5-2010

The Effect of Interface Consistency and Cognitive Load on User Performance in an Information Search Task

Jeremy Mendel
Clemson University, mendel.clemson@gmail.com

THE EFFECT OF INTERFACE CONSISTENCY AND COGNITIVE LOAD ON USER PERFORMANCE IN AN INFORMATION SEARCH TASK

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Applied Psychology

---

by
Jeremy Mendel
May 2010

Accepted by:
Dr. Richard Pak, Committee Chair
Dr. Leo Gugerty
Dr. Jason Thatcher

ABSTRACT

Although interface consistency is theorized to increase performance and user satisfaction, previous research has found mixed and often non-significant results. The source of this discrepancy may be due to varying levels of task difficulty employed in these past studies. This study attempted to control the task difficulty using cognitive load theory. Interface consistency was manipulated along with intrinsic cognitive load and extraneous cognitive load. Interface consistency was manipulated along three dimensions: physical, communicational and conceptual. Intrinsic cognitive load was manipulated by asking participants finance (high load) questions and travel (low load) questions. Unnecessary and irrelevant extra hyperlinks were used to manipulate extraneous cognitive load. These hyperlinks were either present (high load) or absent (low load) in the websites. Forty eight participants searched for answers to 24 questions across four separate websites. Results indicated interactions between consistency and the two types of cognitive load. These interactions suggest that the effects of consistency are dependent upon the difficulty of the task. Specifically, consistency may be especially important for difficult tasks with high cognitive load.

ACKNOWLEDGEMENTS

I would first like to thank my advisor and committee chair, Dr. Pak; it was his guidance that made this project possible. I would also like to thank my other committee members, Dr. Gugerty and Dr. Thatcher, for helping me to improve both my writing and my skills as a researcher. I would like to thank Nicole and Margaux for their support and for being great officemates.

Finally, I would like to thank my Mother for everything she has done for me. Her support and encouragement helped me persevere through many difficult times.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

INTRODUCTION

Interface consistency can be described as the look and feel of an interface and is considered one of the core aspects of usability (Nielsen, 1989). Interface consistency is central to design because it allows users to generalize knowledge and transfer it to other aspects of a system or even to other systems (Bayer, 1992). In essence, interface consistency is the extent to which two interfaces or systems share a common look, layout, and functionality. Interface consistency research is as old as the field of human factors itself. Some of original studies in Human Factors examined the controls of World War II aircraft (Chapanis, 1953; Fitts & Jones, 1961). Aircraft controls of the era were seemingly designed without concern for consistency, with one researcher calling the controls “fiendishly inconsistent” (Chapanis, 1953). One study of the control design attributed over half of all pilot errors to inconsistent controls (Fitts & Jones, 1961).

The presence, or absence, of consistency may also affect user safety. In an analysis of civilian and military helicopter accidents over water, the research concluded that a 25% to 35% mortality rate involving underwater escape was due to inconsistent helicopter door and window jettison mechanism designs (Brooks & Bohemier, 1997). Placement of the release mechanisms varied from mid-chest level to behind the hip of the pilot. Assuming the user was able to find the mechanism, the latch designs were not standardized. Of the 35 helicopters studied, 23 different release mechanisms were employed. The inconsistent position and design of the jettison controls may have caused operator confusion during emergency, high-workload, high-stress situations (Brooks &

Bohemier, 1997). A consistent design (i.e. similar control location and function) of these mechanisms would allow pilots to operate these controls more “automatically”, or with less conscious effort (AlTaboli & Abou-Zeid, 2007; Proctor & Vu, 2006; Schneider, Dumais & Shiffrin, 1984; Schneider & Shiffrin, 1997). The problem of inconsistent interfaces is not restricted to these helicopters; inconsistencies permeate into many other designs.

The difficulty facing construction workers resulting from interface consistency mirrors the helicopter door latch problem. Many of the construction vehicles use a similar control layout but inconsistent control manipulations to operate. For example, an excavator and a skid steer vehicle both have two control sticks and pedals. Although the designs of these two vehicles appear similar, the operation of these machines is inconsistent. The excavator requires the operator to push the pedals to drive the vehicle, while the skid steer requires the operator to manipulate the control sticks to drive. While usability professionals tend to agree that consistency is important, how to best define it remains debated (Grudin, 1989; Nielsen, 1989; Shneiderman, 1998).

Dimensions of Consistency

In an effort to further clarify the concept of consistency, researchers have operationalized specific dimensions of user interaction with the system that contribute to consistency: physical, communicational and conceptual (Adamson & Wallace, 1997; AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2000; 2004; Rhee, Moon & Choe, 2006).

Physical consistency

Physical consistency considers the visual or graphical appearance of an interface or object (AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2000; Rhee et al., 2006). One example of physical consistency is the location of an automobile speedometer. Nearly all automobiles have an analog speedometer with a similar design right above the steering column. This standard is so prevalent that drivers are able to gauge speed in a different car without first studying the speedometer's design. In contrast, emergency brakes in vehicles are often inconsistent. Some vehicles use a hand-operated lever found in between the two front seats while other vehicles use a foot-operated pedal.

Communicational consistency

Communicational consistency is the level of consistency between the way the user interacts with the system and the way in which the system presents information to the user (e.g., Ozok & Salvendy, 2000). For example, pulling back on an airplane's throttle will always decrease the engine's output. An example of communicational inconsistency would be to make a button sometimes engage the air brakes and other times the same button would engage the autopilot.

Conceptual consistency

Conceptual consistency refers to how a user thinks about an interface and its match to how the system presents the interface. Conceptual consistency has been described as the consistency of the metaphor applied to the system and how it is represents components of an interface (AlTaboli & Abou-Zeid, 2007; Kellogg, 1987; Ozok & Salvendy, 2000; Rhee et al., 2006). An example of this is the menu bar found in most windows-based applications. This menu bar uses similar, if not identical headings across programs (e.g. File, Edit, etc) and each menu contains similar commands such as File>Save or Edit>Copy. Newer versions of Microsoft use an inconsistent design which relies on the tab-based system rather than the traditional Windows menu.

Incomplete consistency (only addressing some of the dimensions of consistency) can be detrimental to user performance (Finstad, 2003; Rhee et al, 2006; Satzinger & Olfman, 1998). Returning to the example of the controls of the excavator versus the skid steer, these machines illustrate incomplete consistency. While the interfaces of the two machines are physically consistent, the interfaces are not communicationally consistent. These discrepancies can be detrimental in high-stress conditions similar to the post-World War II aircraft studies (Fitts & Jones, 1961).

Review of Interface Consistency and Human Performance

Consistency between two interfaces may encourage learned skills to be transferred to new systems. The presence of consistency may also help the user predict system responses (AlTaboli & Abou-Zeid, 2007; Nielsen, 1989; Rhee et al., 2006). Consistency can also contribute to the development of expertise through automatic attention responses (Schneider & Shiffrin, 1997). The results of interface consistency can be seen in a shortened learning process, reduced working memory demand and increased efficiency (Bayer, 1992; Nielsen, 1989; Proctor & Vu, 2006). The theorized benefits of consistency are shorter task completion time, reduced error-rate, and higher user satisfaction (e.g., Rhee et al., 2006).

Although interface consistency is theoretically beneficial, empirical results of consistency are unclear. In testing, interface consistency studies found positive effects (AlTaboli & Abou-Zeid, 2007), non-significant effects (e.g. Rhee et al., 2006), and even detrimental effects (Finstad, 2003). In addition to questionable performance benefits of interface consistency, the concept of interface consistency has been criticized as too vague when specifying what makes an interface consistent (Grudin, 1989; 1992). Consistency, or inconsistency, is ultimately based on individual opinion making it difficult to objectively achieve consistency. Supporting Grudin, empirical evidence has shown that some consistent interfaces can cause users to over-generalize functions within the interface (Finstad, 2003). Users interacted with different iterations of a web browser in this study. Participants searched for information, changed advanced browser settings, and saved webpages as HTML files. In this case, interface consistency was detrimental to performance (longer completion time and more errors). Finstad argued that the source of the errors was that subjects over-generalized prior knowledge to the new interface. However, in the study, some of the "consistent" interfaces actually demonstrated incomplete consistency. Incomplete consistency may have led users to incorrectly perceive the interface as consistent, therefore making inappropriate generalizations to the new interface.

Another study examining the effect of consistency on computer-based applications found mixed results (Satzinger & Olfman, 1998). The study manipulated what was referred to as the "action language syntax" and "visual consistency" of the interface. "Action language syntax" refers to the consistency of the commands labels (e.g. F1 = help). This manipulation had a positive effect on performance. However "visual consistency", which is analogous to the physical consistency dimension, negatively affected performance. Satzinger & Olfman concluded that task variety induced through visual inconsistencies improved performance by helping users distinguish between the two interfaces.

Other studies have found no significant effect for overall completion time, error-rate or satisfaction. One such study used simple web-navigation tasks (Rhee et al., 2006). Participant's tasks included clicking, data entry, reading comprehension, and word searches. Another study using similar tasks found a significantly beneficial effect of consistency on error-rate, but failed to see any effect on completion time or satisfaction (Ozok & Salvendy, 2000).

The literature reviewed thus far suggests a conflicted view of consistency. Some studies show positive effects of consistency (AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2003), some show negative effects of consistency (e.g. Finstad, 2003), and some studies show mixed results or none at all (e.g. Rhee et al., 2006). A closer look at the methodology employed in these studies could help explain the contradictory findings.

The manipulation of consistency is a methodological consideration that has varied widely between past studies. One study examined only the effect of physical consistency and found a positive effect (AlTaboli & Abou-Zeid, 2007). This study manipulated the colors, fonts and locations of items on a website. Participants were asked to perform general web-based tasks like point-and-click, reading comprehension and form filling. In this case, participants in the consistent condition presumably performed tasks with fewer errors and reported higher levels of satisfaction as a result of the consistency. Another study examined all three types of consistency across different iterations of an online e-learning website but did not find any significant results of consistency (Rhee et al., 2006). A methodological issue in this particular study may have been that each dimension of consistency was manipulated individually. For example, one iteration of the system was physically inconsistent while still communicationally and conceptually consistent. A different iteration of the system was just conceptually inconsistent.

Another source of these discrepancies may be due to the nature of the tasks used, specifically, the lack of control over task difficulty. If one study used a harder task than another, this could help explain the contrary findings. One study that did not find a significant effect on performance required participants to perform routine internet tasks like form-filling and information searching (Rhee et al., 2006). In contrast, another study that did find significant differences had participants perform similar web-based tasks but also included more advanced tasks like enabling JavaScript (Finstad, 2003). The difficulty of a task can be quantified in many ways. For example, how many steps are required, the type and amount of cognitive processing required, or the level of demands placed on working memory. One method used to measure task difficulty is based on cognitive load theory.

Cognitive Load

Cognitive load can be defined as the burden placed on working memory during problem solving and learning (Ayres, 2006). In the current context, it could also be used to characterize any task's demands on limited resources such as working memory.

