire data set (log of manual activity and eye activity) which consists of several millions of points in less than an hour. With 20 subjects, each completing 10 tasks, the data reduction program generates 200 data points. The data obtained from the data reduction stage is averaged per task per travel reservation system. Hence, a set of 20 points is generated where each point denotes the average count of events per task per reservation system.

5.2.1 Results and Evaluation

Figure 4 illustrates the average task-completion-time per task per system. Compared to System *B*, System *A* has a jittered trend, yet it follows a similar slope. In addition, the task completion time for System *A* is more than twice than the completion times for System *B*. The standard deviation values computed for System *A* are higher than the standard deviation values of System *B*. System *A* and System *B* implement the same application, yet from the data presented in Figure 4, it appears that System *B* subjects learn faster than System *A* subjects. Furthermore, the figure demonstrates that System *A* subjects are less productive than System *B* subjects. Hence, it is safe to conclude that System *B* is more operable and learnable than System *A*.

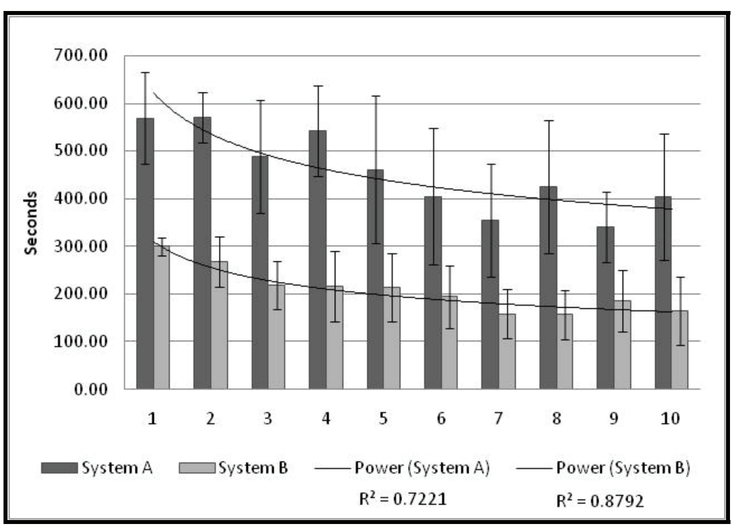

Figure 4. Average Task Completion Time

Figures 5 and 6 provide additional evidence of the usability model's soundness. Figure 5 depicts the average Mickeys per task for System B, and indicates a high correlation with the time and effort usability model. Figure 6 depicts approximate eye physical effort by using the product of average saccade amplitude and the number of detected saccades. A strong fit to a power law curve was observed (R2=0.88) with learning point reached after the 5th task. Like Figures 4 and 5, Figure 6 indicates an agreement with the effort-based usability model. Moreover, a spike in activity with respect to task 9 can be used as an example of the capability of the metrics to discover potential interface shortfalls. Using the usability model to discover or pinpoint usability issues is currently under investigation.

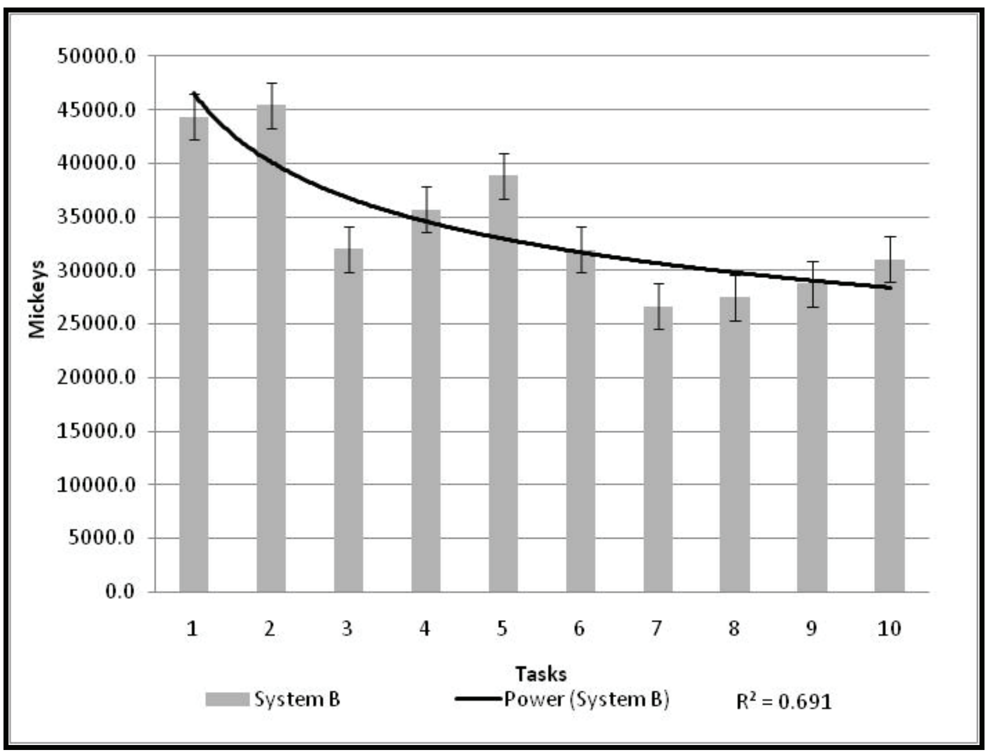

Figure 5 Average Mickeys for System B

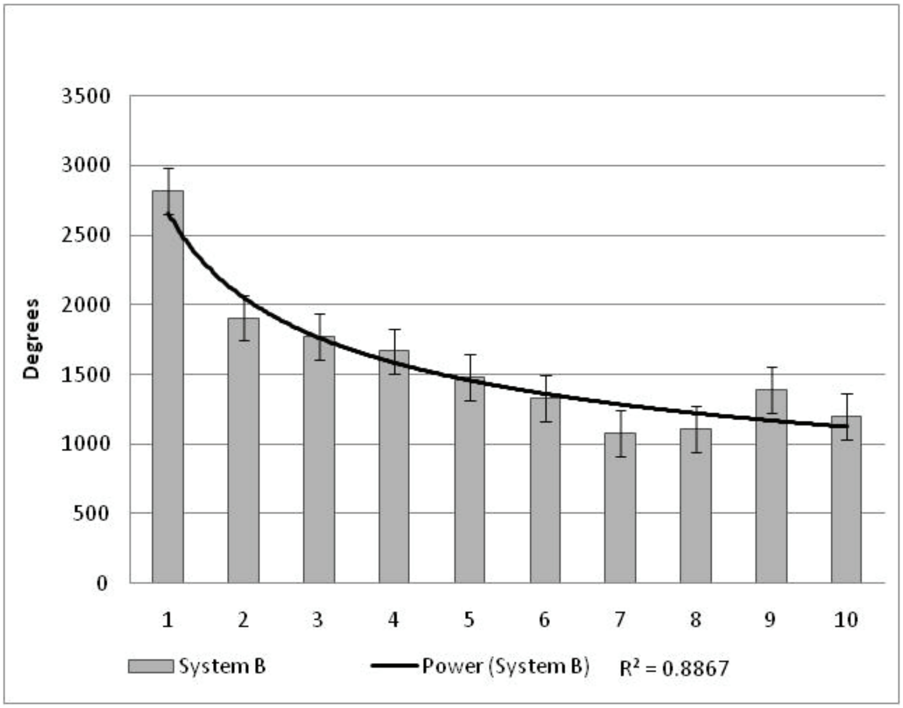

Figure 6. Approximate Eye physical effort for System B.

6 Conclusions and Future Research

This paper has presented an innovative framework for measuring and evaluating software usability. The framework is comprised of two major elements:

- An effort based usability model that is centered on learning.
- A coherent approach for designing and conducting software usability tests.

A learning centered model provides a vehicle to evaluate software usability and adds the capability of a direct comparison of two or more equivalent systems or different versions of the same system. Another advantage of using a learning centered model of usability is that it provides evaluators with a method of predicting subjects' performance. With the usability validation framework presented in this paper, test engineers have information to design the tasks and procedures necessary to conduct a high quality formative evaluation. The experiments presented in this paper provide objective evidence of the effectiveness of the framework.

In conducting this research, it became apparent that one of the major challenges confronting software developers is discovering the specific cause or causes of usability issues. Discovering a cause for a usability issue requires providing a developer with a set of techniques to pinpoint the specific quality characteristic and software element responsible for the user performance, such as interface component placement, instructions, and help facilities. The framework presented in this paper makes a major step to indentifying the software elements involved in the use case scenario that is the basis to the tasks. A future research topic is developing a set of techniques, utilizing the usability model, to pinpoint issues within a task, and the devices providing additional insight into the discovery of the cause of anomalies.

Another major gap in the tools for software designers that might improve the usability of the software is a better feedback mechanism. Goals, Operators, Methods, and Selection rules (GOMS) provides designers with a technique to evaluate the usability of their designs. Integrating GOMS into the effort-based usability model and the validation framework is yet another topic of future research (John and Kieras 1996).

Another direction of future research is to consider a dynamic scenario where the system adapts to the user and enables user specific improvements in usability at run time. This would permit designers to use a “flexible” interface.

APPENDIX A GOALS OR TASKS

A.1 TEMPLATE

A.1.1 GOAL

Dr./Ms./Mr. ____________ is presenting a paper at the ______________ conference being held in _______________ at the _________________. He/she is presenting his/her paper at 10A.M., but he/she must be there for the opening session at 8:30 A.M. The conference will end at 6P.M. on ____________ and Dr./Ms./Mr. _____________ must be there for the closing session.

