Evaluation of Remote Usability Techniques in an Elderly Cohort

by

Chia-Hua Yu

A Capstone

Presented to the Oregon Health & Science University's
Department of Medical Informatics and Clinical Epidemiology

in partial fulfillment of
the requirements for the degree of
Master of Biomedical Informatics
August, 2010

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's Capstone Project of

Chia-Hua Yu

"Evaluation of Remote Usability Techniques in an Elderly Cohort"

Has been approved

Holly B. Jimison, PhD
Capstone Advisor

Contents

List of Figures

Abstract

Background The rising cost of health care has motivated the exploration of different models of care. One such model is the health coaching paradigm. Focusing on health maintenance beavhiors, health coaches may provide more attentive and timely help at lower cost than traditional primary care models. Health coaches can be supported by internet technologies, allowing the coach to communicate with more participants than would be possible with in-person appointments. The Oregon Center for Aging and Technology Living Lab combines the health coach approach with information technology to deliver timely coaching interventions among an elderly population to promote cognitive health. The current study examines the usability of a client-facing interface using remote usability methods. An additional goal is to examine the feasibility of remote usability methods among an elderly population as well as to evaluate the usability of the interface.

Methods This study utilizes the think aloud protocol, a discount usability technique. Participants were recruited from the Living Lab Participants pool. Participants were contacted via the Skype internet telephone application, and then transferred to a web conference application for the usability testing. Participants were guided through five tasks, during which they were asked to verbalize their thoughts while using the system. The usability test session was recorded on computer using screen capture software and analyzed after the test session.

Results The users had considerable difficulty completing the usability tasks, implying usability issues. In addition, the think aloud protocol proved to be a challenging task. The combination of a new computer interface and a novel task may have hindered task completion.

Discussion The participants exhibited a tentative interaction style, as several subjects expressed some discomfort in using computers. This computer-related anxiety may have affected their ability to complete tasks, as some degree of exploratory behavior is required to complete the tasks. Contributing to the lack of task completion is the visual complexity of some of the screens. Several design enhancements are suggested to increase the usability of the system.

Acknowledgements

I am indebted to my advisor Dr. Holly Jimison who has provided insight and guidance. I would also like to give my sincere thanks to Bill Hatt and Michael Chan for their excellent technical skills and and good craic. I would also like to thank Andi Lushaj for testing many remote usability tools prior to the start of the project, and Nicole Larimer for assistance with participant recruitment. This project would not have been possible without the support of my parents. Finally, I reserve special thanks to Lacey Young, who has provided me with a boundless source of positivity and lots of patience.

To my parents

Chapter 1

Introduction

1.1 Aging and the cost of health care

The population of the United States is aging. The Institute of Medicine (IOM) estimates that the population of adults over 65 years old will double between 2005 and 2030 (The Institute of Medicine, 2008). The Institute estimates that by 2030, this cohort will account for almost 20% of the US population. Moreover, the population of people over the age of 80 is expected to double by this time. This increase in the average age is not limited to the United States. The National Institutes on Aging (NIA) predicts a 140% increase in the population of people aged 65 and over among developing countries (The National Institute on Aging, 2007). The NIA raises a number of concerns associated with this trend; among them is the effect of population aging on health care delivery systems.

A key concern is that a greater elderly population will result in increasing health care expenditures. As Gray (2005) pointed out, a number of studies report that older segments of the population spend more on health care relative to younger cohorts. A congressional report found that on average, people over 65 spent four times the amount on health care relative to those under 65 (Jenson, 2007). The same report showed that the expenditures increase with age even in the 65 and over cohort. Given that health care costs increase as one ages and that the proportion of elders is increasing world wide, this led some researchers to conclude that the demographic shift will effect sizable increases in health care spending. For instance, Alemayehu and Warner took cross-sectional data insurance claim data for a single year and modeled longitudinal data for a hypothetical patient (Alemayehu and Warner, 2004). Alemayehu and Warner found that 60% of an individual's lifetime health care expenditures are spent after the age of 65. These results echo the findings of an earlier study using data from Medicare and the National Medical Expenditure Survey (Spillman and Lubitz, 2000). These researchers found that medical expenditures increased significantly after the age of 65. While Spillman and Lubitz found that acute care expenses decreased, long-term care costs (such as nursing home costs) rose to the point of off-setting any reduction in acute care expenses. Reports such as these raise concerns that the aging of the population will cause a large overall increase in health care spending, outstripping the ability for active workers to cover those costs. The association between a growing elderly population and the resulting increase in health care costs has gained enough traction such that it has become "common wisdom" (Chernichovsky and Markowitz, 2004, p. 543).

In response, a number of studies provide evidence that aging is not the sole reason, nor a significant reason, for the increases in cost. Using longitudinal health care data from Switzerland, Zweifel and colleagues showed that age is not a significant predictor of per-capita health care expenditure (1999). They found that independent of age, an individual's proximity to death was more indicative of increases in health care expenditure, regardless of whether an individual was 65 or 90 years old. Thus, controlling for remaining life span eliminated the effect of age on health care cost. A report by Gray corroborates these results. Reviewing subsequent analyses using different data, several of the studies he cites suggest the importance of time to death relative to the age of the individual (2005). He notes, "while time to death is a better predictor of health expenditure than age, it is in turn a crude measure of health status" (p. 18). Reinhardt (2003) refutes the use of cross-sectional data to divine the outcome of a cohort. Reviewing literature that draws on U.S., Canadian, and Australian data, the aging of the population is not found to account for more than 1% of the increase in health care spending.