Cognitive load theory describes the total cognitive load of a task in two core parts: intrinsic and extraneous cognitive loads. Intrinsic load is the difficulty of the task materials. Extraneous cognitive load is the added and unnecessary difficulty induced by the method of presentation (Ayres, 2006; Bannert, 2002; Paas, Tuovinen, Tabbers, & Van Gerven, 2003).

Intrinsic cognitive load

Intrinsic load deals with the cognitive demands or the complexity of the material to be learned (Chandler & Sweller, 1996; Sweller, Van Merrienboer, & Paas, 1998). For example, learning calculus is difficult because of the inherent complexity of the material. To make it more manageable, pieces of what is needed to learn calculus are taught beginning in elementary school in the form of basic arithmetic. This serial learning process (a form of part-task training), where steps are mastered individually, is one technique used to reduce the intrinsic load of a task (Chander & Sweller, 1996).

Another source of intrinsic cognitive load is the amount of element interactivity present in the material (Bannert, 2002; Sweller & Chandler, 1994; 1996). Returning to the example of calculus, it is difficult because the learner must combine so many previously learned procedures ranging from basic arithmetic to order of operation rules. Since the material itself is so complex, it is crucial that the presentation of the material be efficiently designed to avoid further taxing the individual's limited working memory. This aspect of the task, the manner in which the material is presented, is known as extraneous cognitive load.

Extraneous cognitive load

Extraneous load is the added difficulty presented by the method in which the material is presented (Ayres, 2006; Bannert, 2002; Paas et al., 2003). Extraneous load can be detrimental to learning and performance via the "split-attention effect," which occurs when material requires an individual to deal with multiple, disparate sources of information (Gerven, Paas, & Schmidt, 2000; Sweller & Chandler, 1991). Returning again to the calculus example, when a student must use a textbook chapter to find the correct procedure, the back of the textbook to find the correct formula, then an entirely separate workbook to work out the problem, the student must shift attention between three different places. In contrast, if the procedure, formula, and workspace were in closer proximity, it would reduce the extraneous cognitive load.

Another mechanism for the presumed detrimental effects of extraneous load is the redundancy effect, which is when the user/learner must process material that is redundant (Gerven et al., 2000; Sweller & Chandler, 1991). In the calculus example, this would happen if a diagram showing the process of solving simple arithmetic also included redundant written step-by-step instructions.

Some research suggests that the effects of intrinsic cognitive load and extraneous cognitive load loads are interactive in nature; an increase in one makes an individual more sensitive to increases in the other (Schnotz & Kürschner, 2007). Sweller & Chandler (1994) demonstrated this interaction between the two loads by showing that extraneous cognitive load, specifically the split-attention and redundancy effects, were significantly more detrimental in tasks with high intrinsic cognitive load.

Cognitive load of the task as an explanation for conflicting interface consistency results

Previous interface consistency research has not manipulated or considered the difficulty of the tasks used in the studies. Although many studies employed web-based tasks, these tasks varied in the knowledge required. In one study, participants manipulated advanced settings in the browser like turning on JavaScript or viewing the source code of a page (Finstad, 2003) which could be considered a relatively difficult/advanced task. Other studies required participants to perform relatively simple tasks like clicking, data entry, reading comprehension and word searches (e.g. Rhee et al., 2006). These between-study variations in task difficulty make it hard to draw general conclusions about interface consistency effects.

Cognitive Load Theory would predict that poor interface design (more specifically interface inconsistency) would increase the extraneous cognitive load of the user. If interface consistency is one aspect that makes up the extraneous load, then the total cognitive load imposed by the task would moderate the effect of interface consistency (Chandler & Sweller, 1991; Sweller & Chandler, 1994). In the 1994 study, it was shown through multiple experiments that for tasks with a low intrinsic cognitive load, participants were not affected by increased levels of extraneous cognitive load. The explanation for this finding was that "easier" tasks required less working memory thus leaving more cognitive resources to deal with extraneous load before the participant was overloaded.

The link between interface consistency and cognitive load is not entirely novel. Researchers have previously theorized that the amount of load imposed by an interface is affected by the design of an interface (Chalmers, 2003; Saadé & Otrakji, 2007; Szabo & Kanuka, 1998). If this is the case, the varied amount of cognitive load in these tasks may have caused the conflicting results of past studies. Saadé and Otrakji (2007) found a correlation between screen design and cognitive load (as measured by a questionnaire). The correlation suggested that "good screen design", which can include consistency, was associated with reduced subjective cognitive load. If cognitive load is affected by screen design, specifically interface consistency, then it should be controlled to understand the affect of consistency.

To summarize, past interface consistency research, such as Rhee et al. (2006) and Finstad (2003), have ignored the cognitive load of the participant's tasks; furthermore, they were relatively simple tasks. The tasks may have had such a low level of intrinsic cognitive load that participants were easily able to deal with the additional load imposed by the inconsistent interfaces thus showing no effects of interface consistency on performance. Without controlling for varying levels of cognitive load imposed by the tasks in these studies, it is unclear if the tasks were difficult enough to produce an effect. Perhaps by controlling the difficulty of the task, the effect of interface consistency can be better understood.

Current Study

The current study manipulated the level of cognitive load in the task as well as interface consistency. The rationale was that when the level of cognitive load was higher, the positive effects of interface consistency would become apparent as described by Nielsen (1989). When the level of load is low, there would be no significant effect of interface consistency (e.g. Rhee et al., 2006). Interface consistency was expected to interact with cognitive load by reducing the memory load. The goal was to determine under what conditions is consistency beneficial. Specifically, does task difficulty (cognitive load) influence the affect of consistency?

For the present study, participants answered 24 questions across four separate websites. Two websites were designed as consistent with each other and two other websites were inconsistent with each other, manipulating all three dimensions of consistency (physical, communicational, conceptual). Consistency was manipulated between-group meaning that participants either used four consistently designed websites or switched between the different designs. The extraneous cognitive load of the task was manipulated by designing separate websites with extra hyperlinks interspersed throughout the website. These links were either present (high load) or absent (low load) from the body of the webpage. Extraneous cognitive load was also manipulated between-group meaning that the links were either always present or always absent across all the websites a participant used. Finally, the intrinsic cognitive load of the questions was also manipulated. This manipulation was accomplished by asking participants questions from two different topic domains, finance and travel. The “harder”, high intrinsic load questions, involved finance information, while the “easier” low intrinsic load questions used travel information. Intrinsic cognitive load was manipulated within-group meaning that all participants answered half travel questions and half finance questions.

Hypotheses

Interaction

An interaction between the level of interface consistency and the amount of cognitive load imposed by the task was hypothesized (see Figure 1). It was predicted that high levels of intrinsic and extraneous cognitive load (e.g. finance questions with the extra links), would result in a significantly larger reduction in performance due to the inconsistent interface designs when compared to the lower cognitive load conditions. These performance declines were expected to be demonstrated in longer task completion times, more errors, more pages visited during questions and longer average times spent on each page. Under low load conditions, interface consistency was not expected to affect task performance. We expected this pattern because the increased cognitive load leaves participants more vulnerable to other increases in difficulty (e.g. inconsistently designed interfaces). Subjective ease-of-use scores were predicted to reveal a similar interaction between consistency and extraneous load. In this case, individuals would be more likely to report unfavorable scores in the inconsistent and high extraneous load condition. An interaction with the intrinsic load was not measured for the ease-of-use scores since intrinsic load was manipulated within-group.

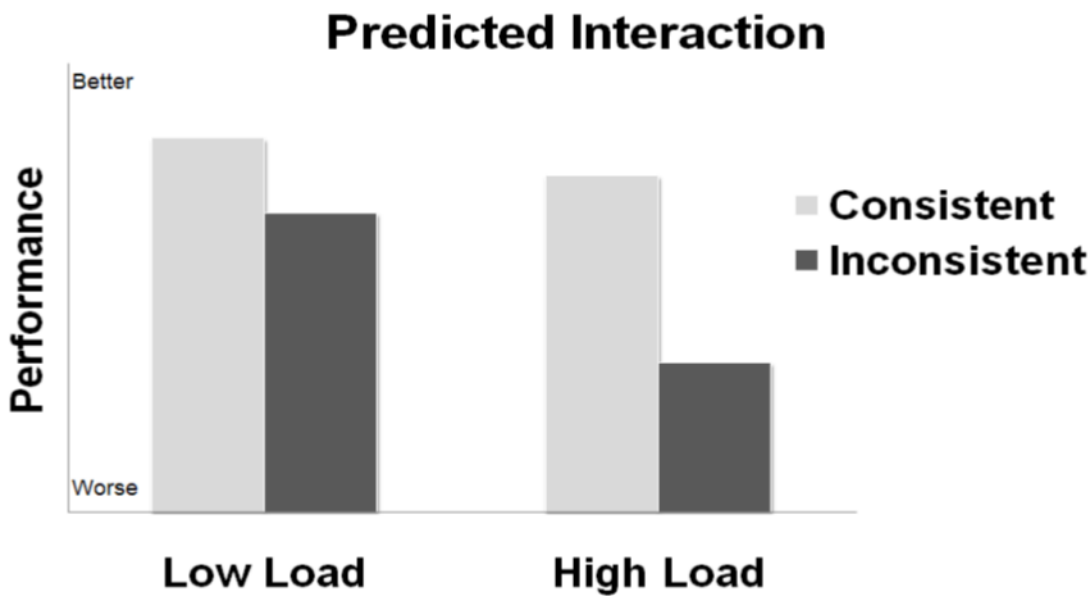

Figure 1. Predicted Interaction between interface consistency and task cognitive load.

Main effects

Consistency was expected to generally improve user performance in line with most previous literature. Specifically, consistency was expected to reduce completion time, errors, pages visited, and time spent on each page. Similarly, low intrinsic cognitive load (travel questions) and low extraneous cognitive load (no extra hyperlinks) were expected to improve performance when compared to the higher load conditions. Regarding the subjective user satisfaction, participants in the consistent and low extraneous load conditions were expected to report better ease-of-use. Because intrinsic cognitive load was manipulated within participants, its influence on the ease-of-use survey was not assessed.

METHODS

Participants

A total of 48 undergraduate students were recruited at Clemson University through the online participant recruitment system. Participants received class credit in exchange for participation. Participants with an error-rate over two standard deviations from the condition mean were removed. One participant from each of the four conditions exceeded the value and was removed. From the original 48 participants, 44 were included in the analyses.

Table 1
Participant demographic frequencies by condition

Condition	Consistent; Low Extraneous Load	Inconsistent; Low Extraneous Load	Consistent; High Extraneous Load	Inconsistent; High Extraneous Load
Mean Age (SD)	18.8 (*SD* = 1.4)	18.5 (*SD* = 0.9)	19.0 (*SD* = 1.5)	18.2 (*SD* = 0.6)
Male	5	6	1	4
Female	6	5	10	7
H.S. Education	10	8	6	10
Some College	1	3	5	1

Note: Pearson Chi-Squared showed no significant differences between groups.

