This study investigated how following standard usability guidelines impacts participants' perceptions of a website. In particular, the study focused on how participants reacted to (deliberately introduced) problems with the usability of a website. Our hypothesis was that these problems would negatively impact both task performance and affective response (causing, for example, participants to view the website as less desirable, or less trustworthy). A 2x2 between-subjects experiment (N = 80) was performed. The data collected (navigational flow, task performance, and affective response) provide insight into the consequences of designing for usability, and can help to guide designers struggling to navigate the tension between performance (usability) and emotional response (affect).

Networked information sources such as the World Wide Web play an increasingly important role in communication and dissemination of information. The popularity of the Web has drastically altered the nature of the relationship between media messages and users (Mastro, Eastin, & Tamborini, 2002), with messages now being designed and disseminated in a variety of different formats (Haubl & Trifts, 2000). Such diverse formats point not only to a lack of uniformity but may also result in users perceiving messages as complex, thereby leading to inability to process them (see Alty, Knott, Anderson, & Smyth, 2001). These concerns have prompted several researchers to argue for consistency of design based upon common guidelines for facilitating information processing (Cogdill, 1999; Marchionini, 1995; Shneiderman, 1998).

These guidelines predominantly focus on improving usability—a measure of how easy a system is to use (Nielsen, 2003). Advocates of usability argue that systems must be carefully designed to meet such criteria that can lead to improved user performance (Norman & Draper, 1986). The bulk of usability research has concentrated on criteria pertaining to task performance—the effectiveness and efficiency of a system in supporting users’ work. These findings are grounded in cognitive analyses of users’ interactions with the system. While such findings have led to major advances in design, this perspective largely ignores the attitudinal and emotional effects that interfaces designed to meet usability criteria may induce (Norman, 2002).

This restricted focus is especially surprising considering the growing body of research that has documented users’ psychological responses to structural features of new media (Reeves & Nass, 1996; see also Bucy, in press; Eveland & Dunwoody, 2001; Sundar & Wager, 2002; Lang, Borse, Wise & David, 2002; Sundar, 2000; Sundar, Kalyanaraman, & Brown, 2003; Sundar & Kalyanaraman, 2004). Since usability criteria can have a major impact on the design and structure of user interfaces, it is likely that messages employing these criteria may evoke
commensurate affective responses. This notion is consistent with recent assertions by scholars who have suggested that usability research should consider not only users’ cognitive responses but also their affective responses (Norman, 2002; Tractinsky, Katz, & Ikar, 2000).

The current investigation attempts to make a modest contribution in this direction by examining—along multiple dimensions—how people find information on Websites designed to have varying levels of usability. That is, by including not only typical task performance measures, but also less commonly-employed cognitive and affective measures, we attempt to provide a holistic view of individuals’ experience of a user interface.

The remainder of the paper is organized as follows: First, we review the literature on usability and interface design. Based on the literature review, we propose hypotheses and detail methods of an experiment designed to test those hypotheses. After presenting the results of the experiment, we conclude with a discussion of implications for research and practice.

Literature Review

Conceptualizing and Measuring Usability

Usability is a measure of how easy a user interface is to use (Nielsen, 2003). It is central to the design of interactive systems, particularly “information-abundant” ones such as Websites (Shneiderman, 1997). Laboratory studies have demonstrated that improving the usability of a system can lead to major gains in user performance. Case studies of usability work in practice provide evidence of improved productivity, increased sales, and high return-on-investment (ROI). Given these benefits, it is apparent that usability will play an increasingly important role in system design. The need to provide usable interfaces will constrain design choices and influence the features and interaction styles of virtually all complex Websites.

However, usability is not a single, monolithic construct. It has multiple aspects, including the three primary ones of efficiency, effectiveness, and satisfaction (ISO, 1998), as well as subsidiary ones such as learnability, memorability, and susceptibility to error. Efficiency measures the resources (typically time) expended to accomplish the user’s goals. Effectiveness measures the accuracy and completeness of the task result. Satisfaction measures users’ attitudes toward the system. However, while these measures may seem intuitively connected, no general relationship has been validated, and correlations among the measures may vary widely depending on task domain and context (Frojkaer, Hertzum, & Hornbaek, 2000). Despite such concerns, these usability measures are widely used to inform both research and practical system design.

According to the “mix of attributes” model (Eveland, 2003), it is critical to identify and study the various attributes of media so as to understand how they affect people, and how to design media appropriately to match peoples’ natural capabilities. Viewing technology and media from the perspective of usability helps to identify key attributes to study, as well as measures by which to evaluate the impact of those attributes. These attributes can then be linked to particular theoretical mechanisms to develop more comprehensive explanations of human behavior. In particular, it is apparent that the usability of an interface can have important effects on performance and attitudes.

