Intl. Journal of Human–Computer Interaction, 29: 404–418, 2013

ISSN: 1044-7318 print / 1532-7590 online
DOI: 10.1080/10447318.2012.715277

Development of a Weighted Heuristic for Website Evaluation for Older Adults

Kyle R. Lynch[1], Diana J. Schwerha[1], and George A. Johanson[2]
[1]*Department of Industrial and Systems Engineering, Ohio University, Athens, Ohio, USA*
[2]*Department of Educational Studies, Ohio University, Athens, Ohio, USA*

Older adults are the fastest growing population of Internet users. As websites acquire a greater number of older visitors, it is vital that they are designed with this demographic in mind. Older users typically have different user characteristics than younger users; they may have changes in perceptual abilities, motor skills, cognitive abilities, mental models, and confidence in the use of technology. This research documents the development of a new weighted heuristic measure for evaluating the usability of websites for older adults and its validation with performance testing. Results from a repeated measures analysis of variance indicated that websites with different heuristic classifications were significantly different with respect to performance metrics and System Usability Scale ratings. Conclusions point to the need for web design that takes into account preferences and abilities of older web users.

1. INTRODUCTION

Members of the baby boomer cohort were born between 1946 and 1964 and continue to increase their use of technologies. In 2010, 74% of the baby boomer cohort used the Internet, 81% used a cell phone, and 46% connected wirelessly to the Internet (Rainie, 2010). In addition, older adults are the fastest growing age group of Internet users (Hanson, 2009; Kurniawan & Zaphiris, 2005). By 2020, there will be a predicted 55 million people older than 65 years of age, a 36% increase for that decade (Administration on Aging, 2010). Having effective use of the Internet by seniors is important because it can facilitate older people in finding healthcare information, setting reminders for important events, and using interactive games to promote memory training (Czaja & Lee, 2008).

The ability for Internet sites to be designed for this cohort is thus critical to our communication, business, and health systems. Although much research has been done on developing guidelines for web developers, the literature on website heuristics for older adults is lacking (Hart, Chaparro, & Halcomb, 2008; Ji et al., 2010; Taveira & Choi, 2009). For companies and organizations to fully satisfy the users in this emerging demographic of elderly Internet users, they must develop websites that are compatible with older adults' unique wants and needs (Zaphiris, Ghiawadwala, & Mughal, 2005).

Current research, however, has indicated that technology use by older adults is complex, and often factors involving its adoption are described by interactions rather than by discrete factors (Hanson, 2010). Technology adoption can be a function of cognitive abilities, perceptual abilities, and the user's perception of the benefit of learning a new technology. Because of this, user preferences that are age specific and the ability to determine features that have possible interaction effects are of great value to designers.

The purpose of this research was to develop and test a weighted heuristic that would evaluate website usability for older adults. Male and female individuals older than 50 years of age participated in the experiment. For this scale development study, a younger group was not used because the authors were not interested in results that reflected comparison, only in results that reflected the usability preferences of an older cohort. Furthermore, the measure is not designed for use with a younger group and might well function differently in this case. Three websites having different levels of usability, as defined by the heuristic, were tested with performance metrics (clicks, errors, and time on task) and System Usability Scale (SUS) ratings. Hypothesis 1 was that a significant difference would exist in the performance metrics (clicks, errors, time) among each of the websites. Confirmation of this would demonstrate that the categorization of website usability by heuristic score was predictive of performance. Hypothesis 2 was that there would be a relationship between task performance and SUS scores. Confirmation of Hypothesis 2 would demonstrate Nielsen and Levy's findings that performance and preference are correlated. This is one of the bases of human–computer interaction and verifies the usability tests as accurate predictors of performance on user interfaces. The contribution to the literature for this

This research study was (partially) supported by the National Institute for Occupational Safety and Health Pilot Research Project Training Program of the University of Cincinnati Education and Research Center Grant #T42/OH008432-05.

Address correspondence to Diana J. Schwerha, 278 Stocker Center, Department of Industrial and Systems Engineering, Russ College of Engineering and Technology, Ohio University, Athens, OH 45701. E-mail: schwerha@ohio.edu

project is in its development of a weighted heuristic (weighted according to preferences of older adults) and its validation of the heuristic by performance metrics.

2. LITERATURE REVIEW

The following review of the literature spans the topics of normal aging changes, website usability metrics for older adults, usability evaluation methods, and usability studies that have used older adults. It discusses the theoretical background and context of the development of the heuristic described in this article.

2.1. Normal Aging Changes That Can Affect Computer Use

Older adults are the fastest growing demographic of Internet users today (Hart, Chaparro, & Halcomb, 2008), and their special interface requirements need to be considered while designing websites. As individuals age, their perceptual, motor, and cognitive skills can change. In addition, they may have different mental models as to how information should be organized, and they may have different levels of confidence. The main challenges that affect an older adult's ability to use the Internet are decreased memory, slower thought-processing speed, fine motor control issues, hearing and vision loss, and general lack of technical knowledge (Hart, Chaparro, & Halcomb, 2008; Smith, Sharit, & Czaja, 1999). These changes can inhibit a senior's ability to see, read, understand, and use websites effectively.

The change in a person's vision capabilities is a vitally important consideration when designing a website for older adults. If the website is difficult to see, it will have severe consequences on the performance, and the user may not desire to stay on the site. Research has demonstrated that older users dislike fonts on websites that were perceived as being too small to read through (Hart, Chaparro, & Halcomb, 2008). Some of the main effects of the natural aging of the eyes include decreased focus on near objects, different color perception, decreased contrast sensitivity, and decreased perceived brightness (Arch & Abou-Zhara. 2004). Many common diseases can further amplify these vision problems. These changes have obvious effects on how well an older person can use a website.

Another challenge in operating input devices for seniors is the reduction in fine motor control ability. Research has shown that elderly subjects with probable Alzheimer's disease or Mild Cognitive Impairment have increased deterioration of their fine motor skills (Yan, Rountree, Massman, Smith Doody, & Hong, 2008). Other work has demonstrated that that older adults made slower mouse movements than younger adults, but their movements also had less variation (Chaparro, Boran, Fernandez, Choi, & Kattel, 1999). The fine motor control problems experienced by the elderly can have a large negative impact on how well they can effectively use a mouse and keyboard to navigate a website (Morrell, 2002).

Cognitive abilities can also be different between older and younger adults. As people age, their working memory and thought processing slow down. In a research study that examined 403 randomly selected subjects from ages 68 to 78, it was shown that 38% of senior citizens experienced Age-Associated Memory Impairment and 27% fulfilled Aging-Associated Cognitive Decline criteria (Hanninen et al., 1996). These disorders are typical among older populations, and even more severe memory-related problems can occur as a result of dementia or Alzheimer's. The decrease in the brain's ability to process and recall information makes it increasingly difficult to learn new tasks and apply previous knowledge (Mead, Lamson, & Rogers, 2002). Learning how to use and navigate a webpage, for example, could be very arduous for someone suffering from these cognitive impairments.

Aside from those more severe age-related changes in cognition, normal age-related changes in fluid and crystallized intelligence can affect a person's ability to navigate the web. Tasks, such as using search and dynamic context changes, that require fluid intelligence rather than crystallized intelligence may be more difficult to learn (Hanson, 2010). Conversely, research has indicated that for the older population higher scores in fluid intelligence are predictive of web use and indicate individuals who may be more likely to participate in email, games, and online shopping (Czaja et al., 2006; Czaja, Sharit, Hernandez, Nair, & Loewenstein, 2010). Strong navigation structures that reduce the ability to learn on the go can serve to mediate possible changes in fluid intelligence.