Dr./Ms./Mr. ________________ is traveling from ______________, and would like a non-stop flight to ___________________.

The conference is at the __________________ hotel on ___________ to ____________, but Dr./Ms./Mr. __________ feels that this hotel is outside of the range of his/her budget of ________ for the travel. Because of the high cost of the hotel he/she wants to stay at a hotel within ____________ miles of the conference center with the following amenities:

1. ___________________________
2. ___________________________
3. ___________________________
4. ___________________________

He/she will need a car to get around at conference city. Again, because of budget constraints, he/she does not want to spend more than ________/day for the car.

A.1.2 DIRECTIONS

Using the web browser already opened, make a flight, hotel, and car rental reservation for Dr. Waterford based on the below information. You should make every attempt to comply with the budget, distance, amenities, and travel time constraints given. Both the departure and return flights *must* be non-stop. Ensure that the airline and hotel reservation is for one adult only. Do not open additional browser windows/tabs, and do not navigate away from System A/System B. You may, however, click on any links provided by System A/System B if they are necessary for, or related to your search.

A.2 GOALS

A.2.1 GOAL 1

Dr. Vornoff is presenting a paper at the *Pikes Peak* conference being held at the Broadmoor hotel in Colorado Springs, Colorado. He is presenting his paper at 10:00 am on Thursday, October 16, but he must be present for the opening session at 8:00 am on Wednesday, October 15 and remain for the duration of the conference, which ends at 3:00 pm on Friday, October 17. He has a travel budget of $800.

Dr. Vornoff is traveling from Salt Lake City, Utah and insists on a non-stop flight to Colorado Springs. Since he feels that the Broadmoor is out of his price range, Dr. Vornoff would like a room at a less-expensive hotel within 10 miles from the conference. This hotel should have the following amenities:

1. Exercise room
2. Internet (wireless or wired)
3. Restaurant/dining room

Dr. Vornoff will need to rent a car during his stay in Colorado Springs. He does not want to spend more than $50 per day, or $180 total for the car rental.

A.2.2 GOAL 2

Dr. Jones is presenting a paper at the *Yellow Brick Road* conference being held at the Hyatt Regency hotel in Wichita, Kansas. She is presenting her paper at 10:00 am on Thursday, October 30, but she must be present for the opening session at 9:00 am on Tuesday, October 28 and remain for the duration of the conference, which ends at 3:00 pm on Friday, October 31. She has a travel budget of $900.

Dr. Jones is traveling from Houston, Texas and insists on a non-stop flight to Wichita. Since she feels that the Hyatt Regency is out of her price range, Dr. Jones would like a room at a less-expensive hotel within 8 miles from the conference. This hotel should have the following amenities:

1. Restaurant/dining room
2. Internet (either wired or wireless)
3. Exercise room

Dr. Jones will need to rent a car during her stay in Wichita. She does not want to spend more than $50 per day, or $250 total for the car rental.

A.2.3 GOAL 3

Mr. Smith is presenting a paper at the *Big Metal Arch* conference being held at the Omni Majestic hotel in St. Louis, Missouri. He is presenting his paper at 10:00 am on Tuesday, October 21, but he must be present for the opening session at 8:00 am on Monday, October 20 and remain for the duration of the conference, which ends at 4:00 pm on Friday, October 24. He has a travel budget of $1400.

Mr. Smith is traveling from San Antonio, Texas and insists on a non-stop flight to St. Louis. Since he feels that the Omni Majestic is out of his price range, Mr. Smith would like a room at a less-expensive hotel within 10 miles from the conference. This hotel should have the following amenities:

1. Restaurant/dining room
2. TV with premium cable channels
3. Exercise room

Mr. Smith will need to rent a car during his stay in St. Louis. He does not want to spend more than $70 per day, or $350 total for the car rental.

A.2.4 GOAL 4

Dr. Waterford is presenting a paper at the *Paul Bunyan* conference being held at the Minneapolis Grand hotel in Minneapolis, Minnesota. He is presenting his paper at 11:00 am on Wednesday, October 15, but he must be present for the opening session at 9:00 am on Tuesday, October 14 and remain for the duration of the conference, which ends at 4:00 pm on Friday, October 17. He has a travel budget of $1000.

Dr. Waterford is traveling from Albuquerque, New Mexico and insists on a non-stop flight to Minneapolis. Since he feels that the Minneapolis Grand is out of his price range, Dr. Waterford would like a room at a less-expensive hotel within 10 miles from the conference. This hotel should have the following amenities:

1. Wireless Internet
2. Restaurant/dining room

Dr. Waterford will need to rent a car during his stay in Minneapolis. He does not want to spend more than $70 per day, or $250 total for the car rental.

A.2.5 GOAL 5

Ms. O'Hara is presenting a paper at the *Tara and Twelve Oaks* conference being held at the Marriott Marquis hotel in Atlanta, Georgia. She is presenting her paper at 3:00 pm on Thursday, September 25, but she must be present for the opening session at 9:00 am on Wednesday, September 24 and remain for the duration of the conference, which ends at 4:00 pm on Friday, September 26. She has a travel budget of $1000.

Ms. O'Hara is traveling from Shreveport, Louisiana and insists on a non-stop flight to Atlanta. Since she feels that the Marriott Marquis is out of her price range, Ms. O'Hara would like a room at a less-expensive hotel within 6 miles from the conference. This hotel should have the following amenities:

1. Exercise room
2. Room service
3. Internet (wired or wireless)

Ms. O'Hara will need to rent a car during her stay in Atlanta. She does not want to spend more than $75 per day, or $300 total for the car rental.

A.2.6 GOAL 6

Dr. Frank-N-Furter is presenting a paper at the *Time Warp* conference being held at the Westin Tabor Center hotel in Denver, Colorado. He is presenting his paper at 2:00 pm on

Tuesday, October 7, but he must be present for the opening session at 8:00 am on Monday, October 6 and remain for the duration of the conference, which ends at 3:00 pm on Friday, October 10. He has a travel budget of $1200.

Dr. Frank-N-Furter is traveling from Columbus, Ohio and insists on a non-stop flight to Denver. Since he feels that the Westin Tabor Center is out of his price range, Dr. Frank-N-Furter would like a room at a less-expensive hotel within 12 miles from the conference. This hotel should have the following amenities:

1. Exercise room
2. Internet (wired or wireless)
3. Restaurant/dining room
4. TV with premium channels

Dr. Frank-N-Furter will need to rent a car during his stay in Denver. He does not want to spend more than $75 per day, or $350 total for the car rental

A.2.7 GOAL 7

Mr. Petty is presenting a paper at the *Stock Car Racing* conference being held at the Dunhill hotel in Charlotte, North Carolina. He is presenting his paper at 1:00 pm on Tuesday, September 23, but he must be present for the opening session at 9:00 am on Tuesday, September 23 and remain for the duration of the conference, which ends at 5:00 pm on Friday, September 26. He has a travel budget of $1000.

Mr. Petty is traveling from Detroit, Michigan and insists on a non-stop flight to Charlotte. Since he feels that the Dunhill is out of his price range, Mr. Petty would like a room at a less-expensive hotel within 12 miles from the conference. This hotel should have the following amenities:

1. Wireless Internet
2. Restaurant/dining room

Mr. Petty will need to rent a car during his stay in Charlotte. He does not want to spend more than $65 per day, or $320 total for the car rental.

A.2.8 GOAL 8

Mr. Buffett is presenting a paper at the *Reuben Sandwich* conference being held at the Hilton Garden Inn hotel in Omaha, Nebraska. He is presenting his paper at 11:00 am on Wednesday, October 22, but he must be present for the opening session at 8:00 am on Monday, October 20 and remain for the duration of the conference, which ends at 4:00 pm on Friday, October 24. He has a travel budget of $1200.

Mr. Buffett is traveling from Chicago, Illinois and insists on a non-stop flight to Omaha. Since he feels that the Hilton Garden Inn is out of his price range, Mr. Buffett would like a room at a less-expensive hotel within 8 miles from the conference. This hotel should have the following amenities:

1. Room service

2. Exercise room
3. Internet (wired or wireless)

Mr. Buffett will need to rent a car during his stay in Omaha. He does not want to spend more than $55 per day, or $325 total for the car rental.

A.2.9 GOAL 9

A.2.9.1 GOAL 9A

Ms. Kilcher is presenting a paper at the *Who Will Save Your Soul* conference being held at the Captain Cook hotel in Anchorage, Alaska. She is presenting her paper at 9:00 am on Friday, October 31, but she must be present for the opening session at 8:00 am on Tuesday, October 28 and remain for the duration of the conference, which ends at 3:00 pm on Friday, October 31. She has a travel budget of $2400.

Ms. Kilcher is traveling from Salt Lake City, Utah and insists on a non-stop flight to Anchorage. Since she feels that the Captain Cook is out of her price range, Ms. Kilcher would like a room at a less-expensive hotel within 10 miles from the conference. This hotel should have the following amenities:

1. Restaurant/dining room
2. Exercise room
3. Wireless Internet

Ms. Kilcher will need to rent a car during her stay in Anchorage. She does not want to spend more than $80 per day, or $380 total for the car rental.

A.2.9.2 GOAL 9B

Ms. Kilcher is presenting a paper at the *Who Will Save Your Soul* conference being held at the Captain Cook hotel in Spokane, Washington. She is presenting her paper at 9:00 am on Friday, October 31, but she must be present for the opening session at 8:00 am on Tuesday, October 28 and remain for the duration of the conference, which ends at 3:00 pm on Friday, October 31. She has a travel budget of $2400.