Usability and performance

Since the emergence of human-computer interaction as an extension of applied ergonomics (Shackel, 1996), task performance has been considered a key component of usability. Researchers have identified specific aspects of interface design that impact performance (Cogdill, 1999; Shneiderman, 1997; Nielsen, 1994). In some cases, particular physiological or mental mechanisms have been identified to account for the performance effects. For example, Fitts’ Law explains the performance of various interface designs in terms of the human motor systems’ capacity (MacKenzie, 1992). Also, interfaces that provide easily recognizable cues are known to perform better because they place less stress on human memory (Shneiderman, 1998). Many of these aspects will strongly influence how people communicate using technology. In the context of the World Wide Web, two aspects of usability are particularly important: labeling and feedback.
Labeling and disorientation

The advent of hypertext has enabled the creation of large, interconnected information spaces. However, these systems are often incoherently organized and difficult to navigate (Bucy, in press; Dalal, et al., 2000). Such lack of navigability can lead to users becoming disoriented (May, Sundar & Williams, 1997) and “lost in cyberspace” (Conklin, 1987). This disorientation then manifests affectively as frustration and confusion (Bucy, in press; Ahuja & Webster, 2001), and cognitively as a high cognitive load (Eveland & Dunwoody, 2001). Cognitive load refers to the activity or workload imposed on mental processes. These affective and cognitive effects contribute to poor performance.

The exact mechanism which causes performance to suffer is not fully understood, but researchers have suggested an explanation based on the limited capacity model of information-processing (Lang, 2000). This model conceptualizes human attention and memory as part of an information-processing system (e.g. Card, Moran, & Newell, 1983). Like a computer, the system has finite processing power and can easily become overloaded. For example, three distinct subprocesses of memory have been identified—encoding, storage, and retrieval (Lang, 2000). If an individual cannot or does not allocate sufficient resources to one of these subprocesses, memory performance will suffer in predictable ways. Structural features of media can elicit automatic responses which shift resources to one subprocess at the expense of others. For instance, rapid cuts in a television broadcast can stimulate encoding, reducing the resources available for storage and retrieval (Lang, 2000).

Analogous to this view, the high cognitive load imposed by disorientation may reduce the resources available for other types of processing, such as comprehension and elaboration (Eveland & Dunwoody, 2001; Ahuja & Webster, 2001). Users therefore require more time to complete tasks (because comprehension is slower), and make more memory errors (because elaboration is impeded).

One design choice which may influence disorientation is labeling. Design guidelines strongly urge the selection of hyperlink labels that are clear, descriptive, and match users’ understanding of the content (Rosenfeld & Morville, 2002; Nielsen, 2000). However, despite the importance of labels for orienting users, many sites provide labels that are vague or non-descriptive. Hypertext and usability theory suggests that these labels will lead to disorientation and high cognitive load, which will then influence users’ task performance and memory. Based on this, we predict that a Website with clear descriptive labels will enable relatively high task performance and memory, compared to a Website with non-descriptive labels.

H1a: Users will complete the task (take less time to locate information on the Website) more rapidly when descriptive labels are provided than when non-descriptive labels are provided.

H1b: Users will complete the task more accurately (provide more correct responses) when descriptive labels are provided than when non-descriptive labels are provided.

H1c: Users will answer questions from memory more accurately when descriptive labels are provided than when non-descriptive labels are provided.

System feedback and transparency

A second aspect of interface design that can affect usability is transparency, or the need to provide users clear feedback about the system’s status (Conn, 1995; Shneiderman, 1998; Nielsen, 1994). Complex systems often have long delays during computation or data transfer, during which users cannot control the interface. Moreover, these delays (or latencies) can have significant effects. For example, Ramsay, Barbesi, and Preece (1998) had participants browse Websites (on a variety of topics). They found that introducing high latencies led users to rate informational Websites as less interesting, even though the content of the sites remained the same. Thus latency can act as a structural feature that induces psychological responses.

Whenever latencies exceed more than a few seconds, it is important to provide feedback about the system’s status to avoid confusing and frustrating users (Nielsen, 1998). Myers (1985), for example, found that adding feedback to an interactive database with random latencies caused users to view the database more favorably. So feedback, too, can act as a structural feature that induces psychological responses.

Feedback can be textual (e.g., a description of the system’s status), visual (e.g. a progress bar), or auditory (e.g., a tone that sounds when a process is complete). One of the key benefits of providing visual feedback is that it provides something for the user to focus on while the system is processing (Nielsen, 1994). In fact, during a latency period, the system is by definition nonresponsive, so the feed-
back mechanism is the user’s only connection to the system. If such a mechanism is absent or ineffective, users’ perceptions of the system are likely to be negatively impacted.

This is especially true on the Web since many Websites have long latencies—particularly for users with low-bandwidth connections (due to slow data transfers), and in some cases for all users (due to slow system responsiveness). However, current Web browsers generally do a poor job of estimating latencies and providing clear feedback during data transfer, and many Websites fail to provide adequate feedback during computation (Nielsen, 1998). HCI theory suggests this lack of feedback will frustrate and confuse users (Ramsay, et al., 1998). Under the same limited-capacity rationale explored in the previous section on labeling, it follows that the lack of feedback will increase cognitive load, and influence task performance and memory. Based on this conceptual rationale, therefore, we predict that a Website with informative feedback will enable relatively high task performance and memory, compared to a Website without feedback.

H2a: Users will complete the task more rapidly (take less time to locate information on the Website) when feedback is provided than when no feedback is provided.

H2b: Users will complete the task more accurately (provide more correct responses) when feedback is provided than when no feedback is provided.

H2c: Users will answer questions from memory more accurately when feedback is provided than when no feedback is provided.