Another common impairment for seniors is loss of hearing. Gradual hearing loss due to aging is referred to as presbycusis. It is estimated that 30% of people older than age 65 have significant hearing impairment (Dugale, 2010). In addition, presbycusis makes it difficult to hear distinct, high-pitched sounds and sounds at very low volumes (National Institutes of Health [NIH], 1997). This can cause problems for seniors if there are audio or video elements to a website that do not include subtitles (Namatame, Kobayashi, & Harada, 2003).

The last important limitation for seniors is their lack of technical knowledge and the occasional fear of technology. Research has shown that elderly subjects continuously failed to perform simple computer tasks and needed clarification on computer terminology (Fidgeon, 2006). For example, seniors regularly failed to scroll down the page to get more information and did not understand what was meant when the moderator requested to "open the minimized window." In addition, the elderly group was more likely to click on the parts of the webpage where there was no link and complained that the hyperlinks were not clearly defined on the sites. Older users were also fearful of downloading documents from the Internet due to the fact that they were worried about viruses (Fidgeon, 2006). The lack of general technical knowledge makes it very difficult and frustrating for senior citizens to access and obtain useful information from the Internet.

Because of the limitations common among senior citizens, they are less likely to use the Internet to efficiently obtain desired information (Mead et al., 2002). A study comparing two age groups' task performance found that older users take longer to complete tasks and make more errors (Echt, Morrell, & Park, 1998). Certain disabilities and diseases could increase one or many of these impairments, causing further inability to perform effectively on a computer (Chisholm, Vanderheiden, & Jacobs, 2001). Because of this, there is a large degree of variability in abilities between older people that should be utilized when designing websites targeted at older users.

2.2. Making a Website Senior Friendly

In the last decade or so, there has been much research on what can be done to make websites more senior friendly. Many of the recommendations have been compiled from research involving focus groups, experimentation, and personal experiences. The National Institute on Aging and the National Library of Medicine (NIA; 2002) collaborated to form a checklist of items for making a website senior friendly. This document clearly explains some of the issues that senior citizens have when attempting to find information online. The guidelines in this document are categorized as readability, presentation, incorporating media, and navigation (NIA, 2002). The NIH guidelines, however, were not weighted and do not reflect the relative importance of some usability characteristics over others. The heuristics that were developed in this research were derived from the NIA guidelines, but they have added weighting factors based on older individuals' preferences.

One of the most important aspects of designing a senior friendly website is ensuring that the text is readable. Certain steps can be taken to ensure that people with declining vision can still access the information that is displayed on the site, the most important of which is the size of the text. It is recommended that the text be no smaller than 12 point (NIA, 2002). It is also preferable to have an easy way to adjust the text size for all the content on the site (including menu items). Also, it is important that there is an appropriate amount of contrast between the text and the background (NIA, 2002). It is recommended to have either dark text on a light background or light text on a dark background (NIA, 2002). The most readable is black text on a white background, and this is highly preferred. It is also suggested that the text be double spaced have sufficient spacing between the letters in each word, to avoid the text appearing cramped or crowded (Zaphiris, 2005). In addition, it is important to avoid any scrolling text, as it is hard for seniors to read quickly while text is moving across the screen (NIA, 2002; Zaphiris, 2005).

How a user moves from page to page within a website is important to consider when designing for older users. Proper labeling and the size and position of the links and menus are vital factors in providing the best user experience. Due to the declining motor skills of older adults, it is harder for older adults to move the cursor with great precision (Smith, Sharit, & Czaja, 1999). To address these changes, menu items and links should be large and clear, as well as definitively separated from each other (NIA, 2002; Zaphiris, Ghiawadwala & Mughal, 2005). Complex menu structures such as drop-down and dynamic menus should be avoided (NIA, 2005). These can be confusing for some users and require significant manual dexterity to operate.

Hyperlinks should be underlined and have a consistent color throughout the entire website (Nielsen, 1997). They should also change appearance when the cursor is placed over them and should permanently change to a different color once clicked (Kurniawan & Zaphiris, 2005; NIA, 2002). This helps to show the user where they have already been on the site, and it is especially important for older users with limited cognitive ability. It is also important for users to be able to see where they are on the site in relation to the home page (NIA, 2002). Most websites are organized hierarchically, and the users should be able to clearly what level or sublevel they are currently viewing. In addition, it is suggested that there be a clear, easy-to-find, robust internal search box for quick access to information on the site (Kurniawan & Zaphiris, 2005, NIA, 2002; Zaphiris, Ghiawadwala & Mughal, 2005).

The manner in which content is organized and presented to the viewer should also be specifically addressed to fit older users' needs. Ideally, all important information should be contained "above the fold," meaning that the user doesn't have to scroll down at all to get to the information (NIA, 2002). There should be a minimal amount of information on each page within the site (Kurniawan & Zaphiris, 2005, NIA 2002; Zaphiris, Ghiawadwala & Mughal, 2005). It is much more desirable for the user to use links and menus on the site than to have to search through information displayed on one page. There should be an appropriate amount of white space so that the site looks neat and organized, as opposed to cluttered or empty (Zaphiris, 2005).

The content within the navigation menu should be related to a typical user's tasks and goals (e.g., Is it a shopping website or is it mainly for information seekers?). Similar items should be grouped together, and labels should clearly and simply describe the information in the sublayers contained within (Zaphiris, 2005). In addition, headings should be larger than the main text and should accurately describe the content of the main text (Kurniawan & Zaphiris, 2005; NIA, 2002; Zaphiris, Ghiawadwala & Mughal, 2005).

Accessibility is especially important for users with specific disabilities. Issues such as color deficiency, hearing loss, other specific vision problems, and technical hindrances should be addressed (Chisnell & Redish, 2001). Websites designed for older users should use multimedia sparingly. If they do choose to use multimedia, it is recommended that they provide text equivalents of the content, either with subtitles or a transcript (NIA, 2002). Also, designers should not use color-blind sensitive colors within close proximity (NIA, 2002). There are many tools online to test if a website is visible by people with color deficiencies. In addition, the site should be free of technical terminology and provide a number or help screen for users that require assistance with the website (Zaphiris, Ghiawadwala, & Mughal, 2005).

2.3. Usability Evaluation Methods

Some of the common methods for evaluation of the usability of websites include the SUS; Goals, Operators, Methods, and Selection rules (GOMS); Keystroke Level Model (KLM); Heuristics; Walkthrough method; and the Software Usability Metric (SUMI; Baber, 2005). The different approaches each have their own advantages and disadvantages and are used by different professions to look for varying aspects and dimensions of usability. Recent articles have stressed the use of strategy in website evaluation (Chiou, Lin, & Perng, 2010).

One of the most established usability metrics is the SUS. The scale was originally developed and then revised (Bangor, Kortum, & Miller, 2008; Brooke, 1996). The scale consists of 10 items and utilizes a 5-point Likert scale. The items are simple statements about the how easy the system is to use and whether the user felt comfortable using the system (Brooke, 1996). The subject gives a response from 1 (*strongly disagree*) to 5 (*strongly agree*), and an SUS score is computed on the scale of 0 to 100. Generally, a score above 70 is considered "usable" (Brooke, 1996).

Bangor et al. (2008) performed extensive research into the validity of the SUS. They examined 2,324 SUS scores from 206 usability tests and found that they were highly reliable, with an alpha of 0.91. The authors found a correlation between age and SUS score; however, there was no significant correlation found between SUS and gender. It was found that in general, the SUS score decreases when the age of the user increases. Although the relationship is not very strong it is statistically significant ($p = .03$). This indicates that SUS scores may not be directly comparable between age groups because an older adult might consider a website as equally usable but actually rate the site lower than would a younger adult. Because of the need for usability studies with older adults, more research needs to be done on this topic.

Although the SUS was originally designed for computer systems in general, it can be modified slightly to apply directly to websites. This is done be simply replacing the word *system* with *website* in each item. Bangor et al.'s (2008) article pointed out that there is not sufficient data to suggest that performance metrics are related to SUS scores. One of the hypotheses for this research was to test the relationship between user's SUS score and task performance.