Ms. Kilcher is traveling from Salt Lake City, Utah and insists on a non-stop flight to Spokane. Since she feels that the Davenport is out of her price range, Ms. Kilcher would like a room at a less-expensive hotel within 8 miles from the conference. This hotel should have the following amenities:

1. Restaurant/dining room
2. Exercise room
3. Wireless Internet

Ms. Kilcher will need to rent a car during her stay in Spokane. She does not want to spend more than $80 per day, or $380 total for the car rental.

A.2.10 GOAL 10

Dr. Van Zant is presenting a paper at the *Lynyrd Skynyrd* conference being held at the Omni Jacksonville hotel in Jacksonville, Florida. He is presenting his paper at 11:00 am on Thursday, October 9, but he must be present for the opening session at 9:00 am on Tuesday, October 7 and remain for the duration of the conference, which ends at 2:00 pm on Friday, October 10. He has a travel budget of $1000.

Dr. Van Zant is traveling from Boston, Massachusetts and insists on a non-stop flight to Jacksonville. Since he feels that the Omni Jacksonville is out of his price range, Dr. Van Zant would like a room at a less-expensive hotel within 10 miles from the conference. This hotel should have the following amenities:

1. Internet (wireless or wired)
2. Restaurant/dining room

Dr. Van Zant will need to rent a car during his stay in Jacksonville. He does not want to spend more than $50 per day, or $220 total for the car rental.

APPENDIX B FORMS

B.1 SUBJECT PROFILE

An Effort and Time Based Measure of Usability
Subject Profile

Subject ID: ____________________

Age: __________ Gender (M/F): ____ Race/Ethnicity: ____________

Vision

Do you wear glasses or contact lenses? (Y/N) _______

If yes, then please provide the following information:

What is your vision problem (check all that apply):

□ Near Sighted □ Far Sighted □ Astigmatisms □ Other

Does your correction employ a one or more prisms? (Y/N) _________

Do your glass have a (check all that apply):

□ non-glare coating □ Photo –sensitive

Computer

Approximately how much time do you spend using a computer every day: ________________

Approximately how much time do you spend on the internet: ________

Approximately how frequently do you make on-line travel arrangements? ____________

Approximately how many travel systems have you used? _____________

B.2 POST-GOAL SURVEY

An Effort and Time Based Measure of Usability Survey

Post-Goal Survey

Subject ID: ____________________

Evaluation: ______________ Goal: ______________

How long did it take to complete the goal?

Time End: ________

Time Start: ________

Total Time: ________

Did you complete the goal (Y/N)? _____

Did you meet all of the criteria set forth in the goal (Y/N)? __________

On the seven-point scale with 7 as the most favorable response, 4 the mid-point and 1 the least favorable response please tell us about your experiences during this study:

Experience	Score
General Comfort	
Shoulder Fatigue	
Neck Fatigue	
Eye Fatigue	
Physical Effort	
Mental Effort	

Note: Subject ID, Evaluation, Goal, Time-End, Time-Start, Total time are completed by the observer.

B.3 PARTICIPATION CERTIFICIATE

TEXAS STATE UNIVERSITY®
SAN MARCOS
The rising STAR of Texas

Certificate of Participation:
An Effort and Time Based Measure of Usability

This certifies that ____________________ has participated in the study "An Effort and Time Based Measure of Usability".

____________________ ____________
Research Assistant *Date*

APPENDIX C EVALUATION PROTOCOL

The following is a set of instructions for administering the eye-tracking pilot study. If you have any questions about these instructions, please ask Dr. Tamir, Dr. Komogortsev or Dr. Mueller for clarification.

Text in italics indicates directions that you are to follow. Bolded text indicates instructions that you are to provide to subjects. Please do not substantially deviate from or alter these instructions. Please adhere to these instructions as strictly as possible.

During the course of the experiment, you may be asked questions by subjects. Please do not provide any information other than what is contained in the consent form. If subjects request answers beyond the scope of the consent form, the consent form provides appropriate contact information for such requests.

Functionally blind persons and persons who are physically unable to use a mouse and keyboard while keeping their chin on a chin-rest for fifteen minutes are not eligible to participate in the study. If any ineligible persons volunteer for participation, perform only steps 1-6 and 19.

Please make sure that you read and understand the complete set of instructions before administering the study to any subjects. Do not administer the study until you have been trained to properly calibrate/recalibrate the eye-tracker and start/stop the logging utilities.

1. *Direct the subject to sit in a seat in front of the eye-tracker, then close the lab door most of the way (leaving it open just a crack), and put the "Do Not Disturb" sign on the door.*

2. *State the following:*

    Thank you for volunteering to participate in this study. Before we proceed, I'd like you to carefully review the following statement of informed consent. After reviewing the consent form, if you would like to continue, please sign and put today's date on the line labeled "Subject's Signature" and return the form to me.

3. *Give the subject one copy of "Consent Form: An Effort and Time Based Measure of Usability" that has been signed and dated on the line labeled "Researcher's Signature". After the subject signs and dates the form and returns it to you, sign your name and put today's date on the line labeled "Researcher Assistant's Signature." Place the form facedown on top of the forms in the "Consent Forms" folder.*
4. *Hand the subject one blank unsigned copy of "Consent Form: An Effort and Time Based Measure of Usability".*
5. *Open the coding spreadsheet. Put the subject's name into the next available space. Note the code next to the subject's name. This will be the subject's subject id.*
6. *State the following:*

    This copy of the consent form is yours to keep. We will now proceed with the study. Remember, you may withdraw at any time. If you wish to do so, please let me know and we will discontinue.

    Write the subject's subject ID on a "Subject Profile", hand it to the subject and ask them to complete it and return it to you. When the subject returns the form, place it in the Subject Profiles folder.

7. *If at any point the subject states a desire to discontinue, then immediately stop and skip down to step 19.*
8. *Open Tobii Studio and open the project named "Pilot study." Open a command prompt and in the logs directory, create a new subdirectory named for the subject's subject id.*
9. *On the eye-tracker computer, go to Control Panel, Internet Options, then under "Browsing history" click the "Delete" button, then click the "Delete all…" button, check the "Also delete files and settings stored by add-ons" box, then click "Yes". Next, prepare, but do not start recording, a mouse/keyboard log named* `[subject id]-[exercise #]`. *In Tobii, open a new recording session named* `[subject id]-[exercise #]`.
10. *State the following:*

    Please turn off your cell phone and any other electronic devices that you have with you at this time, and please remove any hats or non-prescription sunglasses that you are wearing.

    We are now going to take some measurements using the eye tracker. Please place your chin on the chin rest and direct your attention to the monitor. You may look at the monitor and blink your eyes as you normally would, but please do not remove your chin from the chin rest or move your head unless you wish to discontinue the experiment.

11. *Direct the subject to place their chin on the chin rest. If necessary, adjust the height of the chin rest so that the subject is looking directly at the monitor. If you have not run any experiments yet, minimize Tobii and state the following:*

In a few moments, you're going to see a circle with a dot in its center on the screen. Please follow the dot with your eyes. Try not to anticipate the movement of the dot. Remember, you may look at the monitor and blink your eyes like you normally would. We may repeat this process a number of times.

Now run the accuracy calibration procedure then skip down to step 13. If the error rate for this procedure is not less than 50% or is not less than 3 degrees in one eye, skip down to step 19.

12. *State the following if necessary:*

    In a few moments, you're going to see a circle with a dot in its center on the monitor. Please follow the dot with your eyes. Try not to anticipate the movement of the dot. Remember, you may look at the monitor and blink your eyes like you normally would. We may repeat this process two or three times.

13. *Calibrate/recalibrate the eye-tracker. Do not make more than three calibration attempts or recalibrate more than twice. If the eye-tracker fails to gather any calibration data after three attempts, instruct the subject that they may now remove their chin from the chin-rest and skip down to step 19.*
14. *State the following:*

    Please hold your head still and keep your chin on the chin rest while I read you some instructions.

Until the conclusion of Step 16, make sure that the subject does not remove their chin from the chin rest unless they wish to discontinue. Make sure they do not obstruct the eye tracker with their free hand.

15. *State the following:*

    You are now going to carry out the exercises which will be described on the sheet in front of you to the best of your ability. You will be using the keyboard and mouse in front of you, which you may adjust at this time.

    Try to follow the directions as closely as possible and as best as you can. These exercises are *not* a test of you or your skills. You are not being evaluated on your ability to complete the exercises or your ability to use a computer system.

    In these exercises, you will be given a task with certain requirements. You should try to meet the requirements as closely as possible, but you may complete the assigned task without precisely fulfilling every requirement.

    You may move your eyes from the monitor to the sheet and back, but please do not move your head or remove your chin from the chin rest unless you wish to discontinue. I cannot communicate with you in any way during the exercise. If at any point you are unsure of how to proceed, simply take whatever steps you think may be correct.

    You will be utilizing an actual travel website for these exercises, but you will not be booking any actual travel or making any actual purchases. Please do

not enter any personal information into the system at any time (I will be monitoring as well to make sure that this doesn't happen).

You will be completing a total of ten exercises today, with periodic breaks. This will take approximately two hours in total.

Would you like me to review any of these instructions?

Review the instructions with the subject if necessary, but do not provide any information other than what is contained in these instructions and the consent form.

16. Ask the subject:

Are you ready to begin?

When the subject indicates that they are ready, place the next (or first if you have not run any exercises yet) goal sheet onto the bracket attached to the monitor. Be sure that the sheet does not obstruct the monitor.

State the following:

Please do not touch the keyboard or mouse until I tell you to begin.

Start the Tobii recording and mouse/keyboard logging. State to the subject:

You may begin.