Usability and attitudes

Although performance is a key component of usability, it is only one component. Usability also influences attitudes and emotions. The classical usability model encompasses satisfaction (ISO, 1998), but satisfaction as a construct is still poorly understood (Lindgaard & Dudek, 2003). In fact, satisfaction has often been considered a consequence, rather than a component, of usability (Hassenzahl, et al., 2001). And while systematic, validated measures of satisfaction have been developed (Chin, Diehl & Norman, 1988), they generally focus on ergonomic issues such as screen design, system performance, and learnability.

Further research is clearly needed to better understand how people respond to user interfaces, and how measures of satisfaction relate to measures of performance. More nuanced measures of user attitudes have been developed (Sundar & Kalyanaraman, 2004; Chen & Wells, 1999), and used to show that various formal features of new media such as animation, pictures, and sound influence user attitudes. However, the impact of usability criteria on user attitudes remains largely unexplored.

The nascent literature on “affective design” recognizes the variety and importance of attitudinal responses influenced by usability. Bucy (in press) argues that emotional responses may determine which interfaces (e.g. Websites) people choose to use, as they seek satisfaction beyond just task efficiency. Others have called for a shift from usability to “user experience” (Marcus, 2002; Wright, McCarthy and Marsh, 2000). These authors emphasize the importance of aesthetics in interface design, and stress the need for interfaces that promote engagement, fun, and delight rather than just task success.

Why is design that pays close attention to aesthetics and affective response important? One answer is that aesthetics matter for their own sake; another is that they can improve overall usability. Some design theorists argue that aesthetics matter for their own sake, and that HCI should therefore take explicit account of aesthetics (Hallnäs and Redström, 2002; Höök, et al., 2003). According to this view, HCI should be centered on “expressiveness.” That is, designers should be more interested in the expressions associated with a product or interface than in how well it supports a task. The focus of evaluation should shift from “How does the interface perform?” to “How does the interface feel?” Aesthetics are important because affective, as well as cognitive responses, matter.

These approaches have helped to stimulate thinking about the role of aesthetics in HCI, and may lead to innovative new designs. But they have not contributed to a systematic understanding of how people respond to particular interfaces, nor to improvements in usability per se. Hallnäs and Redström (2002) express interest in aesthetics and expressiveness for their own sake, and seem willing to sacrifice usability if necessary. This critical tradeoff must be considered in any discussion of aesthetics-centric design, since it poses a stark dilemma for designers.

In contrast, user-centered design advocates claim that aesthetics matter because they can improve overall usability (Tractinsky, et al., 2000). That is, “attractive things work better” (Norman, 2002). Evidence for this relationship is limited, but a few studies have shown interesting effects. Kim and
Moon (1998) found that visual design of user interfaces can elicit emotional responses. They focused specifically on trustworthiness, and found that online banking interfaces designed to be visually appealing were rated as more trustworthy than less aesthetically-appealing interfaces. However, they used only screenshots of the interfaces, not working systems, so users did not perform any tasks or interact with a system in any way. Their results therefore have extremely limited generalizability, but are still suggestive.

Tractinsky, et al., (2000) gave users financial-management tasks to perform, using various ATM interfaces. Both the aesthetics and the usability of the interfaces were manipulated. They found that post-task perceptions of system usability depended on interface aesthetics—not on the actual usability of the system. However, they did not show that actual usability (i.e., task performance, as opposed to perceived usability) depended on interface aesthetics. Still, their study provides the most striking evidence of usability’s multifaceted character.

However, what this line of research has failed to consider is that overall attitudes toward a Website may be influenced by usability as well as by “beauty.” In other words, it may not be that “attractive things work better,” but rather that “things that work better are more attractive.” To test this possibility, we propose our final set of hypotheses, which predict that conforming to usability guidelines will improve user attitudes toward the Website.

H3a: Attitudes toward the Website will be more positive when descriptive labels are provided than when non-descriptive labels are provided.

H3b: Attitudes toward the Website will be more positive when informative feedback is provided than when no feedback is provided.

Method

All individuals (N = 80) in a 2 (Feedback) X 2 (Labeling) between-subjects factorial experiment were randomly assigned to one of four experimental conditions: 1) feedback; labeling; 2) no feedback; labeling; 3) feedback; no labeling; 4) no feedback; no labeling. A fourth of the participants were assigned to each condition. After participants were exposed to the Website, dependent measures (pertaining to task performance, memory for site content, and attitude toward the Website) were elicited via both an online and a paper-and-pencil questionnaire.

Participants

Eighty undergraduate participants signed up to take part in the study for extra course credit. All participants signed an informed consent form prior to their participation.

Stimulus Material

Based on a pretest (N = 21), a Website, entitled HealthInfoCenter and containing information about health and nutrition, was created for use as the stimulus material in the experiment. The name HealthInfoCenter was chosen to sound “official” and trustworthy, but also to be sufficiently obscure that participants would not recognize it. A HealthInfoCenter logo was created using a medical icon, along with a neutral color scheme and typography, to convey an appearance of authenticity.

The Website consisted of a home page and five content pages. The home page was an index which linked to all five content pages. Each content page linked only to the home page (that is, there were no contextual links, navigation bars, or other forms of navigation). JavaScript was used to hide all of the Web browser’s navigation features, and to disable the right mouse button, thereby preventing participants from navigating away from the site, or using any tools besides clicking and scrolling.