GOMS is an analytical approach to evaluating human–computer interaction. The technique was developed in 1983 as a means to break down complex tasks into separate subtasks (Card, Moran, & Newell, 1983a). Goals are what the user wants to achieve through computer interaction, and they are explained in the simplest terms as an overall purpose of the computer system. Goals can be broken down further to smaller subgoals, which can be broken down even further until an appropriate stopping point is reached. Operators are the actions that are performed in order to reach the goals. Examples of operators include menu selections and button presses. Methods describe how the users use the operators and subgoals to accomplish a given task. Typically there is more than one set of methods available to choose from. In this case, the selection rules are utilized. Selection rules are the process of choosing which method to use to accomplish the task (John & Kieras, 1996). The GOMS method is typically used to predict how long it will take users to perform certain tasks on a computer by finding the times it takes to complete each individual subgoal and action and summing them together to get a final time. The drawbacks of GOMS is that it does not take into account any types of error or any outside influence, such as fatigue or surroundings (Stanton, Salmon, Walker, Baber, & Jenkins, 2005). GOMS provides an accurate task description in optimal settings only.

KLM is a variation of the GOMS method. This technique is similar to GOMS, where the goals are broken down into smaller and smaller subgoals to find the total time it takes to accomplish a task (Card, Moran, & Newell, 1983b). The subgoals are accomplished by subtasks, which are assigned to a standard time. For example, the standard time for a typist to press a keystroke is 0.12 s, to point a mouse is 1.1 s, to click a button is 0.1 s, and to move your hand from the mouse to the keyboard is 0.4 s (Kieras, 2001). The total time to complete a task is found by summing all the subtasks. Once again, this approach is useful in determining task completion time in optimal conditions but does not take into account external influences or human error.

Heuristic evaluation is a usability inspection method for evaluating human–computer interaction. This is done providing multiple desirable characteristics that a website should have (called heuristics) and evaluating how well or how poorly the website meets them (Nielsen, 1990). The guidelines need to be carefully worded so that they are not too vague to be able to correctly critique problem areas or too specific to apply to all websites being evaluated. In typical evaluations, three to five evaluators go through a checklist to see if the heuristics are met (Nielsen, 1994). This method is useful for quick and inexpensive evaluation; however, it is very subjective and open to the evaluator's interpretation of the guidelines. The method does not take into account the evaluator's computer experience and can therefore be biased (Wilson & Corlett, 2005).

The heuristic evaluation was expanded by Keevil (1998) to form a quantitative usability measurement technique. This technique used heuristics taken from the literature and phrased them in questions that could be answered by "yes" or "no" (Keevil, 1998). If the question could not be answered, it is marked "N/A." A usability score is then assigned based on the responses by dividing the number of "yes" responses by the sum of the "yes" and "no" responses. This produces a percentage that can be interpreted as a usability score from 0 to 100 (Keevil, 1998). This technique is useful, as it assigns a clear numerical value to the website in question. However, it does not differentially weight each heuristic. For example, the responses to the questions "Is the text is double spaced?" and "Is the text easily distinguished from the background?" contribute the same score to the final Usability Index, when, in reality, they may have different effects on a site's usability. Also, some of the questions are hard to answer in a simple "yes" or "no" fashion. For example, if a heuristic is present in some aspects of the site, but absent in others, is the evaluator to mark a "yes" or "no"? The method used in our research is an adaption on this evaluation technique, with additional attempts to correct the weaknesses in the model.

Other methods include the Walkthrough method and the SUMI. In the Walkthrough method, users are asked to perform specific tasks or to achieve specific goals using a computer interface. The users provide continuous feedback by speaking aloud saying exactly what they are doing, what they are looking at, what they are thinking, and what they like and dislike about the interface. This method is also a very quick and easy way to evaluate a website or other computer interface. It is good at finding glaring problems and common difficulties and misunderstandings faced by users; however, it is also very subjective to the specific users and the tasks they are required to attempt (Wilson & Corlett, 2005). SUMI is a user report technique aimed at evaluating the user satisfaction of a website or software program against a database of other rated sites. This is done by distributing a 50-item survey to at least 15 people and having them fill out and submit the forms after using the site. The survey is a series of statements, each with three possible responses: *agree, don't know,* or *disagree* (Baber, 2005). Once the surveys are completed, they are compiled and scored against the other websites that have previously been evaluated. This provides for a benchmark of how well the website stands out against other similar sites and is useful for highly competitive companies to gain an edge on their counterparts (Wilson & Corlett, 2005).

2.4. Usability Studies With Older Adults

Studies that have been done on website design and usability specifically for older adults are varied in their scope and applications (Becker, 2004; Chisnell & Redish, 2005; Czaja, Sharit, & Nair, 2008; Ellis & Kurniawan, 2000; Hart, Chaparro, & Halcomb, 2008; Nahm, Preece, Resnick, & Mills, 2004; Taveira & Choi, 2009). Chisnell and Redish (2005) discussed a method of evaluating websites by taking on different personas and applying the characteristics of those personas to the usability testing. The researchers constructed a set of heuristics from the literature and used the personas to evaluate multiple popular websites. The research helped to locate serious usability issues across the web's most used sites.

Other research has examined how aging affects mouse control in older users. Smith et al. (1999) tested 60 adults from three different age groups. The participants completed 20 tasks that involved pointing, clicking, double clicking, and dragging. The performance was measured by cursor movement time, movement distance, movement speed, and errors. The results of the experiment clearly showed that cursor control is much more difficult for older users (Smith et al., 1999).

Currently, the usability of health-related sites seems to be of much interest (Becker, 2004; Czaja et al., 2008; Nahm et al., 2004; Taha, Sharit, & Czaja, 2009). Studies have determined usability problems in sites specifically designed for older adults (Czaja et al., 2008). Another study used a heuristic evaluation and a simple usability test to evaluate three health-promoting websites for usability. The study employed a heuristic evaluation for the websites using four design experts. Next, 10 older adults were recruited to conduct a usability test and comment on their likes and dislikes of each site. The research revealed that the older users had issues with small font size, an overabundance of information, and lack of clear instructions (Nahm et al., 2004).

Another study aimed at the usability of health-related websites examined 125 different sites for how easy they were for seniors to use. In addition to evaluation based on the National Institute on Aging guidelines, the research also examined the reading complexity and translation capabilities of the sites for uneducated and non-English-speaking older adults. The research found the most common design issues in the websites were small text, drop-down menus, lack of adjustable text size, lengthy home pages, and lack of assistance. In addition, only 12% of the sites offered multilingual capabilities, and more than 30% required a college education in order to comprehend the content (Becker, 2004).

Hart, Chaparro, and Halcomb (2008) performed a study in which a panel of experts evaluated 40 websites using a heuristic evaluation. Each evaluator rated 25 heuristics on a 4-point scale of compliance from *never* to *always present*. The score was calculated by the percentage of heuristics that were rated 3 or 4. This score was then used to select three websites for evaluation. Next, a usability test was performed in which 21 older adults performed tasks on the three websites and were monitored for how long each task took, how many mouse clicks it took to complete the task, and task success rate. The results show that there is a drop in mean success rate as the usability score decreases.

Our research builds on the work of previous studies. It includes the heuristic approach and is indebted to the work of Nielsen and Hart, but it expands previous work by the addition of weighted factors. In addition, our heuristic was validated by performance measures that indicated that higher heuristic scores were indicative of higher performance measures. To our knowledge, no other research has developed a weighted heuristic based on older individuals' preferences and validated it with performance metrics.

3. METHODS

This research had two objectives: (a) to develop a weighted heuristic for older adults based on their preferences for website characteristics and (b) to test heuristics using performance metrics. Because the target user group is older adults, a statewide organization aimed at providing services to older adults and having twelve sites was chosen for the heuristic development.