If the subject asks for assistance, simply state: "I apologize, but I cannot help you." Do not assist the subject with the exercises in any way whatsoever, even if they request assistance. Do not let the subject enter any personal information at any point. The exercise is considered to be completed once a "login to complete this order" message is displayed on-screen, the

Web interface is non-responsive for two minutes, no progress is being made toward the goal for two minutes, or the subject states that they are finished with the exercise. Once the subject completes the exercise or ten minutes have elapsed (whichever comes sooner), stop the logging and recording, and inform the subject that the exercise is complete and they may now remove their chin from the chin-rest.

Write the following in the appropriate fields on an "After Goal" form (please write all times in 24-hour/military format): Subject's subject ID, start time, stop time, elapsed time, website used, and goal number. Now hand the form to the subject and ask them to complete the remaining fields and return the form to you.

17. *State the following:*

    We will now continue with the next exercise.

18. *Repeat steps 9, 11-14, and 16-17 for exercises 2-10. If at any point the subject seems frustrated or upset, assure the subject that they are doing fine and remind them that they are not being personally evaluated or tested.*
19. *State the following:*

    Thank you very much for participating in this study. This concludes your participation. Please take your copy of the consent form with you, and thank you again.

    If the subject desires proof of participation, sign and date a "Proof of Participation" form and give it to the subject. Inform the subject that they may show or not show this certificate to anyone completely at their discretion.

Dismiss the subject. If the subject wishes to discuss the study, you may do so with her or him at this time.

If the subject completed the experiment, then on the Coding Spreadsheet, in the "Completed experiment?" column, put "Yes."

If the subject did not meet the participation criteria, then on the Coding Spreadsheet, in the "Completed experiment?" column, put "No: Ineligible."

If the eye-tracker could not be calibrated for the subject, then on the Coding Spreadsheet in the "Completed experiment?" column, put "No: Failed calibration."

If the subject discontinued the experiment, then on the Coding Spreadsheet, in the "Completed experiment?" column, put "No: " and note the point at which the subject discontinued. If the subject completed any forms, file them in the appropriate folder.

APPENDIX D RAW DATA

D.1 SUBJECT PROFILES

D.1.1 SYSTEM A

ID	Age	Gender	Race	Glasses	Computer Usage	Internet Usage	Travel System Usage	Travel Systems
P0-101	20	M	C	N	2.5	2	2	4
P0-102	25	M	C	N	4	4	1	4
P0-103	26	M	C	Y	8	2	0	0
P0-104	22	M	H	N	1	1	0	0
P0-105	19	F	C	N	1	1	0	0
P0-106	23	F	O	N	6	4	2	4
P0-107	31	F	C	N	3	3	3	5
P0-108	29	M	C	Y	8	6	2	5
P0-110	26	M	C	N	5	3	4	3
P0-111	26	M	H	Y	7	2	0	0

D.1.2 SYSTEM B

ID	Age	Gender	Race	Glasses	Computer Usage	Internet Usage	Travel System Usage	Travel Systems
P1-113	28	M	C	Y	1	1	0	0
P1-114	34	F	H	N	3	2	0	2
P1-115	22	M	C	Y	4	3	2	3
P1-116	21	F	B	Y	4	4	2	2
P1-118	22	F	C	N	2	2	0	0
P1-119	22	M	C	Y	1	1	0	3
P1-121	34	M	O	Y	5	4	5	6
P1-122	24	M	B	Y	2	1	1	3
P1-124	34	M	H	N	6	3	2	4
P1-125	29	F	B	N	16	10	2	2

D.2 RAW DATA

D.2.1 MANUAL

D.2.2 SYSTEM A

ID	Mickeys	Clicks	Keystrokes	Transfers
p0-101-01.txt	59925	67	65	10
p0-101-02.txt	62224	75	83	8
p0-101-03.txt	34203	61	65	13
p0-101-04.txt	47073	55	82	9
p0-101-05.txt	57189	74	39	7
p0-101-06.txt	32966	32	74	9
p0-101-07.txt	49919	58	62	5
p0-101-08.txt	37287	42	36	19
p0-101-09.txt	44172	53	70	5
p0-101-10.txt	71298	69	68	11
p0-102-01.txt	37652	42	97	10
p0-102-02.txt	36324	89	45	15
p0-102-03.txt	37272	57	56	17
p0-102-04.txt	37528	56	95	23
p0-102-05.txt	27762	86	64	16
p0-102-06.txt	31408	49	52	16
p0-102-07.txt	37260	72	56	11
p0-102-08.txt	33649	64	54	15
p0-102-09.txt	21441	28	70	14
p0-102-10.txt	42831	69	49	10

ID	Mickeys	Clicks	Keystrokes	Transfers
p0-103-01.txt	84223	57	101	31
p0-103-02.txt	115581	115	100	17
p0-103-03.txt	114091	73	78	14
p0-103-04.txt	103254	98	112	17
p0-103-05.txt	109210	110	71	12
p0-103-06.txt	87817	92	101	16
p0-103-07.txt	111071	85	113	16
p0-103-08.txt	86968	96	101	10
p0-103-09.txt	55983	73	107	18
p0-103-10.txt	169197	134	182	26
p0-104-01.txt	26870	44	203	33
p0-104-02.txt	39259	43	55	14
p0-104-03.txt	44628	34	71	12
p0-104-04.txt	39326	54	109	15
p0-104-05.txt	33140	44	52	7
p0-104-06.txt	41238	49	62	8
p0-104-07.txt	36171	58	78	16
p0-104-08.txt	40414	63	64	7
p0-104-09.txt	28969	49	69	7
p0-104-10.txt	46400	81	82	8

ID	Mickeys	Clicks	Keystrokes	Transfers

ID	Mickeys	Clicks	Keystrokes	Transfers

p0-105-01.txt	36669	71	113	25
p0-105-02.txt	44376	72	73	9
p0-105-03.txt	17934	45	34	7
p0-105-04.txt	45459	71	70	9
p0-105-05.txt	26226	51	43	5
p0-105-06.txt	28374	57	40	5
p0-105-07.txt	27806	68	41	8
p0-105-08.txt	36269	64	97	13
p0-105-09.txt	32839	68	49	6
p0-105-10.txt	25350	42	38	11
p0-106-01.txt	39166	59	118	19
p0-106-02.txt	41825	56	57	12
p0-106-03.txt	36203	57	58	10
p0-106-04.txt	36121	59	127	28
p0-106-05.txt	57417	103	65	14
p0-106-06.txt	33252	51	46	13
p0-106-07.txt	49292	71	62	13
p0-106-08.txt	41580	84	86	18
p0-106-09.txt	48065	93	85	9
p0-106-10.txt	42015	80	69	10

p0-107-01.txt	55565	48	106	14
p0-107-02.txt	44944	55	58	9
p0-107-03.txt	44162	52	79	9
p0-107-04.txt	46771	45	93	24
p0-107-05.txt	41358	57	93	17
p0-107-06.txt	51211	62	87	15
p0-107-07.txt	36871	39	78	12
p0-107-08.txt	41822	58	89	13
p0-107-09.txt	35380	41	94	14
p0-107-10.txt	37572	50	67	19
p0-108-01.txt	30441	33	87	7
p0-108-02.txt	69525	112	56	7
p0-108-03.txt	37546	41	44	6
p0-108-04.txt	28377	54	54	6
p0-108-05.txt	21519	32	41	7
p0-108-06.txt	14316	26	39	17
p0-108-07.txt	21845	28	34	6
p0-108-08.txt	30621	38	34	7
p0-108-09.txt	24196	40	88	7
p0-108-10.txt	12967	25	33	4

ID	Mickeys	Clicks	Keystrokes	Transfers
p0-110-01.txt	57518	51	59	8
p0-110-02.txt	54619	53	71	10
p0-110-03.txt	54163	56	47	4
p0-110-04.txt	55121	73	100	5
p0-110-05.txt	59246	86	102	14
p0-110-06.txt	28497	36	64	10
p0-110-07.txt	17221	32	45	8
p0-110-08.txt	64736	103	141	19
p0-110-09.txt	39181	50	41	6
p0-110-10.txt	48276	84	62	10
p0-111-01.txt	18452	23	44	5
p0-111-02.txt	30461	47	18	6
p0-111-03.txt	15452	28	23	7
p0-111-04.txt	25979	39	64	8
p0-111-05.txt	9669	19	22	8
p0-111-06.txt	14508	33	25	6
p0-111-07.txt	14158	31	21	5
p0-111-08.txt	14605	34	17	5
p0-111-09.txt	10109	21	26	8
p0-111-10.txt	15890	28	21	5

D.2.3 SYSTEM B

ID	Mickeys	Clicks	Keystrokes	Transfers
p1-113-01.txt	25496	30	65	6
p1-113-02.txt	32373	57	29	6
p1-113-03.txt	24352	35	37	8
p1-113-04.txt	40990	70	48	6
p1-113-05.txt	22596	39	22	6
p1-113-06.txt	16164	31	17	6
p1-113-07.txt	14198	30	23	8
p1-113-08.txt	15771	33	15	6
p1-113-09.txt	19102	43	48	8
p1-113-10.txt	17435	41	23	9
p1-114-01.txt	24110	39	89	14
p1-114-02.txt	37744	54	53	9
p1-114-03.txt	22017	30	31	6
p1-114-04.txt	35973	71	39	5
p1-114-05.txt	32478	45	37	11
p1-114-06.txt	31579	43	26	5
p1-114-07.txt	26942	63	28	5
p1-114-08.txt	20788	47	31	6
p1-114-09.txt	33142	54	63	10
p1-114-10.txt	25213	46	30	7