The home page contained the logo, followed by a bulleted list of hyperlinks to the content pages. The content pages contained the logo, a descriptive title, and a few paragraphs of informative text about a particular topic, followed by a link back to the home page. Each page covered only one of the five topics: dieting, fat, cholesterol, vitamins, and alcohol. The pretest informed our choice of content. Furthermore, the actual content featured under each topic was taken from the WebMD site.

In addition, all the pages on the site displayed a question related to the site’s content, as well as a list of multiple-choice answers, or space to type in a free-response answer. The question appeared in a colored box below the main content of the page.

Experimental Treatment Conditions

Four versions of the Website were created, one for each experimental condition. Each version reflected a specific operationalization of selected usability guidelines. (See Appendix A for screenshots of all four experimental conditions).

The human-computer interaction literature contains standardized usability guidelines which practitioners can apply to the development of working
systems. These guidelines are intended to help designers avoid typical interface problems and improve usability without requiring extensive user testing. The stimulus was designed to test whether following (e.g., Cogdill 1999; Shneiderman 1998) or violating (e.g., Tractinsky et al., 2000) these guidelines affects usability and user responses to a Website. The two specific variables (descriptive labeling and informative feedback) employed in this study were carefully manipulated to conform (or violate) to classic usability guidelines.

To provide descriptive labels using clear terminology, two sets of hyperlinks for the home page were created. This manipulation adhered to the criteria outlined by Rosenfeld and Morville (2002). One set used the following descriptive labels: “Quick and healthy tips for diets that work,” “The good, the bad, and the truth about cholesterol,” “Vitamins and your body: healthy and naughty,” “Drink your way to better health,” and “What’s wrong with fat-free diets?” These labels were pre-tested to verify their effectiveness. The other set of hyperlinks used the non-descriptive label “Click here” for each link (see Tractinsky et al., 2000). The pretest mentioned earlier (N = 21) was used to verify the effectiveness of the manipulation.

To provide informative feedback, we created two pages with different loading speeds. Each page caused a ten-second delay from when the participant clicked a link on the home page to when the requested content page appeared. This delay simulated the effect of a slow network connection or overloaded server. We chose ten seconds for the length of the delay as that is the threshold level at which feedback is considered necessary (Nielsen, 1994). One page contained the HealthInfoCenter logo, a descriptive title, the words “Loading, please wait...,” and a progress bar that gave users a visual indication of how long they would have to wait until the next page loaded. This feedback mechanism was manipulated to closely follow the recommendations made by usability researchers such as Nielsen (1994) and Shneiderman (1998). The other condition violated this guideline (see Tractinsky et al., 2000) by not offering any system feedback in terms of how long the user would have to wait till the second page came up on the computer screen. Like before, the efficacy of this manipulation was also verified via the pretest. The second page contained only the logo and title, without a progress bar.

Essentially, participants used the Website by clicking a link on the home page, waiting for the subsequent page to download for ten seconds, and then viewing a content page. Once on a content page, their only option was to click a single link to go back to the home page. The home page loaded instantaneously, and no loading page appeared during that transition.

**Dependent Measures**

**Attitude toward the Website.**

The dependent measure of attitude toward the Website (Aw) was assessed by asking participants to respond to twelve nine-point Likert-type items. The items were adapted from Sundar and Kalyanaraman’s (2004) Website Perceptions scale, and asked participants to respond as follows: “On a 1 – 9 scale, where “1” means the term describes very poorly, and “9” means the term describes very well, please circle the number that indicates how well each term describes the HealthInfoCenter Website you just used,” followed by a series of twelve adjectives: Appealing, Useful, Positive, Good, Favorable, Attractive, Exciting, Pleasant, Likeable, High Quality, Interesting, Sophisticated.

**Task Performance**

The information-seeking task tested participants’ performance at retrieving information from the Website.

During the information-seeking task, ten questions (five multiple-choice and five free-response) were presented as participants browsed the Website. Two questions (one multiple-choice and one free-response) related to each of the five topics. The questions were simple factual queries such as, “Who is the author of Vitamania? A) Jean Mayer, B) Rima L. Apple, C) Meir Stampfer, D) Linus Pauling, E) Jeffrey Blumberg,” or “What type of food is replacing fats in the typical diet?” (answer: simple carbohydrates).

The questions were shown in random order to prevent any systematic learning or ordering effects. The current question was always visible at the bottom of the page, and participants were required to answer each question before proceeding to the next one.

During the reviewing task, participants browsed through the site at their own pace to gain a better sense of the material and prepare for the memory task. Once participants finished reviewing, they clicked a button to proceed to the memory task. A pop-up window prompted participants to confirm that they were ready to proceed, preventing any accidental termination of the review.
The following measures of performance were used:

- Information-seeking time: Total time (in seconds) required to locate the information needed to answer all ten questions.
- Information-seeking accuracy: Number of questions (out of ten) answered correctly.

**Memory**

The information-seeking task tested participants’ performance at answering questions from memory.

During the answering from memory task, ten questions (five multiple-choice, to test recognition, and five free-response, to test recall) were presented in random order. All the answers to the questions could be found on the Website; however, the Website content was hidden, and JavaScript restrictions prevented participants from navigating to previous pages, so participants were required to respond based on their memory of the Website.

The following measures of performance were used:

- Memory accuracy. Number of questions (out of ten) answered correctly.
- Memory time. Total time (in seconds) spent answering the questions from memory.