Reinhardt concludes by commenting on the impact of the supply side of health care on expenditures (2003). He notes that another possible effect of an aging population is the decrease in the proportion of workers relative to retirees. He posits that costs could increase, given a looming labor shortage. Reinhardt suggests the use of health information technology (HIT) solutions as one way to overcome the labor shortage.

Both the IOM and NIA reports enumerate changes in health care needs associated with an aging population, including increases in chronic disease prevalence and long term care utilization (The Institute of Medicine, 2008; The National Institute on Aging, 2007). Despite comprising only 12% of the population, older adults account for a quarter of all physician office visits, more than a third of all emergency service responses and more than a third of all prescriptions. The high utilization of the current elderly population, coupled with the demographic shift, is predicted to cause increases in expenditure as outlined above, and also to stress the health care system if changes are not made in response to the growth of the elderly population.

The authors of the IOM report, *Retooling for an Aging America: Building the Health Care Workforce*, argue that the current care model is insufficient to meet the needs of the elderly: ”Current models of care delivery often fail to provide the best care possible to older adults, and they often do not promote the most efficient use of existing workers” (2008, p. 6). The authors also point to a decline in the number physicians specializing in geriatrics. This problem is compounded by the lower than average compensation received by the physicians, a result of a high Medicare-dependent case mix. Thus the labor shortage results from both decreased entrants into the geriatric field, as well as rapid growth in the elderly population.

Addressing this potential shortage, the IOM advocates changes to the current model of geriatric care. The changes proposed adhere to three principles:

- Addressing health needs comprehensively
- Providing services efficiently
- Engaging elderly patients in their own care (The Institute of Medicine, 2008)

These principles suggest a shift in the way elderly receive medical care. This new model focuses on continuity of care, giving the patient and informal care givers a more active role, as well as involving more diverse members of the health care team. These other caregivers may include nurses, case managers, certified nursing assistants, and a new category of caregiver, the health coach, described below.

1.2 Health Coaching

The third principle listed above, engaging elderly patients in their own care, recapitulates a point the IOM made in an earlier report, *Crossing the Quality Chasm* (2001). This approach is particularly suited for addressing the health concerns of the elderly who may be affected by a host of chronic conditions. Bandura notes, "Biomedical approaches are ill-suited for chronic diseases because they are devised mainly for acute illness. The treatment of chronic disease must focus on self-management of physical conditions over time" (Bandura, 2004, p. 159). Advocating changes to the primary care model, Bodenheimer observes, "people with chronic conditions are their own principal caregivers, and health care professionals—both in primary and specialty care—should be consultants supporting them in this role" (Bodenheimer et al., 2002). One key means of engaging the patient is encouraging him to modify behaviors to promote health. According the Transtheoretical model, patients can undergo several "stages of change" that indicate whether the patient is ready to receive educational and behavioral interventions (Prochaska and Velicer, 1997). If the patient is ready to make a change, collaborative goal setting can effectively help the patient make changes to his habits; if the intervention is delivered at the wrong stage (e.g., the patient may not realize he has a problem, or has already taken steps to change his habits), then the intervention will have limited efficacy. The collaborative nature of goal setting stands in contrast to the typical situation of physicians setting the goals for the patient. This model dictates that the patient and caregiver collaborate on formulating health goals, embodied in action plans.

Action plans are a vehicle for collaborative goal setting between a physician and the patient. Bodenheimer and his colleagues make the distinction between goal setting and action plans (Bodenheimer and Handley, 2009). Goal setting in this framework refers to delineating desirable habits or practices to promote beneficial health states, or to control a chronic condition (e.g. losing 10 pounds, exercising more, or quitting smoking). Bodenheimer uses the term action plans to refer to more specific and concrete goals and patterns of behavior—for example, in lieu of "exercising more", the patient would be told to walk for 10 minutes up and down the block. The rationale for this distinction stems from research on goal setting and human performance in other disciplines (Strecher et al., 1995). Locke and Latham reviewed a number of studies applying goal-setting and behavior change in business settings. Among different factors found to affect job performance, they found that specific goals were more effective in eliciting high performance relative to vague and general goals, or being told to "do one's best" (Locke and Latham, 2002). Goal setting helps to focus the participant's attention and efforts. Strecher and colleagues extend this research to the health care domain, offering guidance to health care practitioners.

Other investigators have endeavored to apply action plans in the management of chronic illness. MacGregor and her colleagues examined the feasibility of action-plans and goal setting for coronary heart disease patients in the context of a routine physician visit (MacGregor et al., 2005, 2006; Handley et al., 2006). The investigators found that most patients were amenable to formulating action plans, estimating that 53% of all enrollees had acted on their action plan during a 3 week follow-up period. Having been trained in setting action plans with patients, over 80% of the physicians in the study indicated that they would continue using the technique. The investigators also found that the physicians spent less than 7 minutes on average in discussing action plans. Physicians also reported that time limited the extent to which they could discuss action plans with patients. Some participants suggested an office visit devoted to collaborative goal setting.

Bodenheimer and Laing advocated inclusion of health coaches in the context of the typical office visit (2007). This new unit of care, called a "teamlet", permits a redistribution of responsibilities and tasks with the aim of increasing physician efficacy, as well as improving patient satisfaction. Current non-physician medical staff such as nurses could be recruited into the new role as the health coach. The primary shift in responsibilities include permitting the health coaches to make routine orders and elicit patient histories. Notably, Bodenhiemer proposes that health coaches help patients generate action plans after the patient has seen the clinician. Given the time constraints that physicians face, this solution presents chronically ill patients with more "face time" with a professional who can provide guidance for better managing their conditions.