Of the participants, 16 were male and the average age was 18.6 ($SD = 1.2$). All participants reported at least three years of experience using computers. Participant demographics and experience with computers is summarized in Table 1 and Table 2. To make certain that all four conditions did not vary significantly demographically or in amount of computer experience, chi-squared tests were conducted. Participants in the four conditions did not differ in terms of age (χ^2 (15, $N = 44$) = 15.7, $p > .05$), sex (χ^2 (3, $N = 44$) = 5.5, $p > .05$) and education (χ^2 (3, $N = 44$) = 5.7, $p > .05$). These four conditions also did not differ significantly in total experience with computers (χ^2 (3, $N =$

44) = 3.9, $p > .05$), highest three-month frequency of computer-use (χ^2 (9, N = 44) = 6.1, $p > .05$) and current three-month computer-use (χ^2 (9, N = 44) = 6.3, $p > .05$).

Table 2

Participant computer experience frequencies by condition

Condition	Consistent; Low Extraneous Load	Inconsistent; Low Extraneous Load	Consistent; High Extraneous Load	Inconsistent; High Extraneous Load
Total experience with computers				
3-5 years total	1	2	0	0
> 5 years total	10	9	11	11
Highest frequency of computer use ever				
Several days per week	1	1	0	0
Daily, infrequently	2	2	3	2
Daily, frequently	8	7	8	7
Daily, most of the day	0	1	0	2
Highest frequency of computer use in the last three months				
1-5 hours a week	0	1	0	1
5-10 hours a week	4	2	3	5
10-15 hours a week	4	4	3	4
> 15 hours a week	3	4	5	1

Note: Pearson Chi-Squared showed no significant differences between groups.

Task

Participants found the answer to a series of 24 questions, half travel and half finance related. To answer these questions, participants navigated through four separate websites with six questions answered on each website. The condition they were assigned to determined which version of the websites they used (see Figure 2). To answer a question, participants "purchased" the item through the website. Instant feedback was given if the question was correct or incorrect; participants did not move to the next question until the current question was correctly answered. The task continued until all 24 questions were answered correctly. Participants took an average of 1105 seconds (SD = 271) to complete all 24 questions.

Design

The study used a 2 (consistency, high/low) x 2 (extraneous load, high/low) x 2 (intrinsic load, high/low) mixed factorial design. Participants were randomly assigned to one of four possible conditions (see Figure 2). To begin, participants completed a demographics and computer experience form. Next, they were given a series of 24 questions to answer on four separate websites. Website presentation order was counter-balanced across participants to control for order effects.

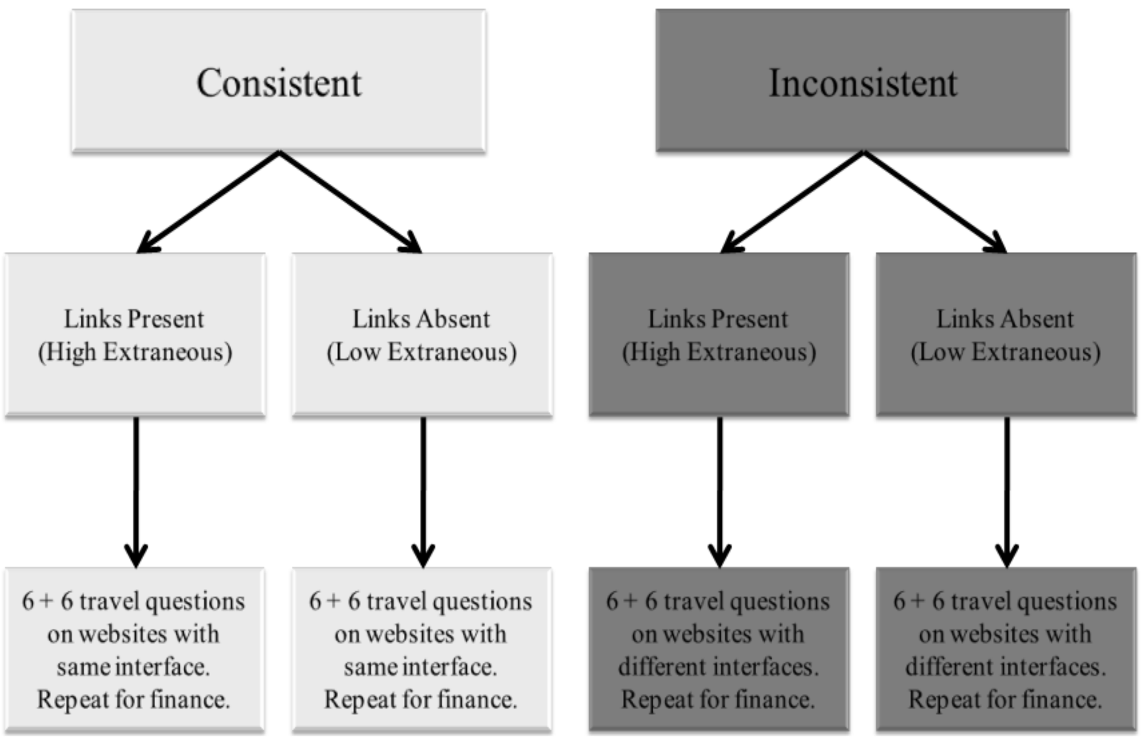

Figure 2. Four possible participant conditions.

Independent variables

The independent variables were interface consistency, extraneous cognitive load, and intrinsic cognitive load of the task. Interface consistency was manipulated between participants. Participants in the consistent condition used four websites with the same interface; participants in the inconsistent condition used four websites with two inconsistent designs. To control interface consistency, a series of inconsistencies was used to alter the original design. These manipulations and justification for using them are seen in Table 3. Screenshots taken from the two designs can be seen in appendix G. The manipulation of the website consistencies was checked immediately after the task by asking each participant how different he or she thought the websites were. An independent samples t-test was used to compare responses from participants in the two conditions and how they perceived the consistency of the websites. This test indicated that participants from the inconsistent conditions reported the websites used as being significantly more different from each other than the participants from the consistent conditions $t(42) = 6.7$, $p \leq .05$, $r^2 = .51$.

Table 3
Example Manipulations of Consistency for Websites

Dimension	Difference	Interface 1	Interface 2	Citation
Physical	Location of navigation Bar	Top, horizontal	Right, vertical	AlTaboli & Abou-Zeid, 2007; Ozok & Salvendy, 2004
Physical	Text spacing	Single Spaced	Double spaced	Bednall (1992); Benbasat and Todd (1993); Ozok & Salvendy, 2004
Comm.	Menu systems	Text hyperlinks for navigation	Image hyperlinks for navigation	Adamson & Wallace, 1997; Ozok & Salvendy, 2004
Comm.	Scroll bars	No scrolling needed	Must scroll to see additional info	
Conceptual	Replacing words with icons	“Four star” hotel written in text	Four stars in an icon form instead of text	Satzinger & Olfman, 1998
Conceptual	Alphabetized list sorting	Categorically sorted	Randomly sorted	Ozok & Salvendy, 2004

Extraneous cognitive load was manipulated by including extraneous hyperlinks within the website for the high cognitive load condition. This design is similar to what is seen on the website Wikipedia where extra, tangentially, related links are scattered throughout the text. These extra links were manipulated between participants. Extra hyperlinks were either present or absent for a participant across all four websites. The theory for this manipulation is that it forces an individual to make a judgment regarding the hyperlink; simply having the extra links requires more decisions to be made (DeStefano & LeFevre, 2007; see Table 4 below). The manipulation is considered an extraneous load since it involves the manner that material is presented.

Table 4
Example of Extraneous Hyperlinks

With Extraneous Hyperlinks	Without Extraneous Hyperlinks
Clemson University is located in upstate South Carolina in Pickens County just north of Interstate 85 and Anderson, South Carolina, along the shores of Lake Hartwell. The University is located just outside of the greater Greenville area and is approximately two hours away from Atlanta, Georgia, Charlotte, North Carolina and Columbia.	Clemson University is located in upstate South Carolina in Pickens County just north of Interstate 85 and Anderson, South Carolina, along the shores of Lake Hartwell. The University is located just outside of the greater Greenville area and is approximately two hours away from Atlanta, Georgia, Charlotte, North Carolina and Columbia.

The intrinsic cognitive load was manipulated within participants by asking all participants both harder questions and easier questions. The harder questions dealt with questions about financial investment. These questions were expected to be more difficult for the participants since the topic is unfamiliar to a typical college undergraduate. The easier questions used travel-based information which participants would likely be more familiar with.

All participants were asked the same 12 finance (high load) questions and the same 12 travel (low load) questions. An example finance question used in the study was "Find the cheapest municipal bond with a yield of at least 1%". An example travel question was "Find the cheapest flight to Los Angeles with free in-flight snacks". Questions were balanced between the two conditions to require the same number of steps to avoid a confounding the manipulation.

The question type was an intrinsic load manipulation since it changes the difficulty of the task itself. The effectiveness of the manipulation was checked immediately after the task by asking each participant which set of questions he or she perceived as more difficult. Regarding the difficulty of the question types, 64% of participants reported the finance questions as being more difficult while only 6% reported that the travel questions were more difficult. The remaining 30% reported that neither type of question was more difficult. A Pearson Chi-Squared test showed that these values significantly differed χ^2 $(6, N = 44) = 23.9$, $p \leq .05$, $r^2 = .54$.

Dependent variables

The dependent variables were task completion time, error-rate, total number of pages visited and the average time spent on each page. These performance measures were calculated separately for the two groups of questions (finance and travel) that participants answered. Additionally, the subjective user satisfaction was collected following the information search task. The metrics used in the study were quantified as follows:

- Completion time: Measured for each group of questions from the beginning of the first question through the final question.
- Errors: Every time a participant answered question incorrectly. Since the participants continue searching until the correct answer is found, a participant can accumulate multiple errors per question.
- Total pages visited: Measured by counting the total number of pages that a participant navigated through while answering a group of questions.
- Average time spent per page: The mean time a participant spent on each page during that set of questions.
- Subjective user satisfaction: Measured on a seven-point Likert scale across 18 questions. Scores were averaged to form a single average ease-of-use score. Questionnaire was adapted from IBM's Computer Usability Satisfaction Questionnaire (Lewis, 1993).

Materials/Apparatus

Five identical workstations with 17" LCD monitors running Windows XP were used in the study. The individual workstations each ran an identical copy of the websites off of the local hard drive. A total of 16 websites were designed using Adobe Dreamweaver CS3 to represent each of the possible conditions. The 16 different designs can be seen in Table 5. The websites were displayed using Mozilla Firefox version 3.5.3. The status bar in Firefox was disabled to avoid signaling which option was correct via the URL (a correct answer would lead to a website with the word correct in the URL).

TechSmith's Morae Recorder version 3.1 was used to administer the questionnaires and to record user performance data.

Participant information was collected using a basic demographics form along with a computer experience questionnaire (see Appendix B). A short survey adapted from the IBM Computer Usability Satisfaction Questionnaires was also administered following the information search tasks to gather user satisfaction data (Lewis, 1995; Appendix D). An example of a questionnaire window is seen in Figure 3.