**Manipulation Check**

Four items were created to check the efficacy of the manipulation. To test the labeling manipulation, participants were asked whether the links on the page helped them to locate information, and whether it was difficult to find information on the site. To test the feedback manipulation, participants were asked whether they found it frustrating waiting for the pages to appear on the screen, and if they found the speed of the site acceptable. These four questions were measured on the same 9-point scale with “1” representing “Strongly Disagree” and “9” representing “Strongly Agree.”

**Behavioral Measures**

HCI theory suggests that interface design should have a strong influence on user behavior (Shneiderman, 1998; Marchionini, 1995). Hence, inclusion of behavioral measures may contribute to a better understanding of the relationship between Website design and usability, as well as attitudes toward the site. We developed custom instrumentation software to monitor participants’ behavior. The software was implemented in Java using the Tomcat application server and MySQL database. Each time a participant clicked on a hyperlink, the software recorded which participant made the request, which page was being requested, and the current time. Using this data we derived the number of “navigational moves” required to find the answer to a question, the total time required to find the answer to a question, as well as the total time spent on each task (information seeking, reviewing, and answering). These measures provide a more detailed view of the Website’s usability and the relationship between user behavior and attitudes.

**Control Measures**

Three questions pertaining to participants’ prior familiarity with the Website (“I’ve heard of HealthInfoCenter before”), content featured on the Website (“I was familiar with the information featured on this Website”) and the degree of attention paid while browsing the site (“I paid a great deal of attention when going through the Website”) were used as control measures. These items were anchored on a 1-9 scale ranging from “Strongly Disagree” to “Strongly Agree.”

Participants were also asked to report the amount of time spent on the Web and their experience with e-commerce. Spending a large amount of time on the Web, and trusting Websites sufficiently to engage in e-commerce were taken to indicate expertise and engagement with the Web as a medium. Participants’ responses to these items were also used as control measures.

Two questions (“I found the information presented on the HealthInfoCenter Website credible,” and “I trust HealthInfoCenter to provide me useful information about my health”) gauged participants’ perceptions of Website credibility.

Based on Sundar et al.’s (2003) suggestion that structural elements that provide feedback can make an interface interactive, the following question was used to gauge users’ perceptions of interactivity: “The way the Website was designed made it interactive.”

**Procedure**

The experiment was administered to groups of students in a campus computer laboratory that contained several Windows computers with Internet connections. Each group consisted of a maximum of eleven participants, and all participants were randomly assigned to one of four experimental conditions, with none of the sessions featuring more than
three participants per experimental condition. Before participants arrived, the experimenter turned on the computers and logged into an introductory page containing a single hyperlink labeled “START,” with instructions not to click on the link until told to do so.

Once all participants had arrived and been assigned randomly to a computer, they were asked to sign an informed consent form. Once all participants had completed the form, the experimenter explained the format of the study. Participants were informed that they would be helping to test a new Website created by Student Health Services, containing information about health and nutrition. Participants were asked to go through the information as they normally would and answer the questions to the best of their ability.

The experimenter then explained that the study procedure involved four different stages. First, participants would use the Website to find answers to specific questions that popped up on the screen. After completing this aspect of the study, they would then be allowed to review the Website to familiarize themselves with information featured on it. Third, after they had finished going through the site, they would be asked to answer questions from memory, without referring to the Website. Finally, they would be requested to fill out a paper-and-pencil questionnaire to elicit their perceptions of the Website. The experimenter displayed a dummy screenshot on an overhead screen to illustrate each stage of the procedure. The software presented these tasks to participants in sequence: first information seeking, then reviewing, and finally answering. Each task had to be completed before proceeding to the next.

After explaining the study, the experimenter asked the participants to click on the “Start” link and begin browsing the Website and answering the questions. Clicking this link caused the software to randomly assign each participant to a condition and to begin recording their use of the Website. As participants finished the final task—answering questions from memory—the software displayed a confirmation message and their participant identification number. At this point, they wrote their number on the post-stimulus questionnaire, and completed the questionnaire. After all participants had completed the questionnaire, they were debriefed, thanked for their participation, and dismissed.

After all participants had left, the experimenter verified that the ID numbers written on the questionnaires matched those reported by the software. The experimenter then closed and reopened the Web browser in preparation for the next experimental session.

Index Construction and Preparation for Data Analysis

The two manipulation check items pertaining to informative labeling (“The links on the page helped me to locate the answers to the questions” and “It was difficult to locate information on the Website”) were combined to form a single index labeled “Perceived Ease of Information Location.” This measure had a high degree of reliability (Pearson’s r = .66, p < .001).

The two manipulation check items pertaining to system feedback (“I found it frustrating waiting for the pages to appear on the screen” and “I’m satisfied with the speed of this Website”) were combined to form a single index labeled “Perceived Satisfaction with Website Speed.” This measure too exhibited a high degree of reliability (Pearson’s r = .56, p < .001).

Participants’ (total number of) correct responses were computed to obtain a measure pertaining to task performance. A “0” was awarded for an incorrect response and a “1” was awarded for a correct response. Accordingly, the “Task Performance” scale ranged from “0” (zero success at locating any information on the site) to “10” (hundred-percent success at locating all information).