To demonstrate the feasibility of recruiting nurses in this new role, Bennett and her colleagues (2005) conducted a study using a 6 to 9 month health coaching intervention. Two nurses received 24 hours of training using motivational interviewing by an experienced trainer. Two groups of elderly participants were randomly assigned to the intervention and control groups—intervention groups were subsequently interviewed by the health coach, while the control group was thanked for their time and continued to receive care from their physician as normal. During the intervention interview the health coaches helped the participants in the experimental condition identify a health goal they wished to work on for the duration of the experiment. The patients were permitted to change their health goal if desired. During the experimental period, the health coaches maintained contact with the participants in the intervention group by phone or email at least once a month. Phone interviews were limited to 15 minutes. At the end of the experiment both groups were administered a survey to assess health state. Despite overall satisfaction with the coaching intervention, patients did not report improved health outcomes in the survey. The authors surmise that while their intervention targeted behavior change, the survey instrument measured changes in health outcomes; thus it was possible that they might have seen significant results had they selected a behavioral measure.

In contrast, Butterworth and her colleagues demonstrated significant improvements in health state as measured by a health assessment survey (Butterworth et al., 2006). They recruited participants from employees of an academic medical center; using a non-randomized design, they recruited participants in the intervention condition through mailed employee communications. A control group was also recruited in high traffic areas. Both groups took a baseline health survey, and a follow-up survey 3 months later. Participants in the intervention received an initial health coaching interview and two follow up contacts. Patients in the intervention group showed a significant improvement in their health state as measured by survey.

Following up and expanding this study, Prochaska et al. (2008) conducted a randomized trial on a similar subject pool: employees at a medical university. Participants were assigned to one of three groups: a health assessment plus one-time feedback regarding behavior change; health assessment plus two follow-up motivational interview sessions; or a health record assessment plus at least three sessions with an interactive software that provides tailored feedback according to the user's stage of change. The investigators found that motivational interviewing and online intervention both produced significant behavior change. Notably, feedback generated by a computer program was efficacious in producing behavior change, without the intervention of a health coach.

Other investigators have explored other means of utilizing information technology to support health coaching. Goldberg et al. (2004) used a web based module to facilitate collaborative action plan formation between patients and physicians. The intervention was expected to enable the patient to better self-manage his or her diabetes. The action plan tool included an assessment tool to gauge the patient's self-efficacy level. However, the use of the tool was voluntary, resulting in low usage among clinicians and low statistical power. However, the authors found that non-physician staff were more likely to use the tool relative to physicians. Calfas et al. (2002) used a computer based assessment tool in primary care environments to promote shared health goal setting between patients and their doctors. Patients in the intervention condition showed successful behavior change as measured by surveys. Among a cohort of diabetic patients, Estabrooks et al. (2005) deployed a CD-ROM based assessment and goal setting tool and elicited successful adherence to self-generated action plans in a cohort of diabetic patients.

While some of the evidence is mixed, health coaching shows promise as a cost-effective means to promote and reinforce behavior change among patients with chronic disease or are at risk. Implementing the IOM and Bodenheimer's recommendations, the foregoing studies show that non-physician staff can effectively perform health coaching tasks, easing physicians' burdens. Furthermore, information technology was shown to to be feasible in the clinical and non-clinical environment. Computers can be also be employed to extend the reach of health coaches by tailoring and to some degree automating the coaching delivery (Shapiro, 2009). The Oregon Center for Aging and Technology (ORCATECH) Living Lab Project integrates these strategies to help elderly participants maintain independence and preserve cognitive health.

1.3 Cognitive Health Coaching

The Cognitive Health Coaching project is one of several projects initiated by the Oregon Center for Aging and Technology (ORCATECH). The main goal of ORCATECH is to improve the health and overall quality of life among the elderly by researching and developing assistive technologies(Oregon Center for Aging and Technology, 2010). ORCATECH provides a number of core resources to researchers including medical and neuropsychological assessment for research participants, coordination of participant outreach and recruitment, and various electronic resources for coordinating research activities. One of the resources provided by ORCATECH, the Living Lab is a population of seniors that have agreed to participate in a variety of experiments incorporating unobtrusive monitoring technologies. Examples include: bed sensors to detect sleep quality; medication monitoring to ensure patient adherence; and computer usage.

Recruiting members of the Living Lab, Jimison and her colleagues have initiated a cognitive health coaching project (Jimison et al., 2007b). The main thrust of the Cognitive Health Coaching Project is to create a modular architecture that can simplify the provision of tailored health coaching. The different modules provide coaching covering a variety of domains, including physical exercise, socialization, and cognitive exercise. Using algorithms that take into account the participant's stage of change, stated health goals, and willingness to share data, the Health Coach System assists the health coaches in formulating action plans. In gathering input to design the health coaching system, focus groups groups revealed that seniors prioritized cognitive function and independence over physical health. Addressing those needs, the investigators developed an internet application that provided a portal for cognitive health coaching and related activities. The interface presented games meant to both stimulate cognition as well as provide measures of cognitive performance to the coaches (see Jimison et al., 2004 or Jimison et al., 2007a for a discussion). Health coaches work with the patients to formulate action plans to reinforce cognitive health. Activities include novelty exercises (e.g., counting backwards from 100 by 7, or brushing one's teeth with the non-dominant hand), socializing with family or other participants via Skype (Skype Technologies, S.A., Luxembourg), or playing games. The first version provided basic communication capabilities that were one way from patient to the coach; messages from the coach were relayed to the participants via email. Feedback for game performance was provided solely through a window showing scores. Also, feedback game participation was given through a series of "badges" (see Figure 2.2). Players would earn badges with increased play.