Table 5

List of Website Versions

Extra Hyperlinks	No Extra Hyperlinks
Travel Interface 1, Question Set 1	Travel Interface 1, Question Set 1
Travel Interface 2, Question Set 1	Travel Interface 2, Question Set 1
Finance Interface 1, Question Set 1	Finance Interface 1, Question Set 1
Finance Interface 2, Question Set 1	Finance Interface 2, Question Set 1
Travel Interface 1, Question Set 2	Travel Interface 1, Question Set 2
Travel Interface 2, Question Set 2	Travel Interface 2, Question Set 2
Finance Interface 1, Question Set 2	Finance Interface 1, Question Set 2
Finance Interface 2, Question Set 2	Finance Interface 2, Question Set 2

Note: The body content remained the same within each question type (finance and travel). For example, all body content was identical across all travel website variations. Screenshots from the two interface designs can be seen in Appendix G.

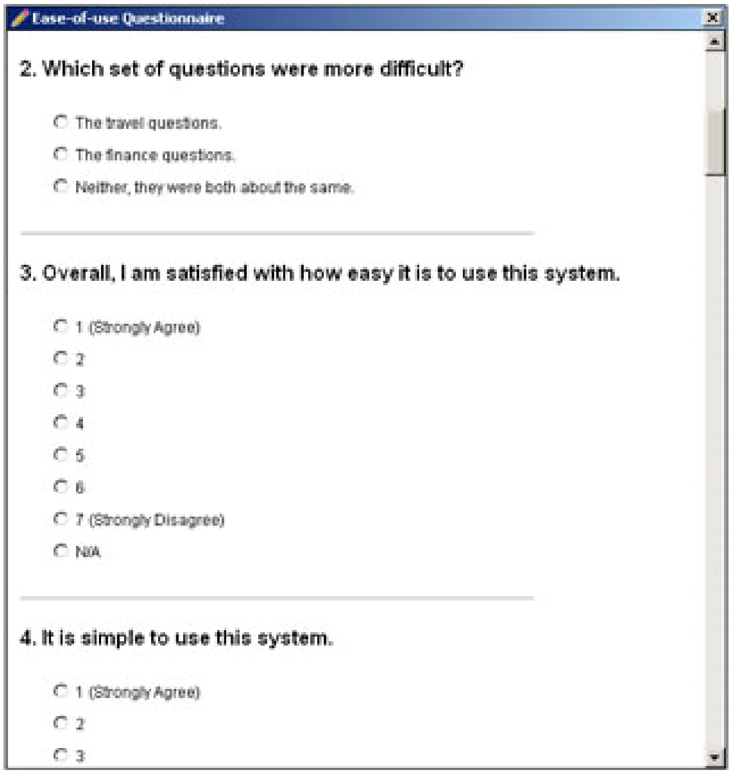

Figure 3. Example image of the ease-of-use questionnaire. This questionnaire was displayed immediately following the information search task.

Procedure

Participants were assigned to one of four conditions prior to arrival (see Figure 2 above). The experimenter began by reading a general overview of the study from the protocol script (Appendix A). Participants were then given a handout containing instructions and the list of questions (Appendix C). Upon receiving the handout, participants were told to read it over until the task was clear. Each individual participant indicated when he or she was ready; the experimenter began the Morae Recorder script. The demographics and computer experience questions immediately opened in a window (Appendix B). Once a participant finished with the demographics and computer experience questions, he or she navigated through the website to answer all 24 questions.

To prevent possible order effects, the order of questions was counterbalanced based on the participant's assigned group. As detailed above in Figure 2, there were four conditions; however, for counter-balancing purposes, there were in essence eight groups. These eight groups are described in Table 6. Once all 24 questions were answered, the participant raised his or her hand and the experimenter ended the recording. When the recording was ended, the manipulation check questions along with the ease-of-use questionnaire automatically opened in a new window (Appendix D). Participants completed the electronic questionnaire using a seven-point Likert scale. Once the ease-of-use questionnaire was complete, participants left the room and collected a copy of the debriefing form on the way out (Appendix E). After all participants finished, the Morae recordings were saved using a filename indicating the participant number and condition.

Table 6
Eight participant groups across four conditions

Group Number	Condition	Question Order	Interface(s) Used
1	Consistent; No added hyperlinks	Travel questions first	Only Interface 1
2	Consistent; No added hyperlinks	Finance questions first	Only Interface 2
3	Inconsistent; No added hyperlinks	Travel questions first	Interface 1 then 2
4	Inconsistent; No added hyperlinks	Finance questions first	Interface 2 then 1
5	Consistent; Added hyperlinks	Travel questions first	Only Interface 1
6	Consistent; Added hyperlinks	Finance questions first	Only Interface 2
7	Inconsistent; Added hyperlinks	Travel questions first	Interface 1 then 2
8	Inconsistent; Added hyperlinks	Finance questions first	Interface 2 then 1

Note: See Appendix G for screenshots of the two interface versions.

Data Reduction

Raw data recordings included the entire time a participant spent browsing the websites during the information search task. The recordings included data both from answering the questions and the transitions between questions. These transitions were an issue since they were not directly relevant to the task. Instead, these transitions measured the time a participant spent reading a question along with any breaks a participant may have taken between questions. Transitions were defined as the time from when a participant found a correct answer until they began searching for the next answer. The data from transition periods were removed and not included in the analyses.

RESULTS

A total of six dependant variables were used to measure task performance and usability: completion time, errors, a composite of time and errors, pages visited, average time per page, and ease-of-use. Descriptive statistics for performance are summarized in Tables 7 and 8. Descriptive statistics for the ease-of-use questionnaire are in Table 9.

Completion Time

To assess condition differences on task completion time, a 2 (consistency or inconsistency) × 2 (high or low intrinsic load) × (high or low extraneous load) repeated measures ANOVA was used (consistency and extraneous load were between-groups). The main effects of consistency and extraneous cognitive load were not significant (*p* = .13 and *p* = .12, respectively). These results indicated that the manipulations of consistency and extraneous cognitive load did not affect a participants completion time. The main effect of intrinsic cognitive load was significant, $F(1,40) = 7.6$, $p \leq .05$, $\eta^2 = .16$, which meant that participants generally took longer to answer the financial (high intrinsic load) questions (*M* = 495.2 seconds, *SD* = 192.0) compared to the travel (low intrinsic load) questions (*M* = 408.0 seconds, *SD* = 150). None of the interactions were

Table 7
Descriptive Statistics for Consistent Conditions

	Low Extraneous Load (no added links)				High Extraneous Load (added links)			
	Low Intrinsic Load (Travel)		High Intrinsic Load (Finance)		Low Intrinsic Load (Travel)		High Intrinsic Load (Finance)	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time (seconds)	352.6	123.1	444.5	124.4	364.9	136.9	519.6	301.4
Errors	1.8	2.2	1.5	1.2	0.7	0.9	3.3	4.2
Page Visits	82.5	28.2	119.5	24.3	73.6	16.2	110.5	13.9
Average Time per Page	4.4	0.8	3.7	0.9	5.0	1.5	4.7	2.6

significant for task completion time ($p > .05$).

Errors

To assess condition differences on number of errors, a 2 (consistency or inconsistency) × 2 (high or low intrinsic load) × (high or low extraneous load) repeated measures ANOVA was used (consistency and extraneous load were between-groups). The three-way interaction of consistency × intrinsic load × extraneous load was significant, $F(1,40) = 4.9$, $p \leq .05$, $\eta^2 = .11$, so the two-way interactions and main effects for errors will not be described.

One source of the three-way interaction was a significant two-way interaction of intrinsic load × consistency, but only in the high extraneous cognitive load condition, $F(1,40) = 6.4$, $p \leq .05$, $\eta^2 = .14$ (Figure 4). Participants answering low intrinsic load questions under high extraneous load using inconsistent interfaces had significantly more errors ($M = 2.8$, $SD = 4.1$) than participants answering the same questions experiencing high extraneous load while using consistent interfaces ($M = 0.7$, $SD = 0.9$). That is, when participants were answering travel questions (low intrinsic load) between consistent websites, they had significantly fewer errors; however, this effect was only observed when the extra hyperlinks were present (high extraneous load).

Table 8
Descriptive Statistics for Inconsistent Conditions

	Low Extraneous Load (no added links)				High Extraneous Load (added links)			
	Low Intrinsic Load (Travel)		High Intrinsic Load (Finance)		Low Intrinsic Load (Travel)		High Intrinsic Load (Finance)	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time (seconds)	429.5	148.6	449.6	94.3	484.8	168.4	566.8	178.7
Errors	1.1	1.3	2.1	3.2	2.8	4.1	2.3	1.8
Page Visits	84.6	24.9	120.5	18.9	98.1	29.4	129.4	19.2
Average Time per Page	5.2	1.4	3.7	0.6	4.9	0.8	4.4	1.1

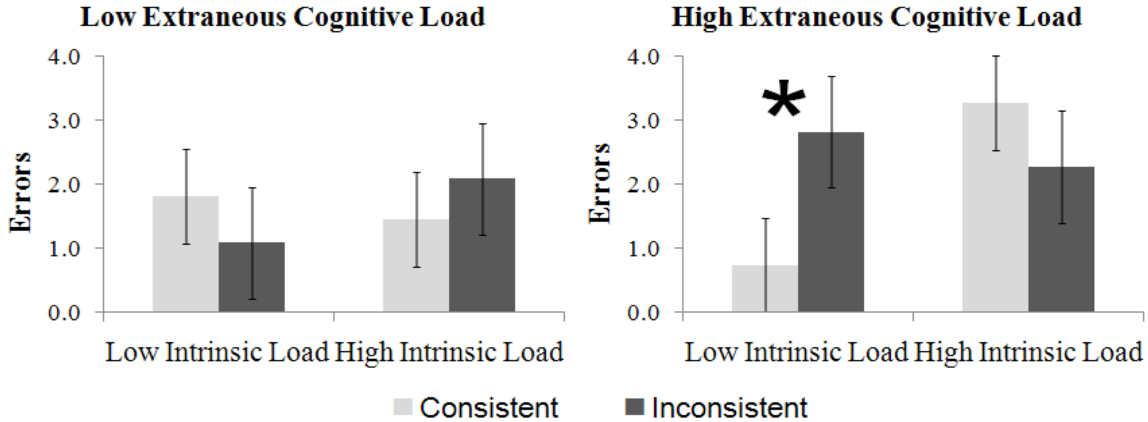

Figure 4. Interactions of intrinsic load and consistency split by extraneous cognitive load. Significant differences are marked with an asterisk.

Another source of the three-way interaction was a 2-way interaction of intrinsic cognitive load × extraneous cognitive load, but only for the consistent interface condition, $F(1,40) = 4.0$, $p \leq .05$, $\eta^2 = .09$ (see Figure 5). Participants in the condition with the extra hyperlinks present (high extraneous load), had significantly more errors when answering the financial questions (M = 3.3, SD = 4.2) when compared to the travel questions (M = 1.5, SD = 1.2). This interaction was only present for participants in the consistent condition.