A similar index was computed to measure participants’ memory for information featured on the site. This scale, labeled “Overall Memory” ranged from “0” (all incorrect responses) to “10” (all correct responses).

The dependent measure of attitude toward the Website was created by additively summing up the twelve items that comprised this index. This index exhibited a high degree of unidimensionality and internal consistency (Cronbach’s alpha = .91).

Also, the two items pertaining to credibility were additively summed up to form an index labeled “Website Credibility.” This measure had a high degree of reliability (Pearson’s r = .79, p < .001).

In addition to the attitudinal measures, behavioral data were gathered using the instrumentation software. The total number of links clicked on during both task performance and review times, as well as the amount of time spent searching for information while completing the task were used as performance measures. In addition, time spent reviewing the Website before responding to the memory
section was used as a separate measure of participants’ navigational behavior during the course of their browsing activity.

The main effects for labeling and feedback variables, as well as the interaction between these two independent variables, were investigated for each dependent variable index through a series of 2x2 factorial analysis of variance (ANOVA).

**Results**

**Manipulation Check**

To assess the efficacy of the labeling manipulation, a t-test was performed to compare ratings of ease of information location for the two labeling conditions. Results revealed a statistically significant effect for the manipulation, t (78) = -4.45, p < .001. Participants exposed to the condition with informative labels perceived it significantly more easy to locate information (M = 8.16, SD = 0.9) compared to their counterparts who were not exposed to informative labels (M = 6.4, SD = 2.33).

A similar t-test was performed to assess the efficacy of the feedback manipulation. The results revealed a statistically significant effect for the feedback manipulation, t (78) = -2.83, p < .01. Participants exposed to the condition with feedback reported greater satisfaction with the speed of the Website (M = 3.75, SD = 1.7) compared to their counterparts who were not offered any feedback (M = 2.68, SD = 1.47).

**Task Performance**

H1a predicted that participants will complete their task (locating answers to ten questions) more rapidly when descriptive labels, compared to non-descriptive labels, are provided. In a similar vein, H2a predicted that participants’ task accuracy will be higher (more correct responses) when descriptive labels, compared to non-descriptive labels, are provided. In a similar vein, H2b predicted that participants’ task accuracy will be higher when feedback, as opposed to no feedback, is provided. When the task performance index (ranging from “0” to “10”) was subjected to a 2x2 factorial analysis of variance, a significant main effect for labeling was obtained such that participants exposed to the descriptive labels scored more correct responses on the task measures (M = 9.73, SD = .67) than did those participants exposed to non-descriptive labels (M = 9.2, SD = .89), F (1, 74) = 7.34, p < .01. Thus, H2a was supported. Neither the main effect for feedback nor the interaction effect between labeling and feedback revealed statistical significance. The former meant that H2b was not supported.

**Memory**

H1b predicted that participants’ task accuracy will be higher (more correct responses) when descriptive labels, compared to non-descriptive labels, are provided. In a similar vein, H2b predicted that participants’ task accuracy will be higher when feedback, as opposed to no feedback, is provided. When the task performance index (ranging from “0” to “10”) was subjected to a 2x2 factorial analysis of variance, a significant main effect for labeling was obtained such that participants exposed to the descriptive labels scored more correct responses on the task measures (M = 9.73, SD = .67) than did those participants exposed to non-descriptive labels (M = 9.2, SD = .89), F (1, 74) = 7.34, p < .01. Thus, H2a was supported. Neither the main effect for feedback nor the interaction effect between labeling and feedback revealed statistical significance. The former meant that H2b was not supported.

**Attitude toward the Website**

H3a predicted that participants exposed to descriptive labels would exhibit more positive attitudes toward the Website compared to those participants exposed to non-descriptive labels. Similarly, H3b predicted that attitudes toward the Website would be more positive when participants were exposed to feedback, compared to those participants exposed to non-descriptive labels. Similarly, H3b predicted that attitudes toward the Website would be more positive when participants were exposed to feedback, compared to those participants exposed to non-descriptive labels. Similarly, H3b predicted that attitudes toward the Website would be more positive when participants were exposed to feedback, compared to those participants exposed to non-descriptive labels. Similarly, H3b predicted that attitudes toward the Website would be more positive when participants were exposed to feedback, compared to those participants exposed to non-descriptive labels. Similarly, H3b predicted that attitudes toward the Website would be more positive when participants were exposed to feedback, compared to those participants.
with no feedback. When the attitude toward the Website index was subjected to a 2x2 factorial analysis of variance, a significant main effect for labeling was obtained such that participants exposed to the descriptive labels exhibited a more positive attitude toward the site (M = 6.16, SD = 1.24) than did those participants exposed to non-descriptive labels (M = 5.42, SD = 1.22), F (1, 76) = 7.53, p < .01. Similarly, participants exposed to feedback rated the site significantly more positively (M = 6.10, SD = 1.08) than did those participants with no feedback (M = 5.50, SD = 1.89), F (1, 76) = 3.47, p < .05. Thus, both H3a and H3b were supported. Like before, the interaction between the two independent variables was not significant.

Additional analyses controlling for time spent on the Web and perceptions of prior familiarity yielded essentially redundant results, suggesting that these variables did not mediate the relationship between the usability measures adopted here and attitudes.