The second iteration of the web application's interface was driven by a desire to include more two-way communication between the player and the health coach within the interface itself (instead of communicating via email or other external means). The interface was redesigned to incorporate more detailed feedback for the users regarding frequency of gameplay as well as their scores (see the Methods section for screenshots). The redesign also incorporated an action plan module that displayed the goals set collaboratively by the patient and the health coach. The patient is able to view targets for the week, and their progress relative to that target. They can also see their game scores and usage data. The enhanced feedback should prove a significant enhancement, as feedback regarding one's action plan progress has been shown to elicit better performance from participants in a number of studies (Bodenheimer and Handley, 2009).

The current study seeks to test the usability of a third version of the patient interface. A mockup of the second version was tested in five participants' homes (Yu et al., in press). While the interface was received positively for the most part, the wording for some interface elements were ambiguous. Moreover the "accordion style" menu was unfamiliar to some. The version under study contains color coding of the participant's progress to facilitate assessment of his or her progress relative to the goals of the action plan. Additionally, the new interface provides the capability for the user to update his or her progress; this feature provides another avenue of communication from the participant to the health coach. Once the action plan is updated, the health coach can see that change reflected in the coach's view of the patient's progress. Changes were introduced to the display of the game usage and score displays: instead of smoothed plots, data points for each episode of game play are displayed as discrete points on the graph (see Methods section for screenshots). The goal of this study is to identify deficiencies in the design so that the developers can make the interface easier to use.

1.4 Usability Testing

1.4.1 Discount Usability

Usability is typically defined as a collection of attributes pertaining to the ease of use of the system. Nielsen (1993, p. 26) lists five qualities central to usability:

- Learnability
- Efficiency
- Memorability
- Error reduction
- Satisfaction

Nielsen asserts that these qualities are best developed with the input of intended end users, a paradigm called iterative design. In designing a product that is usable, it is desirable to involve users at early stages of development(Stone et al., 2005). Early and frequent engagement of the end users allows the developers to make changes to the system when changes do not require extensive re-writes. Frequent elicitation of user feedback is a pillar of iterative design; ideally the end-user's input informs the design at most stages of the system's development. Usability assessments can either be formative or summative. Formative usability refers to testing performed during phases of development, so as to inform the design of the finished product. Summative evaluation is performed to test the usability of a more finished product, typically after it has been released to its intended users. The results of summative evaluation can provide feedback for developers to make further refinements to the finished product.

Because frequent end-user engagement is key aspect of Nielsen's discount usability framework, Nielsen proposes a mode of usability testing that is easy to conduct and does not require expensive equipment (Nielsen, 1993, 1995). More sophisticated usability testing techniques typically require specialized laboratories (Tullis et al., 2002; Thompson et al., 2004; Andreasen et al., 2007). A typical configuration might include a testing room equipped with a computer with software to capture the screen and cameras to capture the users facial expressions. A second room would permit usability testers and developers to discuss the user's performance. Verbal data might be transcribed and coded. In contrast, discount usability utilizes techniques that obviate the need for special facilities and equipment. Discount usability relies on the following techniques: user and task observation; scenarios; simplified think aloud; and heuristic evaluation. Of these, Nielsen considers think aloud to be the most important technique.

Think aloud, also known as verbal protocol analysis, is a technique that has its origin in cognitive psychology as a means to gain access to a subject's cognitions (Ericsson and Simon, 1998). The technique was adapted for usability testing, though the aims are different—while think aloud's usage in cognitive psychology is to understand cognition for understanding its workings, its use in usability tests focuses more on the evaluation of usable interfaces (Boren and Ramey, 2000). Analysis of think aloud data differs between usability and psychology orientations as well. Verbal protocol analysis for cognitive psychology may involve poring over recorded content, coding utterances, and comparing codes with other raters (Trickett and Trafton, 2009). On the other hand, Nielsen states, "In discount usability engineering we don't aim at perfection; we just want to find most of the usability problems" (Nielsen, 1993, p. 17). To that end, he suggests the usability specialist rely on notes taken in the field, and conduct analyses without rigorous transcription or coding.

1.4.2 Seniors and Usability

Assessing the usability of information technologies in the elderly is a pressing concern. More and more seniors are using information technology, particularly internet related technologies. According to surveys conducted by the Pew Internet & American Life project, people aged 70-75 displayed the greatest rapid demographic growth among internet users between 2005 and 2008 (Jones and Fox, 2009). In 2005, 26% of people in this age group were online; by 2008, that number increased to 45%. Among elderly aged 76 years and up, a 10% increase was seen from the 2005 level of 17%. Nielsen observed that senior citizens would show some of the most rapid demographic growth on the web (2002). The emergence of seniors online comes in concert with assertions made by researchers about the potential benefits of technology in improving the quality of life for the elderly population (e.g., Czaja and Hiltz, 2005), as well as the general aging of the baby boom generation, the so-called ”silver tsunami” (Fox, 2006, p. 1). Potential benefits commonly cited in the literature include increased social interaction, independence, and emotional wellbeing, (Czaja and Hiltz, 2005; Czaja et al., 2006; Czaja and Lee, 2007; Hendrix, 2000), and easy access to health information (McCray, 2005; Becker, 2004; Ownby et al., 2008).