3.1. Developing the Weighted Heuristic for Website Evaluation

Although there have been many models developed for evaluating website usability, there is a growing need for a model that quantitatively measures the usability of the site while giving a higher importance level to certain heuristics that rank highly by the user. The first objective of the research was to develop such a model.

The first step in the process of developing a new evaluation technique was to decide which guidelines to use as heuristics. After compiling items from multiple sources, a list of 32 desirable characteristics was created to reflect the most important senior-friendly recommendations. The characteristics were placed into the four categories: Readability, Navigation, Content/Organization, and Accessibility. These characteristics and categories were based on the NIH recommendations (NIA, 2002).

Next, a data collection effort was conducted so that weights could be assigned to each heuristic. It quickly became apparent that some guidelines would have a much larger effect on usability than others. For example, the fact that a website lacked double-spaced text is going to have a lesser effect on usability than a site lacking an internal search feature. To decide the value of the weights for each heuristic, a survey was conducted among 20 people older than 50 years of age. The survey was distributed online via e-mail; people participated under informed consent and were not paid for completion of the survey. The survey listed the 32 heuristics and asked the subjects to rate how important each item was for increasing web usability on a Likert scale from 1 (*not important*) to 5 (*very important*). Using a nonparametric statistical approach, the median response value was selected as the weight for the corresponding heuristic. This weight would be applied in the calculation of a final usability score for the website. A blank version of the weighted heuristic usability evaluation tool is shown in Figure 1 and the heuristic weightings are shown in Table 1.

Of interest, some of the most highly rated characteristics included "The text is easily distinguished from the background," "It is clearly defined where you are on the website," "There is a clear, easy-to-find menu with links to different areas of the website," "Hyperlinks are consistent are consistent and easy to identify," "There is an easy to use internal search feature on the website," "The site is clean and organized," and "There is a phone number or link provided for people who need help with the website." These characteristics can be linked with expected normal changes that could occur with the user. For instance, changing vision could account for the first comment and changing spatial abilities and/or fluid intelligence could account for the characteristics concerned with search and navigation. These results also concur with the literature in this area on preferences of older users (Hanson, 2010; Hart, Chaparro, & Halcomb, 2008; NIA, 2002; Wagner, Hassanein, & Head, 2010).

After the weights were assigned, each of the 12 selected sites was evaluated by two professionals. The heuristic produced percentage scores in the four categories (Readability, Navigation, Content/Organization, and Accessibility) as well as an overall score. The calculation of each score included the evaluator deciding whether the site followed the guideline, followed the guideline in some places but not others, or did not follow the guideline at all. This was referred to as a "presence score" and was marked 2, 1, or 0, respectively. The final score was calculated by multiplying the weight and the presence score of each heuristic and summing the product. This number was then divided by the maximum possible score to produce a percentage called the "Usability Index." These percentages were calculated for each category as well as overall.

The weighted heuristic usability evaluation was performed by two trained evaluators, and the average score was taken to create a final usability index. Out of the 12 websites evaluated, the usability scores ranged from 66.32% to 89.12%, with the median score being 77.59%. These are shown in Table 2.

3.2. Empirical Usability Testing

Participants. Thirty-one male and female individuals finished both parts of the study. Participants were male and female, were at least 50 years old, and had at least some knowledge of how to use a computer and the Internet. The subjects all participated under informed consent and were paid $50 for their participation ($20 was paid after completion of the survey, and $30 was paid upon completion of the usability test and SUS).

Location. All experiments were conducted in the Human Factors Laboratory at Ohio University. They were done on a Dell desktop computer. Usability metrics were measured with Morae software.

Procedure. Participants were required to complete two parts for the experiment: a demographic survey, and the usability experiment. These data were collected on separate days so that the information from the demographic data could be used to inform the choice of tasks for the second part of the experiment. The survey included questions about demographic characteristics (age, education level, computer usage, etc.) as well as for what tasks participants usually used computers and the Internet.

The second part involved subjects participating in an empirical usability test. We chose three sites (low usability, medium usability, and high usability) from the 12 available statewide sites. These sites will be identified as Website 1 (W1), Website 2

Heuristic Website Evaluation

Readability	Present?	Weight	Score
There is sufficient spacing between letters in each word			
The type size at least 12 point font			
The main text is easily adjustable			
The main text double-spaced			
The text easily distinguished from the background			
There an absence of automatically scrolling text			
Total Readability Score:			
Readability Usability Rating:			

Navigation	Present?	Weight	Score
It is clearly identified where you are on the website			
There is a standard layout for each page on the website			
There is a clear, easy to find menu with links to different areas of the website			
The buttons large and easily clickable			
Hyperlinks are consistent and easy to identify			
Hyperlinks show if they have been previously clicked			
The appearance of hyperlinks changes when the cursor is placed on them			
There is an easy to use internal search feature on the website			
There is no underlined text that isn't a hyperlink			
The site use static menus (not drop-down)			
The buttons are clearly labeled			
Total Navigation Score:			
Navigation Usability Rating:			

Content/Organization	Present?	Weight	Score
The heading text is larger than the main body text			
Headings accurately describe the information within the body			
Menu items correspond to typical user tasks and goals			
There is site map for the website			
The site is clean and organized (not cluttered or busy)			
There is a minimal amount of text (only necessary information)			
Sentences are short and straightforward			
Task-related objects are grouped together			
Frequently used links and topics above the fold (you don't have to scroll)			
There is an appropriate amount of whitespace			
Total Content/Organization Score:			
Content/Organization Usability Rating:			

Accessibility	Present?	Weight	Score
There are text equivalents for non-text elements			
Color contrasts are easily distinguishable for those with color deficiencies			
The site is free of technical terminology			
The text and links are a different color from background (not just different hue)			
There is a phone number or link provided for those that need help with the site			
Total Accessability Score:			
Accessability Usability Rating:			
Usability Index:			

FIG. 1. The heuristic evaluation.

(W2), and Website 3 (W3). The particular sites are not identified because some have redesigned their sites since the experiment was finished.

Five tasks were selected for each site (15 tasks total). The selected tasks were all very similar for the different websites, and some were even identical. An example of a typical question is “Find the page that displays information on the Residential State Supplement Program” or “Find the link to the website ‘AARP in Ohio.’” The participants were instructed to find the information that the task requested using only the website

TABLE 1
Weightings for Heuristics

Question	Mdn	Mode	Min	Max
There is sufficient spacing between letters in each word	3	3	1	4
Typeface is larger than 12 point	2	3	1	5
The “main text” size is easily adjustable on the website	3	3	1	4
The “lmain text” is double-spaced	1	1	1	4
The text is easily distinguished from the background	4	4	3	5
There is an absence of automatically scrolling text	3	3	1	5
It is clearly defined where you are on the website	4	4	1	5
There is a standard layout for each page on the website	2	3	1	4
There is a clear, easy to find menu with links to different areas of the website	4	4	3	5
Hyperlinks are consistent and easy to identify	4	4	3	5
There is no underlined text that isn’t a hyperlink (the only text that is underlined are the links)	3	3	1	4
Hyperlinks show that they have been previously clicked	3	3	1	4
The appearance of hyperlinks changes when the cursor is hovering over them	3	2	1	5
There is an easy-to-use internal search feature on the website	4	3	1	5
The site uses static menus (not drop-down or other complex menu structures)	1.5	1	1	4
The buttons are large and easily clickable	2.5	2	2	5
The buttons are clearly labeled	4	4	2	5
The heading text is larger than the main body text	3	3	1	5
Headings accurately describe the information within the body	3	3	3	5
Menu items correspond to typical user tasks and goals	3	3	2	5
There is a site map for the website	3	3	1	5
The site is clean and organized	4	4	2	5
There is a minimal amount of text on the page	3	3	1	4
Sentences are short and straightforward	3	4	1	5
Task related objects are grouped together	3	3	1	4
Frequently used links and topics are above the fold (you do not have to scroll down to see them)	3	3	1	5
There is an appropriate amount of white space	3	3	1	5
There are text equivalents for nontext elements	2.5	2	1	4
Color contrasts are easily distinguishable for people with color deficiencies (color blindness)	3	3	1	5
The site is free of technical terminology	2	2	1	5
The text and links are a different color from the background, not just a different hue of the same color	3	3	1	5
There is a phone number or link provided for people who need help with the website	4	4	1	5

presented to them. They were not allowed to use any external search engines. Before the experiment began, a window for each task was opened so that the participant didn’t have to actually type in the URL to get to the site. The tasks were randomized for each participant in order to eliminate any potential order effect.