**Behavioral Measures**

In addition to the amount of time spent on the task performance measures, we also performed analyses on three other navigational measures. No specific hypotheses guided our exploration of the relationship between the independent variables and these navigational measures.

A 2x2 factorial ANOVA performed on number of links clicked on during task performance revealed a significant main effect for the labeling condition. Specifically, participants in the non-descriptive labels condition clicked on a significantly greater number of links to locate information (M = 23.626, SD = .72) than did those participants who were provided descriptive labels (M = 18.55, SD = .73), F (1, 74) = 26.53, p < .01. Although the main effect for feedback failed to attain significance, the two-way interaction between labeling and feedback was statistically significant, F (1, 74) = 6.45, p < .05. For participants with non-descriptive labels, the presence of feedback appeared to ameliorate the negative effects, because participants with feedback clicked on a lesser number of links, compared to those participants exposed to no feedback. A post-hoc test using Tukey HSD means confirmed the significance of this finding. On the other hand, for participants with descriptive labels, the presence or absence of feedback did not have any impact on their clicking behavior.

Another factorial ANOVA was performed on amount of time (in seconds) spent reviewing the Website (after completion of task performance). The results revealed statistically significant main effects for both the labeling and feedback conditions. Participants exposed to descriptive labels lingered for a longer amount of time (M = 362.64, SD = 184.30) than did those participants exposed to non-descriptive labels (M = 225.25, SD = 179.1), F (1, 74) = 12.30, p < .01. Also, participants spent more time when feedback was provided (M = 328.21, SD = 189.30) than when such feedback was absent (M = 257.15, SD = 192.1), F (1, 74) = 4.06, p < .05. However, there was no significant interaction between labeling and feedback.

The factorial ANOVA on the final behavioral measure, namely, number of links clicked on during (post-performance) Website review time mirrored the findings seen with the measure pertaining to amount of time spent on the site. Specifically, participants in the descriptive labels condition clicked on a significantly greater number of links when reviewing information on the Website (M = 4.52, SD = 1.99) than did those participants who were provided non-descriptive labels (M = 3.15, SD = 2.03), F (1, 74) = 10.29, p < .01. In a similar vein, when participants were provided with feedback, they clicked on a greater number of links (M = 4.30, SD = 2.12) than when they were not provided feedback (M = 3.37, SD = 2.04), F (1, 74) = 5.03, p < .05.

**Other Analyses**

A 2x2 factorial ANOVA on perceptions of Website credibility revealed a main effect for the feedback condition such that participants rated the Website as significantly more credible (M = 7.87, SD = 1.09) compared to those participants who did not receive such feedback (M = 7.20, SD = 1.55), F (1, 74) = 4.03, p < .05. Participants also rated the structure of the Website to be significantly more interactive when they were offered feedback (M = 5.65, SD = 1.44) than when they were not offered any feedback (M = 4.55, SD = 2.41), F (1, 74) = 5.71, p < .05.

In summary, the effects of labeling seem more profound than do the effects of feedback. All the hypothesized findings with the labeling variable received strong support. The feedback variable did not seem to have an impact on users’ task performance. However, the hypotheses pertaining to memory and attitudes toward the site received support. In addition, feedback was shown to have a significant effect on perceptions of Website credibility as
well as structural interactivity. Finally, it appears that usability guidelines also have important behavioral implications, given that violation or conformity of usability guidelines influenced the amount of time spent on the site as well as participants’ clicking behavior.

Discussion

Although usability research has commanded a substantial body of attention, the existing literature has focused on task performance measures while neglecting to take user attitudes into account. Such considerations prompted the current investigation. The major goals of this paper were to examine whether conforming to classic usability guidelines resulted in improved task performance, greater memory for Website content, and more positive attitudes toward the Website. As hypothesized, closely following the guideline of labeling led to more efficient task performance, increased memory for Website content, and more positive attitudes toward the Website. We found similar results for the other variable, namely feedback—but not on any of the typical task performance measures.

Findings from the present study offer several interesting insights that are worthy of discussion. First, it was surprising to note that the element of feedback did not have a substantial effect on users’ task performance. The most obvious implication of this finding is that even widely-recommended usability guidelines need to be examined in different, carefully constructed contexts. Contrary to the existing methodological approach on usability research, our experimental investigation adopted a 2x2 factorial design. Since previous studies have shown that feedback is an important indicator of task performance, we suggest that even commonly-used usability guidelines need to be examined in concert rather than in isolation. Perhaps (and as can be seen from the data), some guidelines (e.g., descriptive labeling) are so psychologically powerful that they obfuscate the effects of other, less-powerful guidelines (e.g., feedback). Clearly, more guidelines need to be examined in similar experimental settings to isolate and delineate specific effects.

A major contribution of this study lies in its demonstration of the relationship between usability and attitudes. As discussed earlier, it appears that conforming to recommended usability norms not only affects task performance (as has been seen in several previous studies), but more importantly, also influences users’ overall impressions of the site. This finding is consistent with several studies examining the psychological effects of formal features in new media and suggests that users’ attitudes can be influenced by careful manipulation of usability guidelines. Obviously, several other usability guidelines need to be examined to further establish the scope of this relationship but the findings reported here certainly warrant further investigation.