ID	Mickeys	Clicks	Keystrokes	Transfers
p1-115-01.txt	31187	37	69	12
p1-115-02.txt	46576	42	43	8
p1-115-03.txt	30152	23	60	11
p1-115-04.txt	46615	51	50	7
p1-115-05.txt	48553	50	44	8
p1-115-06.txt	24463	31	37	8
p1-115-07.txt	36871	30	68	9
p1-115-08.txt	29570	52	38	8
p1-115-09.txt	28419	34	51	20
p1-115-10.txt	54749	46	48	8
p1-116-01.txt	58692	49	40	7
p1-116-02.txt	53099	51	34	6
p1-116-03.txt	49427	50	46	6
p1-116-04.txt	27417	36	50	10
p1-116-05.txt	63310	46	41	7
p1-116-06.txt	46449	47	35	6
p1-116-07.txt	19149	23	53	5
p1-116-08.txt	26410	35	35	6
p1-116-09.txt	16792	27	42	5
p1-116-10.txt	24773	33	49	6

ID	Mickeys	Clicks	Keystrokes	Transfers
p1-118-01.txt	60113	85	39	5
p1-118-02.txt	73785	60	20	6
p1-118-03.txt	35543	43	41	9
p1-118-04.txt	49887	60	37	16
p1-118-05.txt	32291	41	42	7
p1-118-06.txt	45429	53	22	6
p1-118-07.txt	32288	45	22	4
p1-118-08.txt	18049	34	17	5
p1-118-09.txt	34756	57	68	9
p1-118-10.txt	23404	40	30	4
p1-119-01.txt	49385	44	46	7
p1-119-02.txt	43051	40	33	11
p1-119-03.txt	30424	40	36	5
p1-119-04.txt	22290	38	48	6
p1-119-05.txt	15720	24	31	5
p1-119-06.txt	18162	23	26	6
p1-119-07.txt	26415	23	44	6
p1-119-08.txt	34465	40	40	7
p1-119-09.txt	24121	34	43	8
p1-119-10.txt	21723	35	34	4

ID	Mickeys	Clicks	Keystrokes	Transfers
p1-121-01.txt	35446	57	164	12
p1-121-02.txt	29580	55	78	21
p1-121-03.txt	52300	71	110	41
p1-121-04.txt	32250	37	107	24
p1-121-05.txt	46909	58	126	19
p1-121-06.txt	29942	49	160	26
p1-121-07.txt	32440	44	133	16
p1-121-08.txt	42678	56	49	7
p1-121-09.txt	31686	44	59	15
p1-121-10.txt	41324	59	170	23
p1-122-01.txt	76382	56	83	14
p1-122-02.txt	63457	89	39	19
p1-122-03.txt	34762	41	58	8
p1-122-04.txt	21982	36	57	7
p1-122-05.txt	40291	55	49	23
p1-122-06.txt	29434	35	45	19
p1-122-07.txt	16221	29	28	12
p1-122-08.txt	19055	38	63	9
p1-122-09.txt	33602	40	49	12
p1-122-10.txt	17918	36	33	9

ID	Mickeys	Clicks	Keystrokes	Transfers
p1-124-01.txt	48117	63	109	10
p1-124-02.txt	40552	47	27	6
p1-124-03.txt	18936	28	42	5
p1-124-04.txt	41571	61	92	16
p1-124-05.txt	41717	47	24	6
p1-124-06.txt	35731	47	45	7
p1-124-07.txt	23476	36	23	5
p1-124-08.txt	37088	39	16	7
p1-124-09.txt	36439	52	64	10
p1-124-10.txt	64927	74	49	11
p1-125-01.txt	34708	33	56	9
p1-125-02.txt	34311	48	45	9
p1-125-03.txt	22231	43	69	6
p1-125-04.txt	37765	64	102	8
p1-125-05.txt	44632	54	53	7
p1-125-06.txt	41935	51	38	10
p1-125-07.txt	38480	61	62	9
p1-125-08.txt	30865	52	82	12
p1-125-09.txt	29964	48	61	7
p1-125-10.txt	19175	31	48	6

D.3 EYE DATA

D.3.1 SYSTEM A

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled (deg)	Average pupil dilation (mm)
p0-101-01.tsv	599	63	3.42	944	739	443	85.22	14.78	4498	3.58
p0-101-02.tsv	496	81	3.72	1169	670	474	53.24	46.76	3838	3.55
p0-101-03.tsv	357	81	4	760	564	419	72.43	27.57	3469	3.35
p0-101-04.tsv	446	93	4.42	1096	764	371	42.82	57.18	4607	3.24
p0-101-05.tsv	392	84	4.04	807	637	366	47.26	52.74	3772	3.16
p0-101-06.tsv	289	89	3.91	669	421	478	49.63	50.37	2299	3.14
p0-101-07.tsv	331	89	3.87	769	473	478	42.93	57.07	2348	3.14
p0-101-08.tsv	244	88	4.34	597	372	353	33.03	66.97	2200	3.06
p0-101-09.tsv	319	92	4.27	669	533	330	29.31	70.69	2689	3.06
p0-101-10.tsv	440	86	3.81	947	629	518	69.54	30.46	3626	3.05
p0-102-01.tsv	589	90	4.05	1374	1023	463	80.6	19.4	6490	3.62
p0-102-02.tsv	535	91	4.34	1305	974	439	76.09	23.91	6699	3.48
p0-102-03.tsv	581	91	4.2	1380	1028	451	77.45	22.55	6368	3.42
p0-102-04.tsv	618	92	4.06	1498	1039	484	79.7	20.3	6700	3.41
p0-102-05.tsv	577	90	3.76	1268	891	523	78.12	21.88	5400	3.32
p0-102-06.tsv	464	90	4.26	992	723	515	77.08	22.92	5011	3.37
p0-102-07.tsv	521	92	4.23	1185	856	496	79.17	20.83	5679	3.4
p0-102-08.tsv	435	90	4.59	909	735	460	75.21	24.79	5066	3.38
p0-102-09.tsv	266	86	4.51	594	465	420	76.15	23.85	3415	3.39
p0-102-10.tsv	523	92	4.34	1233	923	464	79.94	20.06	6155	3.41
p0-103-01.tsv	640	82	3.62	453	888	229	21.64	78.36	4367	2.65
p0-103-02.tsv	576	89	3.93	590	884	263	23.77	76.23	4848	2.6
p0-103-03.tsv	578	89	3.68	648	946	260	24.55	75.45	5332	2.57
p0-103-04.tsv	560	85	3.64	401	773	225	19.81	80.19	4602	2.54

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled (deg)	Average pupil dilation (mm)
p0-103-05.tsv	402	87	3.84	425	602	261	23.4	76.6	3351	2.49
p0-103-06.tsv	483	88	3.74	434	730	265	22.46	77.54	4136	2.48
p0-103-07.tsv	467	88	3.5	444	717	261	22.83	77.17	3675	2.45
p0-103-08.tsv	463	88	3.76	529	795	279	26.67	73.33	4542	2.46
p0-103-09.tsv	334	86	4.08	306	492	246	22.47	77.53	2772	2.5
p0-103-10.tsv	595	87	4.19	519	895	242	21.34	78.66	5434	2.53
p0-104-01.tsv	615	71	3.27	677	395	932	52	48	2053	3.37
p0-104-02.tsv	596	86	3.92	1154	694	596	50.49	49.51	4102	3.33
p0-104-03.tsv	602	90	3.68	1020	650	674	52.21	47.79	3903	3.33
p0-104-04.tsv	618	88	3.86	1172	771	592	51.45	48.55	4918	3.34
p0-104-05.tsv	598	90	4.02	1081	770	597	52.29	47.71	5054	3.37
p0-104-06.tsv	625	90	4.22	1071	829	559	46.29	53.71	5262	3.33
p0-104-07.tsv	463	93	4.23	795	734	449	38.86	61.14	4265	3.41
p0-104-08.tsv	606	94	4.46	1132	933	482	42.62	57.38	5850	3.39
p0-104-09.tsv	379	88	4.61	579	514	455	36.47	63.53	3727	3.34
p0-104-10.tsv	525	90	4.69	837	752	471	38.47	61.53	4599	3.35
p0-105-01.tsv	619	67	4.1	900	679	548	84.47	15.53	4982	3.6
p0-105-02.tsv	619	84	4.19	1243	790	596	77.1	22.9	5481	3.18
p0-105-03.tsv	453	69	4.71	733	505	554	73.8	26.2	4011	3.13
p0-105-04.tsv	580	62	4.66	865	653	487	75.29	24.71	4882	3.17
p0-105-05.tsv	439	70	4.85	713	560	491	72.63	27.37	3995	3.12
p0-105-06.tsv	390	79	4.96	744	537	510	70.19	29.81	4240	3.14
p0-105-07.tsv	335	83	4.79	601	428	586	69.51	30.49	3250	3.12
p0-105-08.tsv	493	77	4.95	797	638	536	75.08	24.92	4779	3.24
p0-105-09.tsv	408	78	4.28	712	535	534	71.87	28.13	3570	3.13
p0-105-10.tsv	342	75	4.64	569	417	563	77.38	22.62	3084	3.24
p0-106-01.tsv	591	71	3.71	1149	717	470	57.89	42.11	4303	3.83
p0-106-02.tsv	593	70	3.71	1097	755	410	41.2	58.8	4009	3.66