Using a usability testing program called Morae, performance data were collected on the user’s ability to complete the 15 tasks. The software allows the researcher to observe what is happening on the subject’s monitor from a remote computer and log how long each task takes, how many errors the user makes, and how many mouse clicks it takes the user to complete each task. In this case, an error was considered to be any time the user clicked on a link that is not on the direct path to the desired information described in the task. To accurately measure time on task, the administrator started the clock after the task was read and the website was opened and stopped the clock when the mouse cursor was placed over the requested information. Task performance was measured by these three dependent variables (time on task, number of errors per task, and number of clicks per task). For analysis, these performance metrics were summed (e.g., sum total of all performance times). At the conclusion of

TABLE 2
Usability Scores for the Twelve Websites

Website	Readability	Navigation	Content	Accessibility	Total
Site 1	75.00%	83.57%	100.00%	82.76%	87.31%
Site 2	90.63%	52.14%	88.71%	100.00%	77.46%
Site 3*	84.38%	64.29%	79.03%	100.00%	77.72%
Site 4	81.25%	61.43%	85.48%	100.00%	78.24%
Site 5	62.50%	58.57%	67.74%	89.66%	66.84%
Site 6	75.00%	84.29%	90.32%	100.00%	87.05%
Site 7	81.25%	50.71%	85.48%	100.00%	74.35%
Site 8*	90.63%	91.43%	85.48%	89.66%	89.12%
Site 9*	68.75%	62.86%	53.23%	89.66%	66.32%
Site 10A	68.75%	65.71%	69.35%	100.00%	72.54%
Site 10B	84.38%	85.71%	95.16%	89.66%	89.12%
Site 11	50.00%	65.00%	83.87%	79.31%	70.73%

∗sites that were chosen

the usability testing, the subject was asked to complete a survey based on the SUS for each of the three websites they used in the experiment.

4. RESULTS

4.1. Website Evaluation

The websites were evaluated using the newly created weighted heuristic method as described in the Methods section. Two experienced evaluators independently evaluated each website. The scores were averaged between the two. The results from the evaluation of each site are shown in Table 2. In Table 2 a high usability index was 80% to 100%, a moderate usability index was 70% to 90%, and a low usability index was 0% to 70%. One website was selected from each usability level. Chosen websites are listed with an asterisk.

Most of the websites evaluated were given an overall score between 70% and 80%. To select the best possible websites from the 12 that were evaluated, each site was thoroughly explored to find similar features. This was necessary because there needed to be five consistent tasks across the three websites. All five tasks needed to be able to be completed on each site in roughly the same time frame (for an experienced user) and be about the same number of clicks away from the home page. From these results, the following websites were selected for use in the empirical testing: District 8 (good), District 3 (medium), District 9 (poor). These will be referred to as Websites 1, 2, and 3 (w1, w2, and w3), respectively.

4.2. Descriptive Statistics and Survey Results

Thirty-four participants were selected for part one of the experiment. Three participants did not return for or did not fully complete Part 2, leaving only 31 subjects to be considered in the usability testing. The average age was 66.94 years ($SD =$ 10.04 years), and the average amount of time spent per week on the Internet was 9.19 hr ($SD = 6.67$ hr).

The results of the survey showed much variation between subjects on their work status, computer usage, and health. Out of the 31 participants, only 13 responded that they were currently employed, with the other 18 being retired or unemployed. When asked if they had computer training, 17 subjects responded that they had at least some formal training. When asked to describe their general health, 17 responded “excellent,” eight responded “good,” five responded “average,” and one responded “poor.” Ten respondents said that they did have ailments that affected their computer usage, citing vision problems, hearing loss, and arthritis.

Participants reported that word processing and gaming were the most common activities performed on the computer while offline. The most common uses for the Internet among respondents were to obtain general information, news, and health information. Most participants responded that they get a majority of their health information from newspapers, friends, and doctors. Some respondents also stated that they got a lot of this information from online sources. This information was used to inform the choice of the questions for the heuristic testing.

4.3. Heuristic Testing

To determine whether websites that had been rated to have different usability according to the heuristic ratings actually performed differently, we conducted a repeated measures or within-subjects analysis of variance (ANOVA) for each performance metric (time, clicks, and errors) across the three websites. Because tasks were similar, performance metrics were summed for each of the three websites.

For task time, we first used Mauchly’s Sphericity Test and confirmed that the assumption of sphericity could not be rejected ($p = .354$). We then used repeated measures and

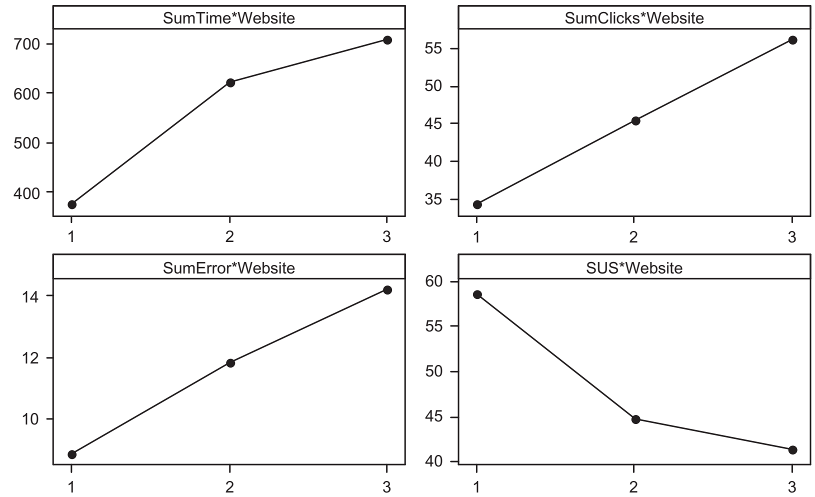

FIG. 2. Scatterplot of SumTime (sec), SumClicks, SumError, SUS (percentage) versus Website.

determined that a significant difference existed in task times between the three websites, $F(2) = 35.808$, $p = .000$. A pairwise comparison of the estimated marginal means showed that tasks times were significantly different with W1 being different from W2 and W3, whereas W2 and W3 were not significantly different (see Figure 2 and Table 3).

It can be seen in Figure 2 that there is a distinction between Website 1 and the others. If data were gathered from more persons, there may possibly have been more evidence to support a significant difference between Websites 2 and 3. Recall that Website 1 was assigned the highest usability index, whereas Website 3 was evaluated with the lowest. From these results, we can conclude that there is a significant difference between Website 1 and the others with respect to total time. This indicates that a high score on the weighted heuristic evaluation can predict a lower task completion time.

The repeated measures ANOVA was also completed with the total number of clicks for each website. Using Mauchly's Sphericity Test, it was confirmed that the assumption of sphericity could not be rejected ($p = .353$). The total numbers of clicks were significantly different among websites, $F(2) = 17.98$, $p = .000$. Pairwise analyses indicated that all three means were significantly different from one another (see Figure 2 and Table 3). Figure 2 indicates that a higher usability index leads to a lower number of clicks needed to accomplish the tasks for a website.