In addition, the findings pertaining to attitudes imply that usability must be considered when evaluating structural features of technology, as usability may be instrumental in affecting users’ attitudes. To evaluate this effect, researchers may consider using more structured, task-centered designs in addition to simply exposing them to more passive Web stimuli. Also, measures of usability need to be collected to complement measures of attitudes and other affective responses to promote richer understanding of human-computer interaction.

A key methodological contribution of our study is to demonstrate the need for holistic approaches to the study of usability, particularly in online environments such as the Web. Task performance and attitudes are interconnected in complex ways, and researchers should seek to tease out these connections and how they interact in different contexts. To accomplish this goal, it is critical to employ comprehensive methodologies that measure multiple aspects of the user experience. In particular, gathering behavioral data using unobtrusive techniques such as clickstream analysis provides an independent assessment of usability and user behavior. This is an important complement to questionnaire data.

The findings reported here also have implications for technological variables such as interactivity. Interactivity is one of the key defining elements of the Web. Inherent in some recent conceptualizations of interactivity is the notion of feedback (see Sundar et al., 2003; see also Heeter, 1989). That is, if a system (in this case, a Website) offers prompt and informative feedback, users exposed to that system will perceive the system as interactively responding to them. This conceptual proposition received empirical support in our study because participants exposed to the feedback condition rated the structure of the Website as more interactive than did those participants exposed to no feedback. Interestingly, it was observed that accurate labeling did not have any effect on participants’ perceptions of interactivity. This is particularly interesting given that some recent research on interactivity has suggested that presence of such elements such as hyperlinks.
may increase perceptions of interactivity (Sundar et al., 2003). A clear recommendation is that structural elements such as hyperlinks need to be examined in a variety of different contexts to better understand their effects on interactivity.

The finding that the presence of feedback increased users’ perceptions of Website credibility is noteworthy. Intuitively, there seems to be little reason to suspect that usability criteria such as feedback would influence perceptions of credibility. However, the results obtained here suggest that some elements of usability can indeed affect users’ perceptions of credibility. It remains to be seen whether users equated the transparency of the system with the trustworthiness of the content, and hence evaluated the credibility of the site on the basis of usability rather than actual content. Therefore, not only does this need to be examined in different Websites (for example, news Websites), but also to see whether other usability guidelines may also manifest similar effects.

Findings from the behavioral measures suggest that usability guidelines have a strong impact on navigability. By conforming to proper guidelines, we showed that the amount of time spent on the site as well as the number of links clicked on decreased when users were tasked with finding information, resulting in more efficient navigation. Interestingly, this pattern was reversed during the review phase. Then, the presence of descriptive labels and feedback induced users to spend more time and also click more links. This suggests that conforming to usability criteria has potential to increase the “stickiness” of a Website.

Our findings also place a formidable constraint on design. The movement toward “user experience” has stressed the importance of affective response to interface design. Many theorists have suggested focusing on the aesthetics of the design to meet this objective (Hallnäs & Redström, 2002). Our data show the limitations of this approach. Since there is often a tradeoff between aesthetics and usability, focusing too closely on aesthetics will only hurt the user experience, since usability has a strong impact on user attitudes. This creates what we call the Designer’s Dilemma: Suppose we want to sacrifice some usability to meet some other goal—how much will this hurt user performance on and response to our site? This dilemma is particularly critical for e-commerce or media sites which often have numerous objectives (brand image, credibility, etc.) that may conflict with usability guidelines. While designers will inevitably have to make painful choices, our data suggest that empirical findings can bear directly on this problem. In addition, relevant data in this area might enable usability specialists to argue more convincingly for designs in step with HCI guidelines. More empirical work is certainly warranted here.

Our data also point to problems in current conceptual understanding of affective responses in HCI. If both usability and aesthetics can contribute to affective response, how can we disentangle these effects? We see a need for more refined experimental designs as well further theoretical consideration of how the basic psychology of affect and cognition is relevant to interface design. In particular, better characterizations of what kinds of affective responses we are interested in and how to measure them would be useful. Of course, it would be valuable to test different user populations, and to look for individual-level characteristics that modulate affective responses. Different people may exhibit different sensitivities to usability or aesthetics, which would argue against adopting universal principles for design.

In sum, although some scholars have expressed concern over the role of usability guidelines in the current technology climate (Norman, 2002; Tractinsky et al. 2000), we believe that the death of usability has been greatly exaggerated. As we have shown, usability matters greatly to individuals’ affective responses as well as to their observed task performance. Developing a more nuanced understanding of usability and its effects on users’ behavior on the Web presents a promising challenge for future research and practice.

References


navigational tools and subject matter expertise on browsing and information retrieval in hypertext. Interacting with Computers, 10, 129-142.


Appendix A
Screenshots of the stimulus material

Screenshot of the loading screen for the no-feedback condition.

*Five things you should know about your diet... but probably don’t*
Five things you should know about your diet... but probably don't

Loading section, please wait...

0%  100%

Screenshot of the loading screen for the informative feedback condition.
Five things you should know about your diet... but probably don't

Please choose a section:

- Quick and healthy tips for diets that work
- The fat of the land
- The good, the bad, and the truth about cholesterol
- Vitamins and your body: healthy and naughty
- Drink your way to better health

Question #1

What type of food is replacing fats in the typical diet?

Answer:

Type your answer here:

Submit answer

Screenshot of the home page for the descriptive labels condition.
Screenshot of the home page for the non-descriptive labels condition.