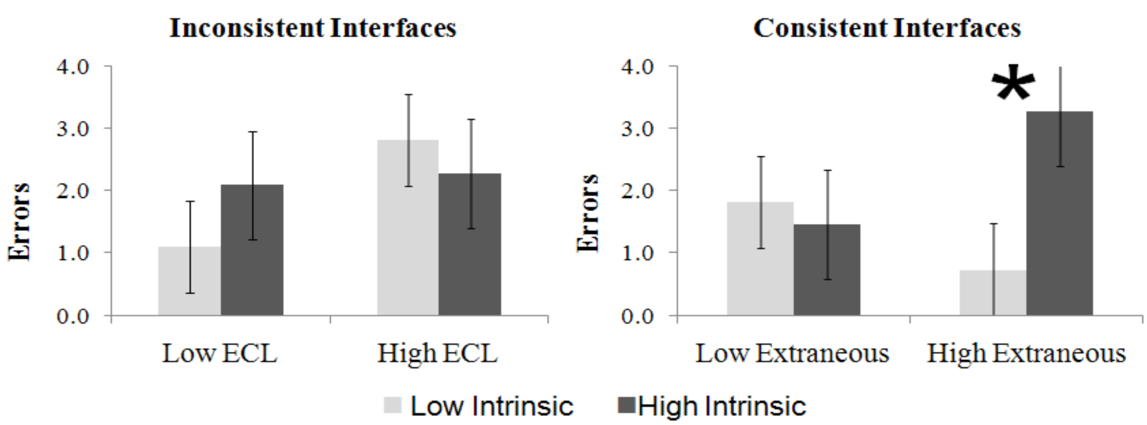

Figure 5. Interactions of intrinsic load and extraneous load split by consistency. Significant differences are marked with an asterisk.

Composite Performance Variable Analysis

A composite variable incorporating time and errors was created to understand overall participant performance (e.g. Pak & Price, 2008). This composite variable was preferable to the regular measures since it favored balanced performance rather than just speed or just accuracy. For each level of intrinsic load (finance and travel), time and errors were converted to standardized, unit-less, z-scores across all four conditions. The standardized values for time and errors were then averaged into a single composite measure for each of the two question domains, one for finance and one for travel. As with the individual measures, lower values on the composite measure indicated better performance.

A 2 (consistency or inconsistency) × 2 (high or low intrinsic load) × (high or low extraneous load) repeated measures ANOVA was used to analyze the condition differences on the new composite score (consistency and extraneous load were between-groups). There were no significant main effects for consistency (p = .18), intrinsic cognitive load (p = .99), or extraneous cognitive load (p = .11). None of the interactions were significant ($p > .05$).

Number of Pages Visited

To assess condition differences on the number of pages participants visited, a 2 (consistency or inconsistency) × 2 (high or low intrinsic load) × (high or low extraneous load) repeated measures ANOVA was used (consistency and extraneous load were between-groups). The main effect for consistency was significant, $F(1,40) = 4.4$, $p \leq .05$, $\eta^2 = .10$. The main effect of consistency suggested that participants browsed significantly more pages when using the inconsistently designed websites (*M* = 216.3 pages, *SD* = 23.3) compared to the participants using consistently designed websites (*M* = 193.1 pages, *SD* = 21.4). There was also a main effect for intrinsic cognitive load, $F(1,40) = 80.2$, $p \leq .05$, $\eta^2 = .67$. Participants visited significantly more pages when answering the financial questions (high intrinsic load; *M* = 120.0, *SD* = 19.9) when compared to the travel questions (low intrinsic load; *M* = 84.7, *SD* = 25.9). The main effect of extraneous cognitive load was not significant ($p = .85$) indicating that extraneous cognitive load did not affect the number of pages a participant visited.

The two-way interaction of consistency × extraneous load, although not significant, was examined closer for any potential simple main effects, $F(1,40) = 3.3$, $p = .075$, $\eta^2 = .08$ (see Figure 6). There was no significant effect of consistency for participants in the low extraneous load (no extra hyperlinks; $p = .86$). In contrast, there was a significant effect of consistency for the high extraneous load conditions (extra hyperlinks present), $F(1,40) = 9.5$, $p \leq .05$, $\eta^2 = .32$. For participants in the high extraneous load condition (extra hyperlinks), those in the consistent condition visited fewer pages (*M* = 92.1, *SD* = 9.2) than those in the inconsistent condition (*M* = 113.7, *SD* = 21.4). None of the other interactions were significant ($p > .05$).

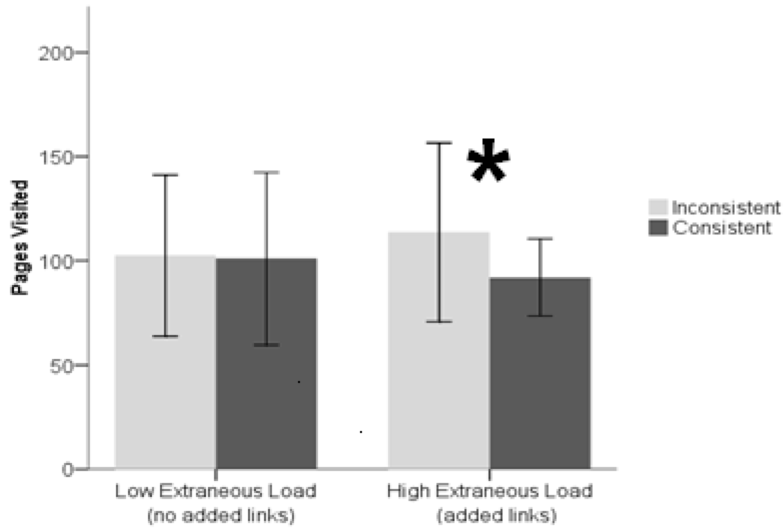

Figure 6. Simple two-way interaction between consistency and extraneous load for number of pages visited. Significant differences are marked with an asterisk.

Average Time Spent per Page

To assess condition differences on the average time participants spent on each page, a 2 (consistency or inconsistency) × 2 (high or low intrinsic load) × (high or low extraneous load) repeated measures ANOVA was used (consistency and extraneous load were between-groups). The main effects of consistency and extraneous load were not significant ($p = .74$, $p = .19$, respectively). The main effect of intrinsic cognitive load, however, was significant, $F(1,40) = 15.1$, $p \leq .05$, $\eta^2 = .27$. This indicated that participants spent significantly more time per page while answering the travel questions (low intrinsic load; $M = 4.1$ seconds, $SD = 1.5$) when compared to the finance questions (high intrinsic load; $M = 4.9$ seconds, $SD = 1.2$). None of the interactions were significant for the average time participants spent per page ($p > .05$).

Subjective Ease-of-use Scores

The subjective ease-of-use questionnaire was scored by averaging responses across all 18 questions resulting in a single average score (Table 9). To assess the condition differences on average ease-of-use score, a 2 (consistency or inconsistency) × (high or low extraneous load) ANOVA was used. The intrinsic load manipulation was not considered since the questionnaire was completed after participants finished both the high and low intrinsic load questions.

Main effects of consistency or extraneous cognitive load were not significant (p = .98, p = .24, respectively). These non-significant main effects indicated that neither the manipulation of consistency nor the extraneous cognitive load manipulation had a significant effect on a participant's average ease-of-use score. The two-way interaction of consistency × extraneous cognitive load was also non-significant (p = .50).

Table 9
Descriptive Statistics for Ease of Use scores

	Consistent				Inconsistent			
	Low Extraneous Load (no added links)		High Extraneous Load (added links)		Low Extraneous Load (no added links)		High Extraneous Load (added links)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Composite Ease-of-use	2.3	1.1	2.9	1.2	2.5	0.8	2.7	1.1

Note: Ease-of-use scores are not separated by intrinsic load since it was a within-group manipulation. Lower is better for Ease-of-Use scores.

In summary, the results of the performance analysis showed that the effect of consistency depended on the difficulty of the task or the cognitive load. Significant interactions were only observed when the extraneous cognitive load was high (extra hyperlinks present). These findings support the prediction that a task must be sufficiently difficult to see the benefits of consistency. Analysis of the subjective ease-of-use questionnaire indicated that the main effects of both consistency and extraneous cognitive load were non-significant. The prediction that consistent interfaces and lower extraneous load would improve ease-of-use scores was not supported.

DISCUSSION

The present experiment examined the effects of consistency under different levels of cognitive load. The goal was to determine if and how the consistency of an interface interacted with the level of cognitive load imposed by a task. Interface consistency is theorized as beneficial to user performance. However, past empirical studies of consistency have not always supported this idea showing a range of results from beneficial (AlTaboli & Abou-Zeid, 2007), to detrimental (Finstad, 2003), to insignificant (Rhee et al., 2006). These results may not be comparable due to the varied tasks used. For instance, if a task used in a study was too easy, perhaps the effect of consistency would have been too weak to detect.

The cognitive load, or difficulty, of the task was controlled by manipulating the intrinsic and extraneous cognitive loads. Intrinsic load was manipulated by asking participants easier and harder questions. Extraneous load was manipulated using a previously successful technique of including extra hyperlinks scattered throughout the content of the page (DeStefano & LeFevre, 2007).

Interactions

The cognitive load literature suggests that the combination of high intrinsic cognitive load and high extraneous cognitive load would be the most detrimental to performance (e.g. Sweller & Chandler, 1994). Similarly, much of the consistency literature asserts that consistency helps to reduce the working memory demand of a task (e.g. Proctor & Vu, 2006). These concepts led to the hypothesized interaction between consistency, intrinsic cognitive load, and extraneous cognitive load. Specifically, a combination of high cognitive load and inconsistency would maximize the cognitive demands of a task leading to poor performance. As hypothesized, the cognitive load manipulations did significantly moderate the effect of consistency in the form of error-rate.

The observed three-way interaction between consistency and cognitive load may help to explain some of the conflicting results from past research. When the difficulty of a task was highest, consistency significantly reduced the error-rate. These findings suggest that tasks must be sufficiently difficult to see significant effects of consistency. This finding is in line with previous literature finding conflicting results. Studies resulting in little or no effect of consistency (e.g. Ozok & Salvendy, 2000; Rhee et al., 2006) may have used tasks that were too easy to see an effect of consistency. Also, as the current results would predict, another study using more complex tasks produced significant results (Finstad, 2003). These results suggest that it is important to consider the task itself before comparing the results of any of the previous research on consistency.

The three-way interaction demonstrated support for a relationship between cognitive load and consistency. One of the sources contributing to the three-way interaction was a two-way interaction in consistency and intrinsic load, but only for the high extraneous load conditions (see Figure 4). Participants made significantly fewer errors in the consistent condition when the extra hyperlinks (high extraneous load) were included. Interestingly, this interaction occurred only for the travel (low intrinsic load) questions.

The direction of results in this aspect of the three-way interaction was surprising since previous cognitive load literature predicted that the maximum effects would be seen during the higher intrinsic cognitive load of the finance questions (e.g. Sweller & Chandler, 1994). An explanation for this seemingly contradictory finding could be that participants approached the financial questions with a different strategy. Overall, participants seemed be more cautious when answering the finance questions; participants spent significantly more time answering the finance questions and browsed through more pages before answering. The strategy might have helped since participants did not significantly differ in error-rate between finance and travel questions. Adapting strategies, while unanticipated, is not surprising. Participants were aware of the increased difficulty for the financial questions as seen in the manipulation check for intrinsic cognitive load; 64% of participants said the high load financial questions were more difficult. Limiting participant time may have at least partially prevented this change in strategy. Preventing this strategy adaption could possibly improve the intrinsic load manipulation by further emphasizing the differences in the two levels.