Finally, a third repeated measures ANOVA was used in the same way to test differences within total errors for each websites. Recall that an error is recorded whenever the subject clicks on a link that does not directly lead to the end result of the task. Because of this, it was expected that the results will be similar to total number of clicks. Using Mauchly's Sphericity Test, the assumption of sphericity was rejected ($p = .013$). Using the Greenhouse-Geisser adjustment (to compensate for the lack of sphericity), total errors were significantly different between websites, $F(2) = 7.208$, $p = .002$. Pairwise comparison for the estimated marginal means indicated that mean number of errors was different between W1 and the others but that W2 and W3 were not significantly different (see Figure 2 and Table 3).

4.4. Relationship Between SUS and Task Performance

SUS was analyzed using a repeated measures generalized linear model analysis to determine if mean SUS score was different between websites. Using Mauchly's Sphericity Test, it was confirmed that sphericity could not be rejected ($p = .131$). Results indicated that mean SUS score was different between websites, $F(2) = 35.808$, $p = .000$. Pairwise analysis indicated that W1 was significantly different from W2 and W3 but that W2 and W3 were not statistically significantly different (see Figure 2 and Table 3).

These results demonstrate that the participants agree with the evaluators' assessment of the websites' usability with regards to W1 being more usable than the other two. A correlational analysis was then done to determine if SUS was correlated with performance metrics and is shown in Table 4.

To determine if SUS could predict task times, regression analyses were run for each website. The dependent variable was total task time, and the independent variables were number of clicks, number of errors, and SUS. The regression analysis contained two blocks. The first block used only number of clicks and number of errors as independent variables. The second block then added SUS as the third independent variable. This was done to measure the effect of adding SUS as a predictor of task time above and beyond the contribution of clicks and errors. It is known that clicks and errors are related to total task time; however, the more difficult challenge for the SUS measure is whether it is able to predict variation in time beyond these. The results of the regression analysis for Website 1 are shown in Tables 5 and 6.

TABLE 3
Pairwise Differences for Dependent Variables

(I) Website	(J) Website	*M* Diff (I-J)	Std. Error	Sig. a	95% CI for Differences Lower Bound	Upper Bound
Measure: Total task time						
1	2	−247.798*	38.047	0.000	−344.275	−151.32
	3	−335.834*	38.666	0.000	−433.88	−237.788
2	1	247.798*	38.047	0.000	151.32	344.275
	3	−88.036	46.242	0.200	−205.294	29.221
3	1	335.834*	38.666	0.000	237.788	433.88
	2	88.036	46.242	0.200	−29.221	205.294
Measure: Total clicks						
1	2	−11.355*	4.155	0.031	−21.892	−0.818
	3	−22.226*	3.363	0.000	−30.752	−13.699
2	1	11.355*	4.155	0.031	0.818	21.892
	3	−10.871*	3.553	0.014	−19.881	−1.861
3	1	22.226*	3.363	0.000	13.699	30.752
	2	10.871*	3.553	0.014	1.861	19.881
Measure: Total errors						
1	2	−3.032*	1.184	0.047	−6.034	−0.031
	3	−5.452*	1.762	0.013	−9.92	−0.983
2	1	3.032*	1.184	0.047	0.031	6.034
	3	−2.419	1.306	0.221	−5.73	0.892
3	1	5.452*	1.762	0.013	0.983	9.92
	2	2.419	1.306	0.221	−0.892	5.73
Measure: SUS						
1	2	13.871*	5.228	0.038	0.613	27.129
	3	17.339*	4.973	0.005	4.729	29.949
2	1	−13.871*	5.228	0.038	−27.129	−0.613
	3	3.468	3.776	1.000	−6.107	13.042
3	1	−17.339*	4.973	0.005	−29.949	−4.729
	2	−3.468	3.776	1.000	−13.042	6.107

Note. Based on estimated marginal means. SUS = System Usability Scale.
[a]Adjustment for multiple comparisons: Bonferroni.
*The mean difference is significant at the .05 level.

The results of the regression analysis show that total errors is a significant predictor of total time for Website 1 and that SUS significantly predicts time above and beyond errors. The variable "total clicks" was not significant. The overall model was found to be significant ($p = .002$). There was a .095 change in the R-square value after adding the SUS, showing that it significantly contributes to the dependent variable, total task time after variation in time due to clicks and errors has been removed. The coefficients for this regression show the magnitude and influence on the total time dependent variable. In this regression, total clicks and total errors have a positive correlation with total time, although the number of total errors has a much larger effect. SUS has a negative impact on total time, showing that a higher SUS score results in lower total task time.

This procedure was conducted again for Website 2. The results are displayed in Tables 5 and 6. This regression analysis shows that, for Website 2, SUS is not a significant predictor of total task time when clicks and errors have been controlled, whereas total clicks and total errors are both significant predictors. Similar conclusions can be drawn from the coefficient values. Higher counts of errors and clicks, and lower SUS scores are predictive of longer task times. The overall regression model was significant ($p < .001$).

The analysis was done once more for Website 3. The results are shown in Tables 5 and 6 and indicate that SUS is not a significant predictor of total task time. The only significant predictor variable in this regression is the total clicks. Once again, the coefficients draw similar conclusions

TABLE 4
Pearson Correlation Coefficients for SUS With Performance Metrics

		Sum Time W1	Sum Click W1	Sum Err W1
SUS Web 1	Pearson correlation	−.625**	−.564**	−.437*
	Sig. (two-tailed)	0.000	0.001	0.014
	N	31	31	31
		Sum Time W2	Sum Click W2	Sum Err W2
SUS Web 2	Pearson correlation	−0.286	−0.076	−0.137
	Sig. (two-tailed)	0.119	0.686	0.461
	N	31	31	31
		Sum Time W3	Sum Click W3	Sum Err W3
SUS Web 3	Pearson correlation	−0.316	−0.314	−0.208
	Sig. (two-tailed)	0.083	0.086	0.262
	N	31	31	31

Note. SUS = System Usability Scale.

TABLE 5
Regression Model Significance

					Change Statistics				
Model	R	R^2	Adjusted R^2	SE of the Estimate	R^2 Change	F Change	*df*1	*df*2	Sig. F Change;
Website 1									
1	.738[a]	0.544	0.512	120.485	0.544	16.726	2	28	0.000
2	.800[b]	0.639	0.599	109.172	0.095	7.104	1	27	0.013
Website 2									
1	.783[a]	0.614	0.586	154.143	0.614	22.241	2	28	0.000
2	.808[b]	0.653	0.614	148.789	0.039	3.051	1	27	0.092
Website 3									
1	.822[a]	0.675	0.652	147.191	0.675	29.059	2	28	0.000
2	.824[b]	0.679	0.643	149.038	0.004	0.31	1	27	0.582

Note. 1 = Predictors: (Constant), Sum Error, Sum Click; 2 = Predictors: (Constant), Sum Error, Sum Click, SUS; a = Predictors in 1; b = Predictors in 2.

as the other regression analyses. In this model, however, the total number of clicks has a much higher weight than the total number of errors. This regression model is significant ($p < .001$).

The results of these three regression analyses are very interesting. It appears that SUS is a significant additional predictor of task performance only when dealing with a highly usable website. It is also noticeable that the SUS scores were, on average, much lower than previous research studies. Even the most usable website had an average SUS score of 58.63, which is well below the recommended benchmark score of 70. As suggested by Bangor et al. (2008), this may be due to the fact that our participants were all in an older age group.