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled (deg)	Average pupil dilation (mm)
p0-106-03.tsv	575	79	4.14	1154	822	418	48.47	51.53	5052	3.68
p0-106-04.tsv	597	71	4.75	1087	857	405	70.33	29.67	6323	3.58
p0-106-05.tsv	649	80	4.34	1252	931	411	41.46	58.54	5550	3.48
p0-106-06.tsv	428	78	5.19	825	604	459	68.13	31.87	4903	3.54
p0-106-07.tsv	423	66	5.13	614	498	490	78.94	21.06	3736	3.57
p0-106-08.tsv	331	79	4.76	645	520	435	75.84	24.16	3368	3.58
p0-106-09.tsv	424	81	4.54	797	649	459	76.8	23.2	4496	3.56
p0-106-10.tsv	392	83	5.3	803	674	412	72.66	27.34	4915	3.48
p0-107-01.tsv	597	96	4.94	1647	1119	447	75.69	24.31	8499	3.18
p0-107-02.tsv	600	95	5.02	1582	1149	435	76.99	23.01	8641	3.12
p0-107-03.tsv	550	93	4.79	1408	1074	414	74.5	25.5	7921	3.08
p0-107-04.tsv	603	92	4.8	1215	960	401	56.24	43.76	7248	3.07
p0-107-05.tsv	514	93	5.13	1341	992	413	70.11	29.89	8004	3.06
p0-107-06.tsv	602	83	4.92	1296	1025	404	65.43	34.57	8190	2.97
p0-107-07.tsv	387	87	5.26	780	668	418	62.27	37.73	5689	2.95
p0-107-08.tsv	557	88	5.23	883	840	354	45.11	54.89	6536	2.95
p0-107-09.tsv	397	89	5.19	601	590	388	47.96	52.04	4684	2.92
p0-107-10.tsv	430	91	5.47	984	768	431	66.41	33.59	6489	2.93
p0-108-01.tsv	391	94	3.78	961	659	496	73.5	26.5	3983	3.36
p0-108-02.tsv	637	96	3.66	1618	1044	531	70.23	29.77	5814	3.3
p0-108-03.tsv	312	95	3.98	761	512	524	74.76	25.24	3323	3.34
p0-108-04.tsv	337	95	3.66	803	521	560	71.94	28.06	3010	3.27
p0-108-05.tsv	201	94	3.66	435	316	547	69.8	30.2	1921	3.28
p0-108-06.tsv	193	95	3.57	418	282	594	71.94	28.06	1603	3.27
p0-108-07.tsv	238	95	4.15	544	385	533	70.77	29.23	2459	3.32
p0-108-08.tsv	282	95	4.22	629	439	555	71.57	28.43	2935	3.33
p0-108-09.tsv	302	94	3.87	615	461	555	74.82	25.18	2735	3.37
p0-108-10.tsv	145	95	4.52	321	226	538	66.63	33.37	1653	3.22

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled (deg)	Average pupil dilation (mm)
p0-110-01.tsv	659	96	4.81	1726	1235	449	77.03	22.97	8426	3.64
p0-110-02.tsv	569	92	4.81	1338	986	454	76.11	23.89	7041	3.73
p0-110-03.tsv	565	95	4.96	1299	918	524	77.07	22.93	6732	3.67
p0-110-04.tsv	605	95	4.74	1401	991	507	74.76	25.24	6888	3.62
p0-110-05.tsv	596	94	5.11	1237	1046	390	54.98	45.02	7203	3.65
p0-110-06.tsv	306	93	4.83	727	549	454	72.14	27.86	3829	3.62
p0-110-07.tsv	183	90	5	397	284	485	74.1	25.9	2173	3.68
p0-110-08.tsv	586	92	4.97	1258	932	494	71.14	28.86	6780	3.63
p0-110-09.tsv	386	96	5.21	891	677	481	72.27	27.73	4850	3.58
p0-110-10.tsv	366	93	5.14	859	650	457	72.5	27.5	4785	3.58
p0-111-01.tsv	394	89	4.15	824	700	334	37.71	62.29	4684	3.35
p0-111-02.tsv	477	89	4.09	839	864	334	36.73	63.27	5369	3.3
p0-111-03.tsv	313	89	4.14	572	539	341	37.87	62.13	3647	3.28
p0-111-04.tsv	461	90	4.03	815	790	356	37.52	62.48	5276	3.27
p0-111-05.tsv	237	88	4.44	468	406	346	39.38	60.62	3035	3.26
p0-111-06.tsv	266	86	4.19	416	441	317	33.67	66.33	2982	3.17
p0-111-07.tsv	197	85	3.78	247	318	307	31.84	68.16	2310	3.2
p0-111-08.tsv	251	87	4.02	426	403	314	32.06	67.94	2638	3.21
p0-111-09.tsv	188	91	4.16	328	317	347	34.14	65.86	2346	3.2
p0-111-10.tsv	274	86	3.97	426	450	319	34.24	65.76	3028	3.2

D.3.2 SYSTEM B

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled deg	Average pupil dilation (mm)
p1-113-01.tsv	304	86	4.11	871	461	296	30.82	69.18	2491	2.88
p1-113-02.tsv	298	97	5.46	120	106	157	6.97	93.03	711	2.78
p1-113-03.tsv	252	87	4.99	489	330	219	21.6	78.4	2079	2.87
p1-113-04.tsv	290	87	4.95	621	368	236	22.22	77.78	2549	2.85
p1-113-05.tsv	207	86	5.27	363	267	221	21.18	78.82	1899	2.9
p1-113-06.tsv	157	91	5.69	367	240	263	25.69	74.31	1720	3.01
p1-113-07.tsv	113	88	5.85	290	159	240	24.51	75.49	1371	3.03
p1-113-08.tsv	110	89	5.17	317	187	253	29.49	70.51	1436	3.05
p1-113-09.tsv	120	87	5.22	342	214	259	30.83	69.17	1338	3.05
p1-113-10.tsv	136	88	5.06	344	220	242	27.38	72.62	1506	3.02
p1-114-01.tsv	308	98	5.43	335	255	179	10.98	89.02	1742	2.77
p1-114-03.tsv	189	97	7.39	81	58	148	6.77	93.23	544	2.72
p1-114-04.tsv	311	95	6.04	166	127	155	7.77	92.23	1110	2.78
p1-114-05.tsv	235	92	6.1	72	55	156	6.17	93.83	426	2.68
p1-114-06.tsv	277	90	6.63	46	60	133	6.12	93.88	771	2.72
p1-114-07.tsv	235	90	6.73	33	53	141	6.33	93.67	495	2.77
p1-114-08.tsv	173	90	5.28	29	53	153	7.07	92.93	420	2.78
p1-114-09.tsv	201	92	4.06	38	66	150	7.59	92.41	510	2.74
p1-114-10.tsv	204	91	6.1	25	49	144	6.68	93.32	458	2.71
p1-115-01.tsv	294	87	4.22	944	602	285	45.48	54.52	4404	3.35
p1-115-02.tsv	215	87	4.47	546	439	245	37.95	62.05	2952	3.31
p1-115-03.tsv	147	88	4.93	499	317	278	47.14	52.86	2399	3.43
p1-115-04.tsv	158	89	4.62	518	324	310	49.77	50.23	2396	3.36
p1-115-05.tsv	197	85	4.26	509	364	316	44.8	55.2	2547	3.35
p1-115-06.tsv	95	90	4.6	317	195	310	50.15	49.85	1391	3.46

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled deg	Average pupil dilation (mm)
p1-115-07.tsv	132	93	4.87	448	277	333	47.94	52.06	2022	3.45
p1-115-08.tsv	141	88	4.68	448	315	285	47.86	52.14	2176	3.48
p1-115-09.tsv	111	88	5.46	339	222	318	50.07	49.93	2042	3.54
p1-115-10.tsv	200	89	4.16	543	391	330	49.52	50.48	2611	3.55
p1-116-01.tsv	315	82	6.55	185	177	183	11.29	88.71	1513	2.45
p1-116-02.tsv	293	77	5.71	158	151	187	12.26	87.74	1457	2.48
p1-116-03.tsv	264	80	4.8	130	148	206	11.47	88.53	1337	2.48
p1-116-04.tsv	154	78	8.57	47	53	169	8.12	91.88	569	2.46
p1-116-05.tsv	290	77	6.15	132	133	176	9.65	90.35	928	2.49
p1-116-06.tsv	189	80	5.33	82	106	188	11.22	88.78	858	2.51
p1-116-07.tsv	118	81	7.02	63	69	167	11.6	88.4	405	2.52
p1-116-08.tsv	123	83	6.21	76	86	183	12.25	87.75	853	2.47
p1-116-09.tsv	111	76	6.62	72	91	178	16.08	83.92	756	2.62
p1-116-10.tsv	129	85	6.47	58	79	170	11.18	88.82	837	2.54
p1-118-01.tsv	306	89	3.94	722	510	258	28.09	71.91	3030	2.86
p1-118-02.tsv	304	87	4.38	490	442	222	23.68	76.32	3666	2.84
p1-118-03.tsv	234	86	5.29	434	347	269	29.98	70.02	3270	2.93
p1-118-04.tsv	230	88	4.78	432	365	273	29.56	70.44	2694	2.85
p1-118-05.tsv	186	87	5.37	322	274	224	24.12	75.88	2355	2.85
p1-118-06.tsv	260	88	4.72	364	355	240	23.01	76.99	2746	2.82
p1-118-07.tsv	160	88	4.84	274	248	239	26.45	73.55	2133	2.87
p1-118-08.tsv	94	86	4.61	143	153	219	24.33	75.67	1365	2.88
p1-118-09.tsv	254	86	4.76	453	385	242	26.86	73.14	3247	2.85
p1-118-10.tsv	129	88	4.36	169	167	236	22.13	77.87	1433	2.84
p1-119-01.tsv	308	91	3.35	961	558	299	36.67	63.33	3475	3.01
p1-119-02.tsv	238	90	3.79	907	498	285	42.61	57.39	3626	2.91
p1-119-03.tsv	183	89	4.25	300	210	268	23.47	76.53	1402	2.91
p1-119-04.tsv	129	84	4.62	152	114	225	17.82	82.18	919	2.94