The other source of the three-way interaction was an interaction between intrinsic load (question type) and extraneous load (extra hyperlinks), but only for participants in the consistent conditions (see Figure 5). As predicted by the cognitive load literature, participants answering the high intrinsic load (finance) questions performed significantly worse under high extraneous load (Sweller & Chandler, 1994). However, this interaction only occurred for participants using consistent interface designs, a surprising finding. The initial prediction was that performance decreases would be maximized when participants were using inconsistent interfaces and experiencing higher levels of cognitive load. This interaction between the two types of cognitive load only for the consistent condition implies that consistency might have generated more cognitive load than inconsistency.

One possible explanation is that interface consistency does, in fact, contribute to higher cognitive load leading to the three-way interaction with intrinsic and extraneous cognitive load. This result was surprising given that consistency is hypothesized to reduce cognitive load; however, some of the past research supports this claim. A previous study showing significant effects of consistency observed an increased task performance time associated with consistency (Finstad, 2003). The author's explanation was that consistency had encouraged users to mistakenly over-generalize between the interfaces. That conclusion is suspect in the context of the current study. In Finstad's study, participants in the consistent condition were given interfaces that appeared consistent but were in essence conceptually incompatible. Participants were encouraged to make generalizations inappropriately when presented with an interface with incomplete consistency (only partially consistent). In contrast, the present study used interfaces that were either entirely consistent or entirely inconsistent; no condition used incomplete consistency. Also, no other results found in the current study supported the notion that consistency was harmful. While some aspects of this three-way interaction are difficult to explain, the significance of the interaction helped support the idea that consistency is affected by the difficulty of a task; furthermore, the types of task difficulty affect consistency differently.

The other interaction of note was the simple two-way interaction between consistency and extraneous cognitive load for the number of pages visited. A significant effect of consistency was only observed when the extraneous cognitive load was high (added hyperlinks). This interaction indicated that consistency had no effect on participants in the low extraneous load (no added links) in regard to the number of pages visited. However, consistency did have an effect when the extraneous load was high (added hyperlinks). In fact, when the extraneous load was high, participants in the inconsistent condition visited significantly more pages than those in the consistent condition.

This two-way interaction is conceptually identical to the previous study that used the added hyperlinks (Saadé & Otrakji, 2007). The interaction in number of pages visited suggests a certain level of navigational confusion or disorientation for participants who experienced the combination of inconsistency and high extraneous load. The inconsistency in the websites seemed to disorient users only when they subjected to the additional cognitive load induced by the superfluous hyperlinks. Saadé & Otrakji's study linked disorientation with aspects of screen design (of which consistency is a part) and cognitive load. The unique addition from the present study was that consistency specifically, rather than the broader concept of screen design, interacted with cognitive load.

Main Effects of Consistency

Participants in the consistent conditions navigated through significantly fewer pages when looking for the answers than participants using inconsistent interfaces. Unlike in some past interface consistency studies (e.g. Rhee et al., 2006), consistency provided a significant performance benefit. One theory of consistency suggests that a consistent design allows users to better predict system behavior (Nielsen, 1989). The increased predictability in the task may have allowed participants to better understand the organizational structure of the websites. This could have allowed participants to more efficiently move through the website as seen by the decreased number of pages visited.

Consistency had no other main effect on the other measures including completion time, errors and the average time spent on each page. These insignificant main effects for consistency are similar to those seen in previous research (e.g. Rhee et al., 2006). The lack of significant influence on the remaining measures, especially time and errors, supported the notion that the effects of consistency are less evident when cognitive load is uncontrolled.

Subjective Ease-of-use

The average ease-of-use scores did not vary significant among the conditions. Similar to another study that failed to see an effect (e.g. Rhee et al., 2006), participants only used a single version of the interface. As a result, participants had no direct means for comparison when judging the ease-of-use. Using a within-group design and administering multiple versions of the ease-of-use questionnaire might have better identified differences between the conditions. Perhaps the questionnaire could be adapted to have participants directly compare the different iterations of the system to determine which is perceived as better.

Limitations and Future Research

It is important to discuss some limitations of the current study. The manipulations seemed to be at least partially effective (as seen in the three-way interaction); however, they may not have been strong enough to elicit more interactions. The cognitive load literature has identified other manipulations of task difficulty that could be used in place of the present manipulations. Recreating the three-way interaction witnessed in the current study using more robust manipulations might expand upon the findings in this study. Ideally, with more effective manipulations, the interaction would also be seen in the other performance metrics.

Additionally, the present study treated consistency as a binary trait rather than manipulating dimensions individually as done in other research (e.g. Rhee et al., 2006). This manipulation makes it impossible to pinpoint which dimensions of consistency interact with the different types of cognitive load. For example, perhaps only physical consistency interacts with extraneous cognitive load. Future research should explore how the dimensions of consistency are affected by different types of task difficulty.

While it is important to understand how consistency affects various systems, future research should also explore how consistency affects disparate users. Researchers should consider how consistency impacts individuals sensitive to increased cognitive load. For example, older adults tend to have a reduced ability to cope with increased cognitive load (Van Gerven, Paas, & Tabbers, 2006). The two types of cognitive load might also have disparate effects on aging populations when compared to other age groups. Intrinsic load is a knowledge-based version of task difficulty since it relies on the difficulty of the materials. This type of difficulty may not have any more effect on older adults than it would on other age groups. Extraneous load, however, represents the more perceptually-based difficulty since it is based on the presentation of the materials. Given older adults general decline in perceptual abilities, extraneous load manipulations may be disproportionately detrimental to older adults. Consistency's interaction with cognitive load suggests that older adults might find consistency especially beneficial in these situations. For instance, consistency might help an older adult navigate a website with banner advertisements (a form of extraneous load) more accurately.

Results from the present study support the notion that when a task is sufficiently difficult, interface consistency is beneficial to performance. Furthermore, the type of difficulty induced may also alter how consistency affects performance. Past research on consistency demonstrated a range of conflicting results with some studies showing no significant performance effects. Studies showing no effects of consistency may have overlooked the importance of task difficulty; perhaps these studies used tasks that were too simple to show any effects of consistency. The present study helped to put forward a concept of when consistency matters: in difficult tasks with added distractions present. In this case, consistency may help to alleviate the difficulty of a task by allowing users to generalize knowledge between systems. When designing for consistency, developers should consider how a system is utilized and the possible difficulties that users might face. Any study of interface consistency must also control task difficulty.

APPENDIX

Appendix A: Protocol Script

Protocol for Thesis

Materials required for each participant:

1. This protocol.
2. Two copies of Informed Consent (per participant).
3. One Debriefing Form (per participant)
4. Noise machine

Arrive at least 15 minutes before participants are scheduled to arrive then do as follows:

1. Place signs in the hallway to direct the participants to the eye tracking lab and prop open lab door.
2. Turn on noise machine.
3. Review website orders for participant group and lay out correct instructions.
4. Start the workstations and get browser ready (on launcher page with bottom bar hidden).
5. Greet participants when they arrive.
6. Record participant names for attendance (as needed for the HPR).

Once participants have all arrived:

- Hello. Thank you for agreeing to participate in this study today. You can expect the entire study to take around 45 minutes to complete. Before we continue, please make sure that your cell phones are set to silent.
- The purpose of this study is to examine how website designs will affect your performance in searching for information.
- First, before we begin, I'll need you to complete this "Informed Consent" form. This form will explain the study and inform you of your rights as a participant. Once you have read it, please sign it along with the duplicate copy; one copy is for you and one is for me.

[Hand them Consent Form and wait for participant to finish reading/signing consent forms]

Study Introduction

- Okay, now we are ready to begin the study. This study consists of three parts, a questionnaire at the beginning, some web-based search tasks in the middle, and then a final questionnaire. All of these parts will take place on the computer.
- The first questionnaire covers basic demographic information along with how much experience you have using computer systems.
- The web-based search tasks are questions that will require you to search both travel websites and financial websites to find the answers. When you do find the correct answer, you just need to purchase the item to record the answer. All the navigation you do on these websites will be recorded so that I can see when you get to the correct answer. Please answer the questions in the order that they are written on the form.
- Finally, after the web browsing portion, raise your hand and I will open the final questionnaire for you. This final questionnaire will ask you questions about your opinions and experience with the websites. Once you complete this questionnaire, you will be finished with the study.
- Next I'll pass out the sheets with the questions. Just follow the instructions on these sheets and they will guide you through the study.
- Does anyone have any questions before we proceed?

[Wait for questions]

- Okay, we'll go ahead and begin. First I'll hand out the sheets and let you read them. When you are ready to begin, raise your hand and I'll start it up for you. Please work as quickly and as accurately as you can.

[Hand out question forms for that specific participant group. Press Alt-Ctrl-Shift-F9 to begin the recording software. The screen will flash black for a moment before the first questionnaire.]

After completing the demographics/computer experience questionnaire, the participants will use the websites to answer all the questions. After they finish all questions, they will be instructed to raise their hands.

[Quietly go over to the participant and press Alt-Ctrl-Shift-F9 again to finish the experiment and bring up the Ease-of-Use questionnaire.]

After participants click "Done" on the Ease-of-Use questionnaire, Morae will prompt you to save the recording. First, walk the participant to the door and thank them again for participating then hand them a Debriefing form. Once they have left, quietly save the recording using the appropriate participant number.]

File naming structure:
P03G03
P (1-6) G (1-8)

Appendix B: Demographics and Computer Experience Questions
(administered electronically)

1. Date of Birth (MM/DD/YYYY)

2. Gender

3. Race/Ethnicity

4. Current marital status

5. What is your current college major?

6. How many years of education have you completed?

7. In which type of housing do you live?

8. What is your primary language?

9. What is your occupational status?

10. If you work for pay, what is your primary occupation?

11. Please check all of the following devices that you have used.
    - ☐a Answering Machine
    - ☐b Cellular Phone
    - ☐c Compact Disk Player
    - ☐d Copy Machine
    - ☐e Cruise Control (in your car)
    - ☐f Fax Machine
    - ☐g Microwave Oven
    - ☐h On-line Card Catalog System (at the library)
    - ☐i Phone-in Banking (e.g., press "1" for "yes")
    - ☐j Video Cassette Recorder
    - ☐k Video Camera
    - ☐l Voice Mail
    - ☐m Automatic Teller Machines
    - ☐n Home Securities Systems
    - ☐o Pay at the Pump Systems
    - ☐p Clock Radio/Alarm
    - ☐q Video Arcade Games
    - ☐r ------ *None of the Above* --------

12. Please check which of the following items you own.

- ☐a Answering Machine
- ☐b Cellular Phone
- ☐c Compact Disk Player
- ☐d Cruise Control (in your car)
- ☐e Fax Machine
- ☐f Microwave Oven
- ☐g Video Cassette Recorder
- ☐h Video Camera
- ☐i Clock Radio/Alarm
- ☐j Home Computer

13. Have you had any experience with computers?

- ☐ Yes
- ☐ No

14. Of the input devices listed below, please indicate **ALL** devices with which you have had experience (check all that apply).