In order to ensure the usability of computer interfaces for the elderly, age-related changes need to be accounted for. These changes include degradation in memory (Craik, 2000), psychomotor skills (Fisk et al., 2009; Jastrzembski et al., 2005; Walker et al., 1997), and vision. Working memory, the amount of information that an individual can hold has been shown to diminish with age. This has led some researchers to suggest changes to device interfaces to provide a better support for individuals with diminished memory, such navigational aids. Other solutions suggest facilitating procedural learning of computer tasks (Mead et al., 1999). Amongst experienced mouse users, differences were found in motor skills between young and elderly users (Walker et al., 1997). Despite evidence showing superior performance on pointing tasks using a light pen, elderly users still lagged behind younger subjects in a combined pointing and typing task (Jastrzembski et al., 2005). Tullis and his colleagues also found user behaviors unique to older adults (Chadwick-Dias et al., 2002). These behaviors included: a more cautious interaction approach; clicking on non-link text items; difficulty in ascertaining their location within the website; and difficult with tabbed navigation elements.

To address the usability needs of seniors, numerous design guidelines have been published. Researchers (e.g. Czaja and Lee, 2007; Nielsen, 2002; Demiris et al., 2001; Chadwick-Dias et al., 2002; Zaphiris et al., 2007, 2009), government organizations such as the National Institute on Aging (NIA) and the National Library of Medicine (NLM) (2002), as well as the World Wide Web Consortium (The World Wide Web Consortium, 2008) have all published guidelines for designers to better accommodate older audiences. These guidelines include prescriptions for legibility (especially using larger font sizes, or permitting one to increase font size), predictability, and simplicity. Findings such as these have led guideline authors to recommend larger on-screen interface elements, and menus that do not require users to hold down mouse buttons. While these guidelines offer some guidance to web designers, there is some evidence that recommendations are not followed rigorously (Becker, 2004). Petrie and her colleagues noted that an understanding and knowledge of guidelines are insufficient to truly understand the needs of the end user (Petrie et al., 2006). As technical innovations continue to reach a broader audience, attention to this field will warrant continued research. The literature offers some practical guidance for conducting usability tests among the elderly (Dickinson et al., 2007).

1.4.3 Remote Usability Testing

Remote usability testing is a testing methodology that entails separation of test subject and test administrator in physical location, time, or both (Castillo Jose et al., 1997). The two major categories of remote testing are asynchronous methods and synchronous methods. Asynchronous methods rely on passive modes of data collection. Data collection techniques include electronic surveys, critical incident reporting, and automated click-through data collection (Castillo Jose et al., 1997). Data can be automatically sent to the test administrator for later analysis. Synchronous methods are closer to traditional usability testing in that the moderator is “present” over a network connection to give instructions to the participant and monitor his or her progress. The moderator can interact with the participant either over a conventional telephone line, internet telephony through the computer, or an internet-based tele-conferencing application.

As internet technologies have matured, the broad dissemination of high speed internet has afforded the transmission of richer multimedia content. The Pew Internet and American Life Project estimates that 60% of American adults use broadband at home to access the internet (Rainie, 2010). Broadband permits the transmission of video feeds at a distance with consumer equipment. Initially, remote usability was implemented without the aid of internet technologies, utilizing local area networks and closed circuit televisions. However, the wide adoption of broadband permits remote user testing over wide geographic areas; this is especially important if the potential users and participants are not readily found locally (Brush et al., 2004) or if they are located on other continents (Dray and Siegel, 2004). The relative ease of recruitment and coordination of testing schedules are prime motivators for remote usability testing. Costs associated with travel can also be obviated using remote usability testing (McFadden et al., 2002). For the current study, eliminating the need to travel is especially important; the participants of the Living Lab may be more reluctant to travel long distances or have mobility impairments. Thus, remote usability methods promise to facilitate user-centered design principles.

Chapter 2

Methods

2.1 Participants

As indicated in the previous section, the Cognitive Health Coaching participants are members of the ORCATECH Living Lab, a cohort of community dwelling seniors in the Portland Metropolitan area who have agreed to participate in a variety of studies that evaluate assistive technologies. The participants are in a variety of living situations—some live independently in the community, while others reside in residential facilities. The average age of the participants is 76.1 $\pm$ 4.9 years. As part of the ORCATECH Living Lab, the participants have a variety of passive monitoring devices installed in their home, including bed sensors, motion detectors and medication tracking devices. The subjects of the current study were recruited from this broader pool of seniors. Information sessions were arranged at OHSU, and the potential recruits were introduced to the interface and novelty exercises. All participants had been equipped with standardized personal computers with internet connectivity, webcams, and the suite of Living Lab games pre-installed. While the majority of participants received standard minitower computers, some participants received laptops due to space constraints. Also, while some software was provided by the experimenters, the seniors were free to install additional software as they pleased.

After the information session, the seniors who elected to enroll in this research project were contacted by the health coaches and maintained contact via email and Skype. For the duration of this project the coaches collaboratively set action plans and engage the seniors in online socialization activities via the internet.

The present participant cohort (n = 5) was selected from this group of seniors. This convenience sample was selected based on availability through the Skype service. A sample size of five subjects was chosen on the basis of optimal sample sizes reported in the literature (Nielsen and Landauer, 1993). Additionally, in collecting think aloud data the researcher is less interested in achieving statistical significance, and more interested in richer and nuanced data than what quantitative metrics can provide (Trickett and Trafton, 2009). Additionally, one of the goals of the discount usability paradigm is to reduce the barriers to conducting usability tests. Thus, a smaller sample can prove more cost effective if major usability problems are uncovered (Nielsen, 1993). All participants had granted prior consent to participate in a variety of activities sanctioned by the Living Lab Project.

2.2 Software

A variety of software was used to prepare and conduct the usability test. Dameware Mini Remote Control (Dameware Development LLC., Covington, LA) was primarily used to enable remote access to the participant's computer. This application is used by other Living Lab research assistants to perform system upgrades and software installations. Once the subject assented to participate in usability testing, the research assistant installed a newer version of the patient interface that incorporated some interface enhancements. Also installed at this time was Blueberry Express (Blueberry Software Ltd., Leamington Spa, UK), a screen recording software. This application recorded both a video of the participant's screen as well as video from the users web camera, permitting simultaneous views of the user's screen and facial expression. Recordings produced by Blueberry were saved on the participant's local hard drive, and subsequently transferred to the test-giver's drive using Dameware.