5. DISCUSSION AND CONCLUSION

Although a moderate amount research in human factors engineering, psychology, and other related fields has been published on website usability for older adults, much more work needs to be done in order to promote technology adoption (Becker, 2004; Bucur & Kwon, 1999; Hanson, 2010; Hart, Chaparro, & Halcomb, 2008; Ji et al., 2010; Taveira & Choi, 2009). Some of the most popular guidelines are from AARP, the National Institute on Aging, W3C, and Nielsen's extensive website usability research (Chisnell & Redish, 2005; NIA, 2002; Nielsen, 1990). In addition, much work has been done on how to conduct heuristic evaluations and other usability tests for all age groups of users (Alonso-Ros, Vázquez-García,

TABLE 6
Regression Coefficients for Sum Time as Dependent Variable for Each Website

Model		Unstandardized Coefficients B	SE	Standardized Coefficients β	t	Sig.
Website 1						
1	(Constant)	222.855	39.298		5.671	0.000
	SumClickW1	1.766	1.46	0.237	1.21	0.237
	SumErrW1	10.177	3.677	0.542	2.767	0.01
2	(Constant)	441.664	89.487		4.936	0.000
	SumClickW1	0.242	1.441	0.032	0.168	0.868
	SumErrW1	10.027	3.332	0.534	3.009	0.006
	SUS_W1	−2.825	1.06	−0.373	−2.665	0.013
Website 2						
1	(Constant)	333.654	51.431		6.487	0.000
	SumClickW2	3.626	1.218	0.462	2.978	0.006
	SumErrW2	10.316	4.02	0.399	2.566	0.016
2	(Constant)	443.764	80.237		5.531	0.000
	SumClickW2	3.666	1.176	0.467	3.118	0.004
	SumErrW2	9.519	3.907	0.368	2.437	0.022
	SUS_W2	−2.289	1.311	−0.2	−1.747	0.092
Website 3						
1	(Constant)	241.195	66.795		3.611	0.001
	SumClickW3	7.982	1.343	0.792	5.943	0.000
	SumErrW3	1.272	3.505	0.048	0.363	0.719
2	(Constant)	283.014	101.035		2.801	0.009
	SumClickW3	7.793	1.402	0.773	5.56	0.000
	SumErrW3	1.212	3.55	0.046	0.341	0.735
	SUS_W3	−0.735	1.318	−0.064	−0.557	0.582

Mosqueira-Rey, & Moret-Bonillo, 2010; Lamminen, Leppnen, Heikkinen, Kmrinen, & Jokisuu, 2009; MacDorman, Whalen, Ho, & Patel, 2011). Despite this, a gap in the literature exists for work that has tested heuristics against performance metrics and quantitative usability studies using older adults. This usability index is the first of its kind to assign a researched weight and presence score to each heuristic and use them to quantitatively rank the usability of websites using performance metrics.

This research has (a) established a method of quantitatively evaluating a website's usability for older users and (b) provided insight into conducting usability tests on various websites designed for older users. The results indicated that the heuristic was predictive in categorizing websites so that performance metrics could be demonstrated to be significantly different between websites. Repeated measures tests showed that W1 had significantly different total task times and total errors than W2 and W3. It was also shown that all three websites had significantly different total clicks. A decrease in a website's assigned usability index corresponded to a lower level of performance, as measured by total task completion time, total number of mouse clicks, and total number of errors.

The relationship between SUS scores and performance was also analyzed. Using repeated measures ANOVA, mean SUS scores were significantly different between websites, with W1 being considered to be more usable than the other two websites. However, SUS was not equally predictive of task time between different websites. Using regression analysis, it was shown that SUS corresponds with total task time above and beyond Clicks and Errors for W1; however, SUS proved to be an insignificant predictor for total task time for Website 2 and Website 3. This demonstrates that SUS was predictive of task time performance only for the website with the highest usability rating. This was an unusual finding, and one that was quite unexpected. One would think that if a participant performed poorly on the given tasks for a particular website, that he or she would perceive it as having poor usability; however, that was not necessarily the case.

We believe that the research point to elements that could be used by website designers. First, the weights for the heuristics (see Table 1) are indicative of which usability characteristics are more or less important to an older individual. Those characteristics that rank high should be present in a website designed for older adults (e.g., easily distinguished text or an easy to find menu). In addition, the weightings also indicate which characteristics are not of greatest importance to older individuals (e.g., having larger than 12 point or having double-spaced text). Of interest, many of the highly ranking elements correspond to navigation. It is common knowledge that spatial abilities decline as age increases, so the fact that our participants rated this highly is comparable to what the literature recommends. It was also interesting to us that our participants still wanted to have a phone number on the web page so that they could call (and presumably speak with someone) who could help them with additional information.

6. FUTURE RESEARCH

As older adults continue to access the web for communication, business, and health reasons, it is imperative that designers have validated tools to help them design usable websites. This research can serve as a starting point for diverse future research projects. Because this was the first study using the weighted heuristic evaluation method, many improvements could be made to the scale. Future studies could be performed on how to better enhance the validity of the model as well as to determine if it could be adapted to work with younger users. It would also be interesting to know which heuristic scores have the biggest effect on website performance.

The research opens up an opportunity for more research in the area of the relationship between SUS and task performance on websites of various usability levels. It appears from this analysis that the less usable a website is, the less accurate its SUS score becomes, which, needless to say, is quite a paradox. It may be that some usability scales have differing degrees of sensitivity and that caution should be used when applying them in situations where a large range in usability characteristics exists. It would be interesting to determine which usability features may interact with others to influence the varying levels of sensitivity provided for by a particular evaluation method.

Further research could also be performed to see the relationship between health issues and web performance. A good project would be to take seniors with various medical conditions and monitor their web performance across sites with varying degrees of usability. It would be interesting to discover which health problems had the largest impact on user performance, and what different website designs could assist these users.

An interesting behavior was observed while running participants through the usability testing. A few of the subjects became extremely frustrated during the course of the testing. They would get very flustered if they could not find the answer quickly and would start to swear or get angry with the test administrator. Future research could look into what causes users to get frustrated while online.

Finally, research could be done on the usability of other systems for seniors using a similar evaluation method. Computer interfaces are becoming more and more common in household appliances, automobiles, and cell phones, just to name a few. With touch screens becoming more widespread across a variety of electronic devices, it would be interesting to research how the heuristics developed for websites can cross over on to other platforms in an attempt to increase the usability and performance on such machine.

REFERENCES

Administration on Aging. (2010). A profile of Older Americans: 2008. Retrieved from http://www.aoa.gov/aoaroot/aging_statistics/Profile/2010/docs/2010profile.pdf

Alonso-Ros, D., Vázquez-García, A., Mosqueira-Rey, E., & Moret-Bonillo, V. (2010). A context-of-use taxonomy for usability studies. *International Journal of Human–Computer Interaction*, *26*, 941–970.

Arch, A., & Abou-Zhara, S. (2004). *How web accessibility guidelines apply to design for the ageing population*. World Wide Web Consortium Sophia Antipolis cedex, France.

Baber, C. (2005). Evaluation in human–computer interaction. In J. R. Wilson & E. N. Corlett, *Evaluation of human work* (pp. 357–388). Boca Raton, FL: Taylor & Francis.

Bangor, A., Kortum, P., & Miller, J. (2008). An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction*, *24*, 574–594.

Becker, S. A. (2004). A study of web usability for older adults seeking online health resources. *ACM Transactions on Computer–Human Interaction*, *11*, 387–406.

Brooke, J. (1996). SUS: a 'quick and dirty' usability scale. In P. W. Jordan (Ed.), *Usability evaluation in industry* (pp. 189–191). Bristol, PA: Taylor & Francis.

Bucur, A., & Kwon, S. (1999). Computer hardware and software interfaces: Why the elderly are under-represented as computer users. *CyberPsychology & Behavior*, *2*, 535–543.

Card, S. K., Moran, T. P., & Newell, A. (1983a). The GOMS model of manuscript editing. In S. K. Card, T. P. Moran, & A. Newell (Eds.), *The psychology of human–computer interaction* (pp. 139–192). Hillsdale, NJ: Erlbaum.

Card, S. K., Moran, T. P., & Newell, A. (1983b). The Keystroke-level model. In S. K. Card, T. P. Moran, & A. Newell (Eds.), *The psychology of human computer interaction* (pp. 256–313). Hillsdale, NJ: Erlbaum.