- ☐a Keyboard
- ☐b Mouse
- ☐c Light-pen
- ☐d Trackball
- ☐e Touch Screen
- ☐f Voice Input System
- ☐g Joystick
- ☐h ------ *None of the Above* --------

15. Indicate the total length of time you have used computers.

- ☐1 Less than 6 months
- ☐2 6 months but less than 1 year
- ☐3 1 year but less than 3 years
- ☐4 3 years but less than 5 years
- ☐5 At least 5 years

16. In the past, what was the highest frequency of your computer use over any 3-month period?

- ☐1 Once every few months
- ☐2 Every month
- ☐3 Once per week
- ☐4 Several days per week
- ☐5 Daily, but infrequently during the day
- ☐6 Daily, frequently during the day
- ☐7 Daily, most of the day

17. How frequently have you used a computer in the last three months?
   - ☐1 Less than one hour a week
   - ☐2 1 hour but less than 5 hours a week
   - ☐3 5 hours but less than 10 hours a week
   - ☐4 10 hours but less than 15 hours a week
   - ☐5 At least 15 hours a week

18. Of the basic computer operations listed below, please indicate all with which you are proficient (check all that apply).
   - ☐a insert a disk
   - ☐b open a file
   - ☐c delete a file
   - ☐d save a file
   - ☐e transfer files
   - ☐f use a printer
   - ☐g ------ *None of the Above* --------

19. Of the items listed below, please indicate all with which you are proficient (check all that apply).
   - ☐a Computer graphics (e.g., Photoshop, Harvard Graphics, AutoCAD)
   - ☐b Database management (e.g., Access, Filemaker, Lotus 123, etc.)
   - ☐c DOS
   - ☐d Electronic mail
   - ☐e Macintosh
   - ☐f Presentation software (e.g., PowerPoint, Freelance, etc.)
   - ☐g Programming package (e.g., Basic, C++, Fortran, etc.)
   - ☐h Spreadsheet (e.g., Excel, Quattro Pro, etc.)
   - ☐i Statistical package (e.g., SPSS, SAS, etc.)
   - ☐j UNIX
   - ☐k Windows
   - ☐l Word processing (e.g., Microsoft Word, WordPerfect, etc.)
   - ☐m Other (please specify) ______________________________

Appendix C: Participant Instructions Handout

Participant Instructions

If you have a question while you completing the study, please ask the experimenter. There are three parts to the study which are as follows:

1. Demographics and Computer Experience Questionnaire
2. Web-based Search Tasks
3. Ease-of-Use Questionnaire

Demographics and Computer Experience Questionnaire

Answer all questions to the best of your knowledge. When finished, click the *Done* button before proceeding to the next section.

Web-based Search Tasks

To answer the questions on the question sheet, you will need to visit four different websites. **If you try to answer the question on the incorrect website, you will not receive credit for that answer.** When you think you have found the correct answer, purchase that item. The system will inform you if your answer is correct.

Please find the correct answer to each question before moving on to the next question! Check off each question after you complete it. Work as quickly and as accurately as you can.

When you are finished with all the questions, please raise your hand and the experimenter will open the final questionnaire.

Ease-of-Use Questionnaire

Please answer all the applicable questions based on your experience with the websites you used. When you are finished, press the "Done" button.

Do not click anything else once you finish, just leave the new window alone.

Finished

You are now finished with the study. Please quietly get up so you do not disturb the other participants. On your way out, collect the "*Debriefing Form*" from the experimenter which will further explain the experiment. If you have any additional questions, feel free to ask the experimenter.

Question Form (24 total)

Travel Buddy Website

- ☐ 1. Find the cheapest flight to Los Angeles.
- ☐ 2. Find the most expensive three star hotel in Detroit.
- ☐ 3. Find the flight with one layover to New York City leaving after 6pm.
- ☐ 4. Find the two star hotel in Honolulu with a sauna.
- ☐ 5. Find the most expensive flight to Detroit with free in-flight snacks.
- ☐ 6. Find the cheapest hotel in New York City with valet parking.

Discount Destinations

- ☐ 1. Find the most expensive flight to New York City.
- ☐ 2. Find the cheapest hotel in Los Angeles.
- ☐ 3. Find the flight to Honolulu leaving after 4pm.
- ☐ 4. Find the three star hotel in Detroit with a Kitchen in the room.
- ☐ 5. Find the cheapest flight to Los Angeles with free in-flight snacks.
- ☐ 6. Find the most expensive hotel in Honolulu with a king size bed.

Continued →

Investing Info Website

- ☐ 1. Find the cheapest municipal bond.
- ☐ 2. Find the most expensive large cap savings and loan stock.
- ☐ 3. Find the medium term (6-10 years) speculative bond with the highest yield.
- ☐ 4. Find the small cap residential construction stock with the most revenue per employee.
- ☐ 5. Find the BB-C rated corporate bond with the highest coupon payment.
- ☐ 6. Find the most expensive biotech stock with a beta of at least 1.0.

Finance Central

- ☐ 1. Find the most expensive speculative bond.
- ☐ 2. Find the cheapest small cap biotech stock.
- ☐ 3. Find the short term (1-5 years) AAA-BBB corporate bond with the lowest yield.
- ☐ 4. Find the micro cap wireless communication stock with the highest profit margin.
- ☐ 5. Find the cheapest municipal bond with a yield of at least 1%.
- ☐ 6. Find the cheapest savings and loan stock with a revenue per employee of at least 500,000.

Appendix D: Ease-of-Use Questions (administered electronically)

1. Overall, I am satisfied with how easy it is to use this system.

2. It is simple to use this system.

3. I can effectively complete my work using this system.

4. I am able to complete my work quickly using this system.

5. I am able to efficiently complete my work using this system.

6. I feel comfortable using this system.

7. It was easy to learn to use this system.

8. I believe I became productive quickly using this system.

9. Whenever I make a mistake using the system, I recover easily and quickly.

10. The information (such as on-line help, on-screen messages and other documentation) provided with this system is clear.

11. It is easy to find the information I need.

12. The information provided with the system is easy to understand.

13. The information is effective in helping me complete my work.

14. The organization of information on the system screens is clear.

15. The interface of this system is pleasant.

16. I like using the interface of this system.

17. This system has all the functions and capabilities I expect it to have.

18. Overall, I am satisfied with this system.

Appendix E: Debriefing Form

Debriefing: Website Design
and Performance

Thank you very much for participating in this study. We could not conduct our research without your help.

This study was designed to examine how users will perform when using different interface designs. There were two different website designs for the different pages. One version had the green background and the other version had the blue background. Some participants only saw the pages with the green background, others saw just the pages with the blue background, and other participants used both versions. We measured the time you took to answer each question, the number of incorrect answers and your subjective feedback about your experience with the websites.

We expect to see that when participants use the two different website versions (the green AND blue versions) they not perform as well since they have to adapt to different website designs. Additionally, these participants will not be as satisfied with the websites.

Our goal is to understand how interfaces used for computers and other applications (like vehicles) can be better designed. Ultimately, we hope to use results from this study to aid in improving the design of systems.

If you are interested, we will share a summary of our results with you by mailing you a newsletter at your request. Because each individual’s data and test scores are completely confidential, there will be no way for us to mail your individual results.

Thank you very much for your time and cooperation. If you have any questions or suggestions about the study please do not hesitate to contact the director of the project:

Dr. Richard Pak
(864) 656-1584

Appendix F: Data Extraction and Calculation Process

Morae data files include a video recording of the screen along with a table detailing website navigation (see Figure 7). Data was exported from Morae in a .csv format then organized in Excel spreadsheets. Participant questionnaires, including the demographics and ease-of-use questionnaire, were exported to Excel and converted to an SPSS-friendly organization.

For the objective performance data, transitions between questions were deleted. These transitions represented question reading time and possible breaks taken between questions which were unrelated to the task. Participant performance was then calculated using formulas in Excel. The specific calculations for each metric were as follows:

- Time was calculated by subtracting the beginning timestamp from the ending time stamp.
- Errors were calculated by counting the number of times the error page was displayed.
- Total pages visited was measured by using Excel to count the number of navigations.
- Average time taken was determined by subtracting the timestamp of when a participant reached a page from the timestamp when a participant navigated to a new page.

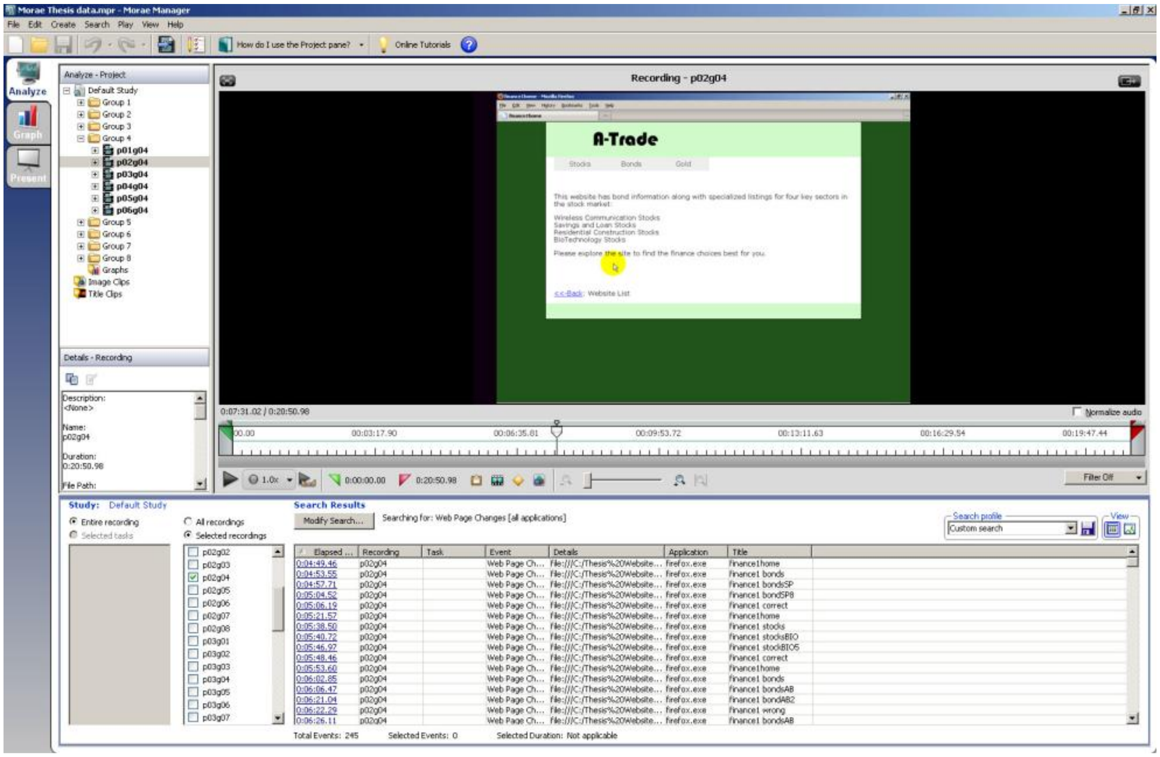

Figure 7. TechSmith's Morae Manager software. Screen capture video is shown on top with individual events shown in a table on the bottom.