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled deg	Average pupil dilation (mm)
p1-119-05.tsv	84	83	4.18	223	139	352	46	54	1068	3.08
p1-119-06.tsv	121	85	3.94	133	96	273	19.06	80.94	1046	2.85
p1-119-07.tsv	95	85	4.11	157	107	311	26.66	73.34	854	2.79
p1-119-08.tsv	170	86	4.57	292	217	273	24.8	75.2	1793	2.87
p1-119-09.tsv	162	87	4.02	344	241	285	30.26	69.74	1621	2.96
p1-119-10.tsv	122	87	3.89	327	221	334	41.99	58.01	1611	2.91
p1-121-01.tsv	247	93	4.72	715	503	221	29.94	70.06	3818	2.98
p1-121-02.tsv	168	92	5.89	332	285	229	24.25	75.75	2446	2.94
p1-121-03.tsv	234	92	5.09	518	408	221	24.49	75.51	3036	2.87
p1-121-04.tsv	134	89	4.95	299	241	243	27.3	72.7	1801	2.86
p1-121-05.tsv	189	91	6.04	392	324	213	24.08	75.92	2866	2.84
p1-121-06.tsv	178	89	5.1	344	282	242	24.76	75.24	2130	2.82
p1-121-07.tsv	176	91	5.03	303	284	242	24.36	75.64	2081	2.75
p1-121-08.tsv	158	92	5.44	316	286	230	25.83	74.17	1960	2.73
p1-121-09.tsv	138	91	5.64	270	231	228	24.17	75.83	1603	2.77
p1-121-10.tsv	151	90	5.76	289	248	221	23.55	76.45	2093	2.83
p1-122-01.tsv	307	93	4.99	932	574	263	31.55	68.45	4744	3.33
p1-122-02.tsv	312	83	5.12	795	537	278	32.7	67.3	4188	3.17
p1-122-03.tsv	265	91	5.45	582	450	237	25.51	74.49	3883	3.28
p1-122-04.tsv	147	94	4.99	402	245	238	26.27	73.73	1976	3.17
p1-122-05.tsv	160	89	5.22	285	211	248	21.92	78.08	1740	3.25
p1-122-06.tsv	140	86	6.47	326	238	230	27.41	72.59	2459	3.24
p1-122-07.tsv	102	40	5.6	95	75	235	26.11	73.89	801	3.25
p1-122-08.tsv	128	78	6.59	187	177	237	24.81	75.19	1619	3.29
p1-122-09.tsv	269	71	5.51	386	336	254	26.72	73.28	3439	3.26
p1-122-10.tsv	123	84	5.98	259	201	245	28.94	71.06	1938	3.26
p1-124-01.tsv	305	84	4.48	336	409	210	19.74	80.26	2697	3.31
p1-124-02.tsv	288	88	4.07	331	382	209	18.84	81.16	2700	3.17

Name	Duration (min)	Validity LE	Average sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled deg	Average pupil dilation (mm)
p1-124-03.tsv	172	88	3.95	189	212	205	17.26	82.74	1676	3.11
p1-124-04.tsv	306	86	4.26	304	372	208	18.03	81.97	2708	3.1
p1-124-05.tsv	304	86	4.02	274	333	201	16.65	83.35	2645	3.06
p1-124-06.tsv	303	78	4.08	326	392	205	20.54	79.46	2883	3.07
p1-124-07.tsv	189	81	3.72	226	237	226	21.15	78.85	1663	3.06
p1-124-08.tsv	263	80	3.93	232	290	214	18.06	81.94	2201	3.02
p1-124-09.tsv	257	82	4.36	260	298	209	18.68	81.32	2229	3.03
p1-124-10.tsv	312	84	4.43	261	371	204	17.68	82.32	2597	3.02
p1-125-01.tsv	306	71	4.79	422	416	242	29.12	70.88	2638	3.82
p1-125-02.tsv	288	82	4.45	476	403	281	30.95	69.05	2801	3.89
p1-125-03.tsv	248	66	3.96	379	363	248	35.68	64.32	2586	3.71
p1-125-04.tsv	304	64	4.68	484	429	261	35.69	64.31	3573	3.71
p1-125-05.tsv	295	75	5.11	361	375	260	28.35	71.65	2812	3.71
p1-125-06.tsv	229	77	5.22	299	286	246	26.6	73.4	2438	3.58
p1-125-07.tsv	263	77	5.3	261	275	263	24.65	75.35	2154	3.62
p1-125-08.tsv	206	76	4.64	187	229	255	24.65	75.35	1537	3.52
p1-125-09.tsv	237	76	4.82	274	264	283	26.71	73.29	2015	3.6
p1-125-10.tsv	137	74	5.57	198	177	229	25.79	74.21	1565	3.46

D.4 SUMMARIZED BY SYSTEM

D.4.1 SYSTEM A

D.4.2 EYE DATA SUMMARY

Goal	Duration	Validity LE	Average. sac. amp.	Sac count	Fix count	Average Fix duration	Fix per	Sac per	Eye path travelled (deg)	Average pupil dilation (mm)
1	569.40	81.90	3.99	1065.50	815.40	481.10	64.58	35.43	5228.50	3.42
2	569.80	87.30	4.14	1193.50	881.00	453.20	58.20	41.81	5584.20	3.33
3	488.60	87.10	4.23	973.50	755.80	457.90	61.31	38.69	4975.80	3.29
4	542.50	86.30	4.26	1035.30	811.90	438.80	57.99	42.01	5445.40	3.25
5	460.50	87.00	4.32	902.70	715.10	434.50	54.94	45.06	4728.50	3.22
6	404.60	87.10	4.38	759.20	614.10	455.50	57.70	42.30	4245.50	3.20
7	354.50	86.80	4.39	637.60	536.10	450.30	57.12	42.88	3558.40	3.22
8	424.80	87.80	4.53	780.50	660.70	426.20	54.83	45.17	4469.40	3.22
9	340.30	88.10	4.47	609.20	523.30	421.50	54.23	45.77	3528.40	3.21
10	403.20	87.80	4.61	749.80	638.40	441.50	59.91	40.09	4376.80	3.20

D.4.3 MANUAL DATA SUMMARY

Goal	Mickeys	Clicks	Keystrokes	Transfers
1	44648.1	49.5	99.3	16.2
2	53913.8	71.7	61.6	10.7
3	43565.4	50.4	55.5	9.9
4	46500.9	60.4	90.6	14.4
5	44273.6	66.2	59.2	10.7
6	36358.7	48.7	59.0	11.5
7	40161.4	54.2	59.0	10.0
8	42795.1	64.6	71.9	12.6
9	34033.5	51.6	69.9	9.4
10	51179.6	66.2	67.1	11.4

D.4.4 SYSTEM B

D.4.5 EYE DATA SUMMARY

Goal	Duration	Validity LE	Average. sac. amp.	Sac count	Fix counter	Fix average duration	Fix per	Sac per	Eye path travelled (deg)	Average pupil dilation (mm)
1	298.80	87.40	4.66	642.30	457.00	243.60	28.55	72.63	3055.20	3.08
2	258.40	88.00	5.07	423.60	349.91	224.10	24.18	76.30	2509.10	3.02
3	218.40	86.20	4.88	368.60	288.73	230.60	25.38	75.56	2277.80	3.04
4	209.60	85.10	5.25	333.10	260.73	231.90	25.09	75.91	1961.10	3.00
5	215.20	84.90	5.23	290.70	244.09	234.40	25.21	75.71	1963.10	3.03
6	188.10	85.40	5.19	259.10	222.00	233.80	24.49	76.52	1816.60	3.01
7	154.30	81.40	5.16	214.60	180.55	240.90	25.30	75.95	1390.40	3.01
8	176.00	85.00	4.99	223.60	226.55	229.90	24.19	76.03	1545.00	3.01
9	167.30	83.50	5.25	276.50	224.27	240.00	26.39	74.29	1874.80	3.04
10	169.50	85.60	4.99	339.20	257.73	249.60	29.10	70.64	2059.50	3.08

D.4.6 MANUAL DATA SUMMARY

Goal	Mickeys	Clicks	Keystrokes	Transfers
1	44363.6	49.3	76.0	9.6
2	45452.8	54.3	40.1	10.1
3	32014.4	40.4	53.0	10.5
4	35674.0	52.4	63.0	10.5
5	38849.7	45.9	46.9	9.9
6	31928.8	41.0	45.1	9.9
7	26648.0	38.4	48.4	7.9
8	27473.9	42.6	38.6	7.3
9	28802.3	43.3	54.8	10.4
10	31064.1	44.1	51.4	8.7

Bibliography

Andre, T. S., Hartson, H. R., Belz, S. M., and McCreary, F. A. (2001). "The user action framework: a reliable foundation for usability engineering support tools." *Int. J. Hum.-Comput. Stud.*, 54(1), 107-136.

Andreassi, J. L. (1995). *Psychophysiology: Human Behavior and Physiological Response*, Hillsdale, NJ: Lawrence Erlbaum.

Beizer, B. (1990). *Software Testing Techniques 2nd Edition*, Boston, MA: Thomson Inter-Science.

Bevan, N. (1999). "Quality in use: Meeting user needs for quality." *J. Syst. Softw.*, 49(1), 89-96.

Blandford, A., Green, T. R. G., Furniss, D., and Makri, S. (2008). "Evaluating system utility and conceptual fit using CASSM." *Int. J. Hum.-Comput. Stud.*, 66(6), 393-409.