While Skype was used to establish an initial contact with the participant, Gotomeeting (Citrix Systems Inc., Ft. Lauderdale) was used for the actual test session. Gotomeeting was chosen for its capacity to stream a video feed of a meeting participant's desktop. Additionally, Gotomeeting supports multiple users on a single meeting session, as well as limited screen video capture capability. During the usability test, instructions were given over the Gotomeeting system. In the event that Blueberry could not be loaded onto the participant's computer, the Gotomeeting recording capabilities were used.

2.3 Test Interface

The earlier versions of the interface provide basic information about high scores, as shown in Figure 2.1 Additional feedback was given in the form of "badges", which

Figure 2.1: This figure shows the main game selection window. High scores from previous games are shown in the panel on the right.

could be obtained through repeated gameplay (see Figure 2.2). The test prototype

Figure 2.2: This screenshot displays the number of badges earned through repeated gameplay. Progress towards the next badge is indicated by the bar below game title.

expanded on the feedback given to the game player. In addition to information regarding gameplay, the new interface aimed to provide a longitudinal view of game performance (as measured by scores). The interface permits three increasing time windows over which to display scores, allowing the participant to view scores and gameplay frequency over the last week, three months, and the entire span of his or her enrollment in the study.

Another feature added to the interface was the capability for the patient to update his or her progress on the action plan screen. This allows the participant to perform some of the data entry (Figure 2.3). This allows greater interactivity in the interface.

2.4 Testing Procedure

A research assistant established contact with individuals who were available on Skype. If the subject was willing to participate in a test, the research assistant provided a brief

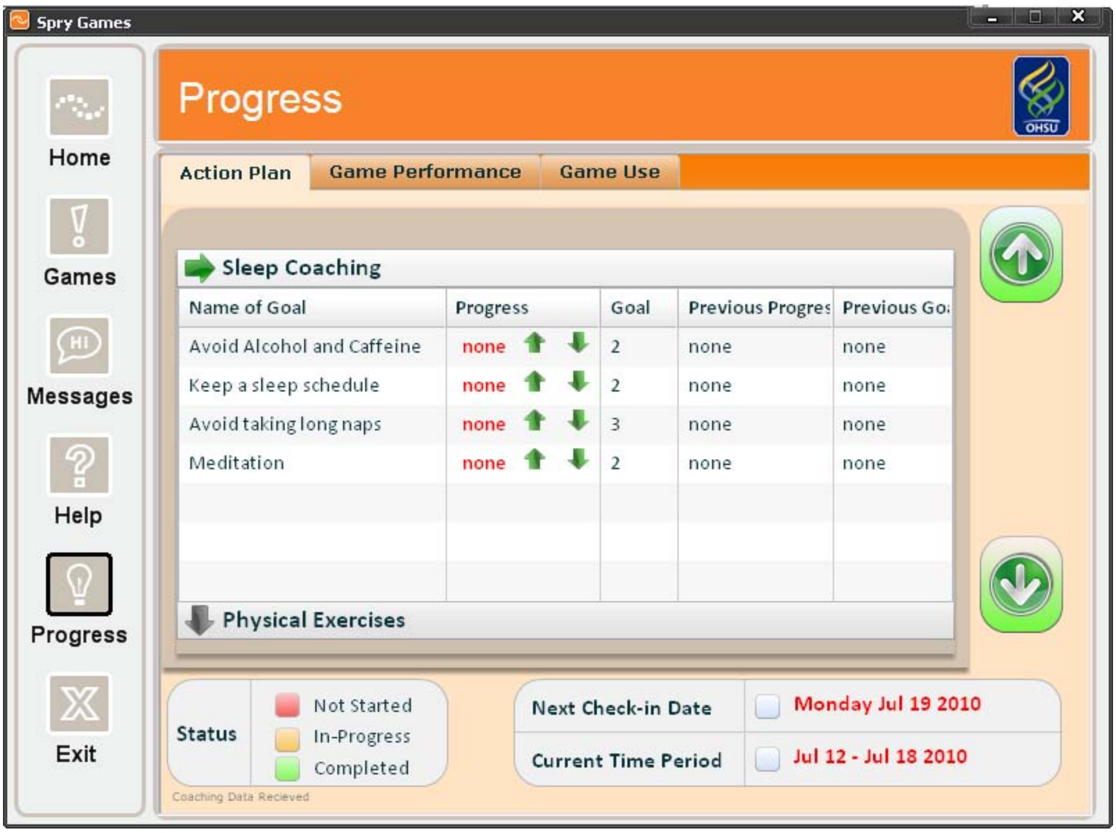

Figure 2.3: The enhanced action plan. The green arrows to the side of the progress field allow the user to record their activities, which the health coach can view at a later time.

summary of the test and its purpose. Next, the assistant used software to establish virtual desktop connection (software described above) to prepare the participant's computer for testing. This included installing screen capture software, loading the new version of the interface, and setting up a web conference with Gotomeeting. Once the participant's computer had the appropriate setup, the facilitator introduced himself and the test to the participant.

The facilitator then explained the idea behind usability testing, and explained the idea of the think aloud protocol. Following advice given by Boren and Ramey (2000), the facilitator explained the difference between the think aloud protocol and explanations directed at another person. The facilitator then demonstrated the think aloud protocol using a website. The facilitator also explained that the interface was the "test subject" and that the participant's responses were not judged in any way. Following the briefing, the tasks were given to the participants one at a time. If the participant seemed lost, the facilitator gave prompts and hints.