Appendix G: Website Screenshots

Figure 8. Example page from Interface version 1 displaying stock information on one of the financial websites.

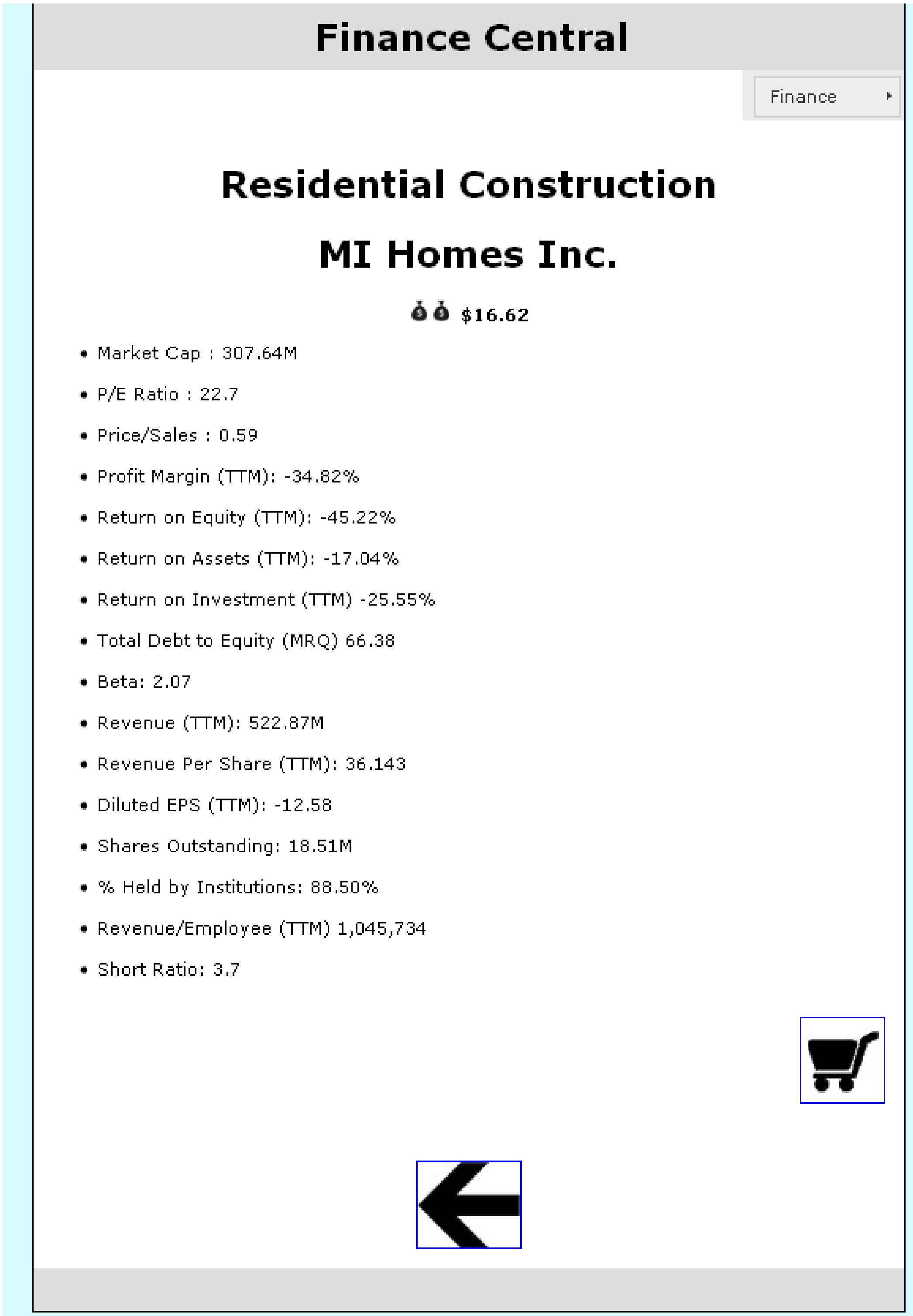

Figure 9. Example page from Interface version 2 displaying stock information on one of the financial websites.

REFERENCES

Adamson, P.J. and Wallace, F.L. (1997). A comparison between consistent and inconsistent graphical user interfaces. Pre-publication Report, University of Northern Florida, Department of Computer and Information Sciences, Jacksonville, Florida.

AlTaboli, A. & Abou-Zeid, M. R. (2007). Effect of Physical Consistency of Web Interface Design on Users' Performance and Satisfaction. In J. A. Jacko (Ed.), *Human-Computer Interaction: HCI Applications and Services* (pp. 849-858). Berlin: Springer.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*, 389-400.

Bannert, M. (2002). Managing cognitive load - recent trends in cognitive load theory. *Learning and Instruction*, 12, 139-146.

Bayer, N. L. (1992). Consistency in Interface Design: An Overview. *IPCC 92 Santa Fe. Crossing Frontiers. Conference Record*, September 29-October 3, (pp. 412-417).

Bednall, E.S. (1992). The Effect of screen format on visual list search. *Ergonomics*, *35*, 369-383.

Benbasat, I. & Todd, P. (1993). An experimental investigation of interface design alternatives: icon vs text and direct manipulation vs menus. *International Journal of Man-Machine Studies*, *38*, 369-402.

Brooks, C. J. & Bohemier, A. P. (1997). Helicopter door and window jettison mechanisms for underwater escape: Ergonomic confusion! *Aviation, Space, and Environmental Medicine*, *68*(9), 844-857.

Brünken, R., Plass, J., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53-61.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293–332.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 1-20.

Chalmers, P. A. (2003). The role of cognitive theory in human-computer interface. *Computers in Human Behavior*, *19*, 593-607.

Chapanis, A. (1953). Psychology and the Instrument Panel. *Scientific American*, 301-306.

DeStefano, D. & LeFevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*, 1616-1641.

Finstad, K. (2003, January). Analogical problem solving in casual and experienced users: When interface consistency leads to inappropriate transfer. Dissertation abstracts international (UMI Dissertation Information Service, No. 3058935).

Fitts, P.M., & Jones, R.E (1961). Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls. In W.H. Sinaiko (Ed.), *Selected papers on human factors in the design and use of control systems* (pp. 332-358). New York: Dover.

Grudin, J. (1989). The Case against user interface consistency. *Communications of the ACM*, *32*(10), 1164-1173.

Grudin, J. (1992). Consistency, Standards, and Formal Approaches to Interface Development and Evaluation: A Note on Wiecha, Bennett, Boies, Gould, and Greene. *ACM Transactions on Information Systems, 10*(1), 103-111.

Kellogg, W. A. (1987). Conceptual consistency in the user interface: Effects on user performance. In H. J. Bulinger & B. Shackel (Eds.), *Human-computer interaction; INTERACT '87IFIP Conference on Human-Computer Interaction* (pp. 389–394). Amsterdam, The Netherlands: Elsevier.

Lewis, J.R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, *7*(1), 57-78.

Mahajan, R. & Shneiderman, B. (1997). Visual and Textual Consistency Checking Tools for Graphical User Interfaces. *IEEE Transactions on Software Engineering*, *23*(11), 722-735.

Nielsen, J. (1989). Executive Summary: Coordinating User Interfaces for Consistency. In J. Nielsen (Ed.), *Coordinating User Interfaces for Consistency* (pp. 1-7). San Francisco, CA: Morgan Kaufmann Publishers.

Ozok, A.A. & Salvendy, G. (2000). Measuring consistency of web page design and its effects on performance and satisfaction. *Ergonomics*, *43*(4), 443-460.

Ozok, A. A., & Salvendy, G. (2003). The effect of language inconsistency on performance and satisfaction on the World Wide Web: results from four experiments. *Behaviour and Information Technology*, *22*, 155–163.

Ozok, A.A. & Salvendy, G. (2004). Measuring consistency of web page design and its effects on performance and satisfaction. *Computers in Human Behavior, 20*, 149-161.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measures as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Pak, R., Price, M.M. (2008). Designing an Information Search Interface for Younger and Older Adults. *Human Factors, 50*(4), 614-628.

Polson, P. (1988). The Consequences of Consistent and Inconsistent User Interfaces. *Cognitive Science and Its Applications for Human Computer Interaction*, 59-108.

Proctor, R. W. & Vu, K. L. (2006). Location and Arrangement of Displays and Control Actuators. In W. Karwowski (Ed.), *Handbook of Standards and Guidelines in Ergonomics and Human Factors* (pp. 399-409). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Reisner, P. (1990). What is Consistency? *Proceeding IFIP Third International Conference*, 175-181.

Rhee, C., Moon, J., & Choe, Y. (2006). Web interface consistency in e-learning. *Online Information Review*, *30*(1), 53-69.

Saadé, R. G. & Otrakji, C. A. (2007). First impressions last a lifetime: Effect of interface type on disorientation and cognitive load. *Computers in Human Behavior*, *23*, 525-535.

Satzinger, J., & Olfman, L. (1998). User Interface Consistency Across End-User Applications: The Effects on Mental Models. *Journal of Management Information Systems*, *14*, 167-193.

Schneider, W., Dumais S.T. and Shiffrin, R.M. (1984). Automatic and control processing and attention. In *Varieties of Attention* (pp. 1-27), New York: Academic Press.

Schneider, W., and Shiffrin, R.M. (1997). Controlled and automatic human information processing: 1. Detection, search, attention. *Psychological Review*, 84, 1-66.

Schnotz, W. & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, *19*, 469-508.

Shneiderman, B. (1987). *Designing the user interface-Strategies for effective human-computer interaction.* Reading, MA: Addison-Wesley.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human–computer interaction.* Tokyo: Addison Wesley Longman.

Smith, S. L., & Mosier, J. N. (1986). *Guidelines for designing user interface software* (Tech. Rep. No. ESD-TR-86-278). United States Air Force Electronic Systems Center, Hanscom AFB, MA.

Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognitive and Instruction*, *12*, 185-233.

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251-296.

Szabo, M. & Kanuka, H. (1998). Effects of Violating Screen Design Principles of Balance, Unity, and Focus on Recall Learning, Study Time, and Completion Rate. *Journal of Educational Multimedia and Hypermedia*, *8*(1), 23-42.

Tanaka, T., Eberts, R. E., & Salvendy, G. (1991). Consistency of human–computer interface design: Quantification and validation. *Human Factors*, *33*(6), 653–676.

Van Gerven, P.W.M., Paas, F.G.W.C., & Tabbers, H. K. (2006). Cognitive Aging and Computer-Based Instructional. *Educational Psychology Review*, 141-157.

Van Gerven, P.W.M., Paas, F.G.W.C., Van Merriënboer, J. J.G., & Schmidt, H.G. (2000). Cognitive load theory and the acquisition of complex cognitive skills in the elderly: Towards an integrative framework. *Educational Gerontology*, *26*, 503-521.