Caulton, D. A. (2001). "Relaxing the homogeneity assumption in usability testing." *Behavior & Information Technology*, 20(1), 7.

Chartette, R. N. (2005). "Why Software Fails"*Spectrum*. City: IEEE.

Crosby, M., Iding, M., and Chin, D. (2009). "Visual Search and Background Complexity: Does the Forest Hide the Trees?", *User Modeling 2001*. pp. 225-227.

Dennerlein, J. T., and Johnson, P. W. (2006). "Different computer tasks affect the exposure of the upper extremity to biomechanical risk factors." *Ergonomics*, 49 vol 1.

Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*: Springer.

Dumas, J. S., and Redish, J. C. (1999). *A Practical Guide to Usability Testing*, Portland, OR, USA: Intellect Books.

Ebbinghaus, H. (1885). "Memory: A Contribution to Experimental Psychology". City.

Fagan, M. E. (1976). "Design and code inspections to reduce errors in program development." *IBM Systems Journal*, 15(3), 182-211.

Fitts, P. M., Jones, R. E., and Milton, J. L. (1950). "Eye movements of aircraft pilots during instrument-landing approaches." *Aeronautical Engineering Review*, 9(2), 24-29.

Fitzpatrick, R. (1999). "Strategies for Evaluating Software Usability." http://www.comp.dit.ie/rfitzpatrick/papers/chi99%20strategies.pdf. accessed on: March, 2009.

Folmer, E., Van Gurp, J., and Bosch, J. (2003). "A Framework for Capturing the Relationship between Usability and Software Architecture." *Software Process Improve and Practice*, 8, 67-87.

Fuhrmann, S., Komogortsev, O., and Tamir, D. (2009). "Investigating Hologram-based Route Planning." *Transactions of Geographical Information Science*, in press.

Goldberg, J., H., Stimson, M. J., Lewenstein, M., Scott, N., and Wichansky, A., M. "Eye tracking in web search tasks: Design implications." *Presented at Proceedings of the Eye Tracking Research and Application Symposium*.

Goldberg, J. H., and Kotval, X. P. (1999). "Computer interface evaluation using eye movements: methods and constructs." *International Journal of Industrial Ergonomics*, 24(6), 631-645.

Gries, D. (1987). *The Science of Programming*: Springer-Verlag New York, Inc.

Hax, A. C., and Majluf, N. S. (1982). "Competitive cost dynamics: the experience curve." *Interfaces*, 12(5), 50-61.

Hilbert, D. M., and Redmiles, D. F. (2000). "Extracting usability information from user interface events." *ACM Comput. Surv.*, 32(4), 384-421.

Hornbaeck, K. (2006). "Current practice in measuring usability: Challenges to usability studies and research." *Int. J. Hum.-Comput. Stud.*, 64(2), 79-102.

Howarth, J., Smith-Jackson, T., and Hartson, R. (2009). "Supporting novice usability practitioners with usability engineering tools." *Int. J. Hum.-Comput. Stud.*, 67(6), 533-549.

IEEE. (1990). "IEEE Std 610.12-1990 IEEE Standard Glossary of Software Engineering Terminology". City: Institute of Electrical and Electronic Engineers: New York, NY.

IEEE. (2002). "IEEE Std 730-2002 Standard for Softwae Quality Assurance Plans". City: IEEE: New York, NY.

IEEE. (2004). "IEEE Std 1012-2004 Standard for Verification and Validation". City: IEEE: New York, NY.

Ikehara, C. S., and Crosby, M. E. (2005). "Assessing Cognitive Load with Physiological Sensors"*Hawaii International Conference on System Sciences*. City, pp. 1-9.

ISO. (1998). "ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability". City: International Organization for Standardization: Geneva Switzerland.

ISO. (2001). "ISO/IEC 9126-1:2001 Software Engineering-Product Quality-Part 1: Quality Model". City: International Standards Organization: Geneva Switzerland.

ISO. (2003). *ISO/IEC 9126-2:2003 Software Engineering-Product Quality-Part 2: External Metrics*. International Standards Organization, Geneva Switzerland.

Jacob, R., and Karn, K. (2003). "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises.", in R. Radach, J. Hyona, and H. Deubel, (eds.), *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Elsevier Science.

John, B. E., and Kieras, D. E. (1996). "The GOMS family of user interface analysis techniques: comparison and contrast." *ACM Trans. Comput.-Hum. Interact.*, 3(4), 320-351.

Kahneman, D. (1973). *Attention and effort*, Englewood Cliffs, NJ: Prentice-Hall.

Kaner, C. (2003). "An Introduction to Scenario Testing." http://www.testingeducation.org/a/scenario2.pdf. accessed on: December, 2008.

Kit, E. (1995). *Software Testing in the Real World*, Reading, MA: Addison-Wesley.

Komogortsev, O., Mueller, C., Tamir, D., and Feldman, L. (2009). "An Effort Based Model of Software Usability"*2009 International Conference on Software Engineering Theory and Practice (SETP-09)*. City: Orlando, FL.

Komogortsev, O. V., and Khan, J. (2009). "Eye Movement Prediction by Oculomotor Plant Kalman Filter with Brainstem Control." *Journal of Control Theory and Applications*, 7(1).

Leveson, N., and Turner, C. S. (1993). "An Investigation of the Therac-25 Accident." *IEEE Computer*, 26 no. 7(7).

Marshall, S. P. "The Index of Cognitive Activity: measuring cognitive workload." *Presented at Human Factors and Power Plants, 2002. Proceedings of the 2002 IEEE 7th Conference on*.

McCall, J. A., Richards, P. K., and Walters, G. F. (1977). *Factors in Software Quality*. Nat'l Tech. Information Service, .

Mills, C., Bury, K. F., Roberts, T., Tognazzini, B., Wichansky, A., and Reed, P. (1986). "Usability testing in the real world"*Proceedings of the SIGCHI conference on Human factors in computing systems*. City: ACM: Boston, Massachusetts, United States, pp. 212-215.

Moore, P., and Fitz, C. (1993). "“Gestalt theory and instructional design, Journal of Technical Writing and Communication." *Journal of Technical Writing and Communication*

23(2), 137-157.

Myers, G. (1979). *The Art of Software Testing*, New York, NY: John Wiley & Sons.

Nielsen, J. (1993). *Usability Engineering*, San Francisco, CA, USA: Academic Press.

Nielsen, J. (2005). "Heuristic Evaluation." http://www.useit.com/papers/heuristic/. accessed on: 1 March 2009.

Nielsen, J. (2008). "Logging Actual Use." http://www.usabilityhome.com/FramedLi.htm?Logging.htm. accessed on: December, 2008.

Pressman, R. (2010). *Software Engineering: A Practitioner's Approach*, New York, NY.: McGraw-Hill.

Ritter, F. E., and Schooler, L. J. (2001). "The Learning Curve"*Internatonal Encyclopedia of Social & Behavioral Sciences*. City: Elsevier Science Ltd.

Royce, W. W. (1970). "Managing the development of large software systems: concepts and techniques"*Proceddings, IEEE WESCON*. City: IEEE Computer Society Press, pp. 1-9.

Rubin, J., and Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan , Design, and Conduct Effective Tests*, Indianapolis, IN, USA: Wiley Publishing, Inc.

Rumbaugh, J., Jacobson, I., and Booch, G. (1999). *The Unified Modeling Language Reference Manual*, Reading, MA: Addison Wesley Longman, Inc.

Seffah, A., Donyaee, M., Kline, R. B., and Padda, H. K. (2006a). "Usability measurement and metrics: A consolidated model." *Software Quality Journal*, 14(2), 159-178.

Seffah, A., Donyaee, M., Kline, R. B., and Padda, H. K. (2006b). "Usability measurement and metrics: A consolidated model." *Software Quality Control*, 14(2), 159-178.

Shneiderman, B. (1998). *Designing the user interface: strategies for effective human-computer-interaction*, Reading, MA: Addison Wesley Longman, Inc.

Skov, M. B., and Stage, J. (2005). "Supporting problem identification in usability evaluations"*Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*. City: Computer-Human Interaction Special Interest Group (CHISIG) of Australia: Canberra, Australia, pp. 1-9.

Solomon, G. (1983). "Television watching and mental effort: A social psychological view", in J. Bryant and D. Anderson, (eds.), *Children's understanding of television*. New York: Academic Press.

Sutherland, I. E. (1963). *Sketchpad: A man-machine graphical communication system*, Massachusetts Institute of Technology, Cambridge, MA.

Tamir, D., Komogortsev, O. V., and Mueller, C. J. (2008). "An Effort and Time Based Measure of Usability"*Proceedings of the 6th International Workshop on Software quality*. City: ACM: Leipzig, Germany.

Tullis, T., and Albert, B. (2008). *Measuring The User Experience: collecting, analyzing, and presenting usability metrics*, Burlington, MA: Morgan Kaufmann.

Vukelja, L., Müller, L., and Opwis, K. (2007). "Are Engineers Condemned to Design? A Survey on Software Engineering and UI Design in Switzerland"*NTERACT 2007*. City: Springer Rio de Janeiro, Brazil.

Winter, S., Wagner, S., and Deissenboeck, F. (2008). "A Comprehensive Model of Usability", *Engineering Interactive Systems: EIS 2007 Joint Working Conferences, EHCI 2007, DSV-IS 2007, HCSE 2007, Salamanca, Spain, March 22-24, 2007. Selected Papers*. Springer-Verlag, pp. 106-122.

Wright, T. P. (1936). "Factors Affecting the Cost of Airplanes." *Journal of Aeronautical Sciences*, 3(4), 122-128.