Chaparro, A., Bohan, M., Fernandez, J., Choi, S., & Kattel, B. (1999). The impact of age on computer input device use. *International Journal of Industrial Ergonomics*, *24*, 503–513.

Chiou, W., Lin, C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995–2006. *Information & Management*, *47*, 282–290.

Chisholm, W., Vanderheiden, G., & Jacobs, I. (2001). Web content accessibility guidelines 1.0. *Interactions*, pp. 35–54.

Chisnell, D., & Redish, J. (2005). Designing websites for older adults: Expert review of usability for older adults at 50 web sites. *AARP*, pp. 1–60.

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., & Rodger, W. (2006). Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (CREATE). *Psychology of Aging*, *21*, 333–352.

Czaja, S. J., & Lee, C. C. (2008). Information technology and older adults. In A. Sears & J. A. Jacko (Eds.), *The human–computer interaction handbook—Fundamentals, evolving technologies and emerging applications* (pp. 777–789). New York, NY: Erlbaum.

Czaja, S. J., Sharit, J., & Nair, S. N. (2008). Usability of the Medicare health web site. *Journal of American Medical Association*, *300*, 790–792.

Czaja, S. J., Sharit, J., Hernandez, M. A., Nair, S. N., & Loewenstein, D. (2010). Variability among older adults in Internet health information-seeking performance. *Gerontechnology*, *9*, 46–55.

Dugale, D. (2010). *Age changes in the senses.* Retrieved from http://www.nlm.nih.gov/medlineplus/ency/article/004013.htm

Echt, K. V., Morrell, R. W., & Park, D. C. (1998). Effects of age and training formats on basic computer skill acquisition in older adults. *Educational Gerontology*, *24*, 3–25.

Ellis, R. D., & Kurniawan, S. H. (2000). Increasing the usability of online information for older users: A case study in participatory design. *International Journal of Human–Computer Interaction*, *12*, 263–276.

Fidgeon, T. (2006, February). *Usability for older web users*. Retrieved from http://www.webcredible.co.uk/user-friendly-resources/web-usability/older-users.shtml

Hanninen, T., Koivisto, K., Reinikainen, K., Helkala, E.-L., Soininen, H., Mykkanen, L., . . . Laakso, M. (1996). Prevalence of ageing-associated cognitive decline in an elderly population. *Age and Aging*, *25*, 201–205.

Hanson, V. (2009). Age and web access: The next generation. In *International Cross-Disciplinary Conference on Web Accessibility* (pp. 7–15). New York, NY: ACM.

Hanson, V. (2010). Influencing technology adoption by older adults. *Interacting with Computers*, *22*, 502–509.

Hart, T. A., Chaparro, B. S., & Halcomb, C. G. (2008). Evaluating websites for older adults: adherence to "senior-friendly" guideline and end-user performance. *Behaviour & Information Technology*, *27*, 191–199.

Ji, Y. G., Choi, J., Lee, J. Y., Han, K. H., Kim, J., & Lee, I. (2010). Older adults in an aging society and social computing: A research agenda. *International Journal of Human–Computer Interaction*, *26*, 1122–1146.

John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique? *Association for Computing Machinery*, pp. 1–30.

Keevil, B. (1998). Measuring the usability index of your Web site. *Proceedings of the ComputerHumanInteraction CHI'98 Conference*, 271–277.

Kieras, D. (2001). Using the keystroke-level model to estimate execution times. Retrieved from University of Michigan website: http://ftp://ftp.eecs.umich.edu/people/kieras/GOMS/KLM.pdf

Kurniawan, S., & Zaphiris, P. (2005). Research-derived web design guidelines for older people. *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, 129–135.

Lamminen, J., Leppnen, M., Heikkinen, R., Kmrinen, A., & Jokisuu, E. (2009). A quantitative method for localizing user interface problems: The D-TEO method. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments*, *5*, 121–145.

MacDorman, K. F., Whalen, T. J., Ho, C., & Patel, H. (2011). An improved usability measure based on novice and expert performance. *International Journal of Human–Computer Interaction*, *27*, 280–302.

Mead, S. E., Lamson, N., & Rogers, W. A. (2002). Human factors guidelines for web site usability: Health-oriented web sites for older adults. In R. W. Morrell (Ed.), *Older adults, health information, and the World Wide Web* (pp. 89–97). Mahwah, NJ: Erlbaum.

Morrell, R. W. (2002). Human factors guidelines for web site usability: Health-oriented web sites for older adults. In R. W. Morrell (Ed.), *Older adults, health information, and the World Wide Web* (pp. 89–97). Mahwah, NJ: Erlbaum.

Nahm, E.-S. P., Preece, J. P., Resnick, B. P., & Mills, M. E. (2004). Usability of health web sites for older adults: A preliminary study. *CIN: Computers, Informatics, Nursing*, *22*, 326–334.

Namatame, M., Kobayashi, M., & Harada, A. (2003). A case study for reaching web accessibility guidelines for the hearing-impaired. *PsychNology Journal*, *1*, 400–409.

National Institutes of Health. (1997). Presbycusis. Retrieved from http://www.nidcd.nih.gov/health/hearing/presbycusis.asp

National Institute on Aging & National Library of Medicine. (2002). Making your web site senior friendly. Retrieved from http://www.nlm.nih.gov/pubs/checklist.pdf

Nielsen, J. (1990). Heuristic evaluation of user interfaces. *Conference on Human Factors in Computing Systems*, 249–256.

Nielsen, J. (1994). Usability inspection methods. *Conference on Human Factors in Computing Systems*, 413–414.

Nielsen, J. (1997). User interface design for the WWW. *Conference on Human Factors in Computing Systems*, 140–141.

Rainie, L. (2010, March 19). Baby boomers in the digital age. Retrieved from http://www.pewinternet.org/Presentations/2010/Mar/Boomers.aspx

Smith, M., Sharit, J., & Czaja, S. (1999). Aging, motor control, and the performance of computer mouse tasks. *Human Factors*, *41*, 389–396.

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods: A practical guide for engineering and design*. Burlington, MA: Ashgate.

Taha, J., Sharit, J., & Czaja, S. (2009). Use of and satisfaction with sources of health information among older internet users and nonusers. *The Gerontologist*, *49*, 663–673.

Taveira, A. D., & Choi, S. D. (2009). Review study of computer input devices and older users. *International Journal of Human–Computer Interaction*, *25*, 455–474.

Wagner, N., Hassanein, K., & Head, M. (2010). Computer use by older adults: A multi-disciplinary review. *Computers in Human Behavior*, *26*, 870–882.

Wilson, J. R., & Corlett, E. N. (2005). *Evaluation of human work*. London, UK: Taylor & Francis.

Yan, J. H., Rountree, S., Massman, P., Smith Doody, R., & Hong, L. (2008). Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. *Journal of Psychiatric Research*, *42*, 1203–1212.

Zaphiris, P., Ghiawadwala, M., & Mughal, S. (2005). Age-centered research-based web design guidelines. *Conference on Human Factors in Computing Systems*, 1897–1900.

ABOUT THE AUTHORS

Kyle R. Lynch earned his B.S. and M.S. in Industrial and Systems Engineering from the Russ College of Engineering and Technology at Ohio University. Mr. Lynch currently serves as an industrial engineer for Owens-Illinois.

Diana J. Schwerha is an Associate Professor in the Department of Industrial and Systems Engineering in the Russ College of Engineering and Technology at Ohio University. Dr. Schwerha's research interests focus on ergonomics and the aging worker, usability of technology for older users, and ergonomics and employee retention.

George A. Johanson is an Emeritus Professor of Educational Research and Evaluation in the Patton College of Ohio University. His research interests include methods of survey research, differential item functioning, and differential person functioning. Dr. Johanson received his Ed.D. from the University of Massachusetts, Amherst.