When usability testing was completed, the facilitator thanked and debriefed the participant. The research assistant used Dameware to obtain the screen capture movie file, which was saved on the participant's hard drive. Finally, the testing software and experimental interface was removed from the participant's computer.

2.5 Tasks

Based on an analysis of the goals of the patient interface, five representative tasks were created to serve as a focal point for collecting usability data. It was determined that besides playing the games, the interface would provide: a means of communicating between the participant and the health coach; a means for the participant to enter rudimentary data concerning his or her progress with respect to the action plan; and a way to get feedback either actively from coaching methods, or passively through the various score and participation displays.

To that end, the five tasks were formulated to capture these basic aims, as listed in Table 2.1.

Task	Interface goals
1. Send a message via the comments window to the health coach	Communication with the health coach
2. Read the third to last message sent to the patient (from the coach)	Communication with health coach
3. Update the number of times a novelty exercise was performed under the action plan	basic data entry
4. Read off the last time Scavenger Hunt was played in the last year	Receiving feedback
5. Read off the highest score received while playing Spelling Bee	Receiving feedback

Table 2.1: Table of the five usability tasks given to the participants, and their mapping to the primary goals of the interface

2.6 Analysis

In keeping with the philosophy of discount usability, the recorded data was reviewed by the author primarily for usability issues. In contrast to verbal protocol analysis where the goal is to uncover themes that reflect cognitive processes of the participant, the emphasis in discount usability is discovery of interface problems (Nielsen, 1993). In analyzing typical verbal protocol data, the emphasis is on the participant—the software may be used to elicit verbal data. In analyzing think aloud data for usability studies, the emphasis is on the system, using the verbal reports as indicators of usability problems. Thus, the author supplemented the field notes collected during the usability session with examination of the videos to detect problems with the interface.

Chapter 3

Results

3.1 Task Completion

The users were not able to accomplish the tasks unaided for the majority of tasks. When the participant was seemed unable to accomplish the task, the test administrator gave hints to the participant in the form of relevant areas to click. The following section details the usability problems encountered organized by task. In general, task completion was hampered by a combination of timid interaction style, participants' physical limitations, and deficiencies in the interface.

A major problem with the test software that affected all tasks was its "time out behavior"—after a period of inactivity, the program would reset itself to the start screen. This would happen in the middle of usability testing, requiring the user to retrace his or her steps. The reset behavior sometimes occured more than once during a session, sometimes distracting the participant and hampering task completion.

3.2 Usability Problems Found

3.2.1 Sending a brief message to the health coach

This task proved to be very difficult for the users. One of the key usability problems with the task was the location of the function. The comment screen was located in the part of the application dedicated to providing help to the user. This juxtaposition may have caused some confusion over what the exact function of the window. Because the window is located in the help section, ostensibly the comments are for the purpose of soliciting technical assistance from the health coach. Yet, the functionality for sending messages to the health coaches is labeled as "Comments", as shown in Figure 3.1. The naming implies that the user is to provide feedback regarding the system, which is also indicated by text in the window. The intended recipients are the health coaches, but that may not be clear from the naming of the section.

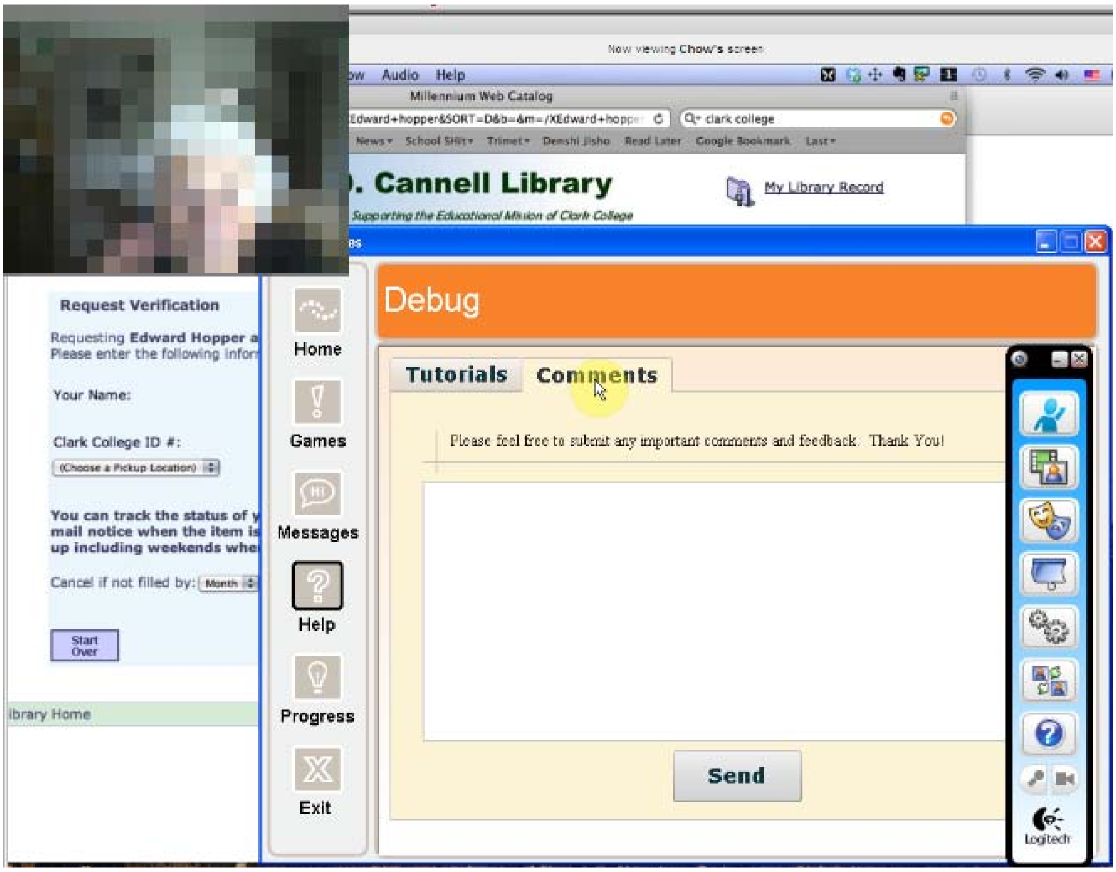

Figure 3.1: The comments field is located under the help section along with the tutorials. While ostensibly used to send feedback to the health coaches, it's location makes it hard to find.

Thus, the mismatch between the intended function of the window and its location within the application could limit its usefulness to end-users and health coaches.

Because of the ambiguous naming, most users were explicitly guided to the help section. When directed to the Help screen, some of the participants were unable to see the tabs at the top of the page. This inability to immediately see the tabs was found in other tasks where the patients were required to find functionality in other screens by using the tabs, described later.

3.2.2 Reading messages through the message window

The second task was to read off a message from the message screen (Figure 3.2.

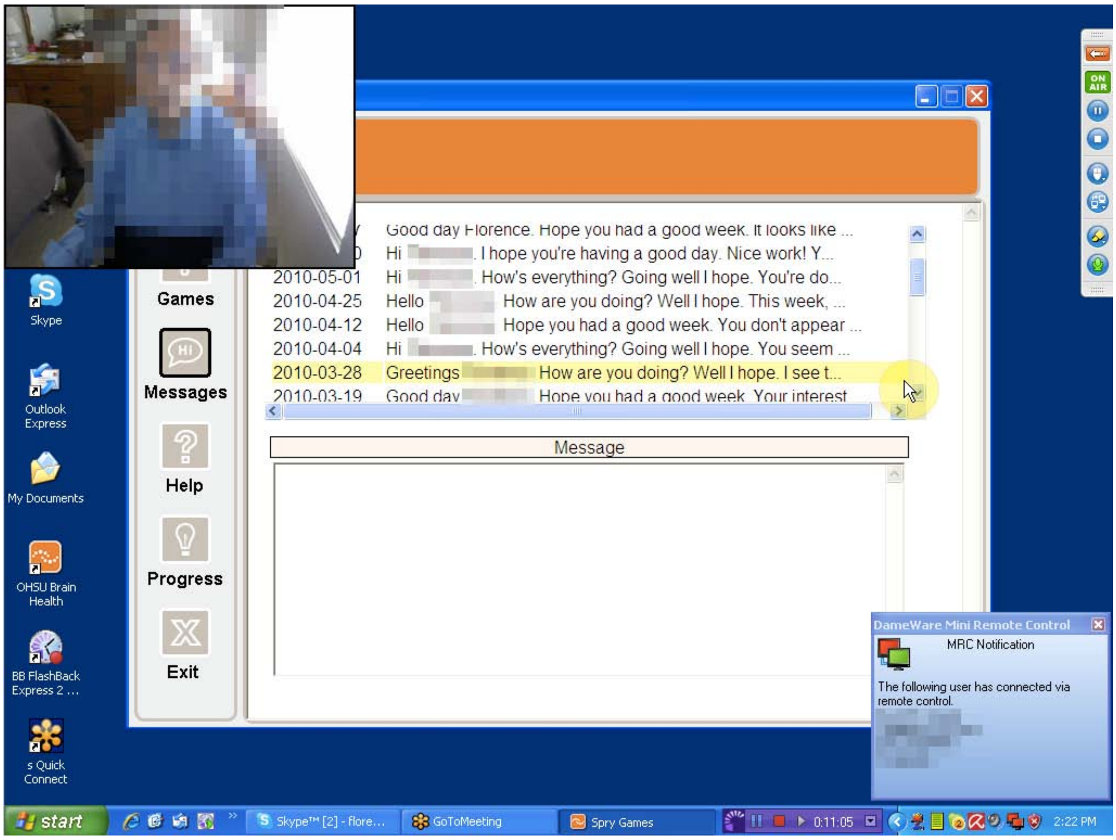

Figure 3.2: The Messages screen displays a running log of all the messages sent to the user by the health coach. The messages are ordered in descending chronological order, with more recent messages toward the top of the display.

The original intent of the message screen was to allow the participants to read off past messages (either hand written or generated automatically) sent by the health coach. In addition to being displayed through this window, the messages were also sent over email and displayed on the welcome screen when the game program is first invoked. The message window consists of two main elements: a top pane that lists the first few words of the message in chronological order, and a bottom pane that displays the full text of a message selected in the top window. The amount of text displayed in this top area varies with the size of display type set by the participant. When the user first navigates to the screen, the bottom pane is initially blank, as shown in Figure 3.2.When a message line is clicked in the square above, the full message appears at the bottom.

When asked to read off a message from the list, several users started to read the segment in the top portion of the window, but stopping when the message was cut off. This behavior implies that they did not know to click on the message line they were reading to display the whole message. Several design issues may contribute to this issue. First, the window below is initially blank. Newcomers to this screen might not know that the bottom portion is for the display of the full message; the blank square could suggest a text entry field or a display field. The second usability issue lies in the lack of affordances for the window's top pane. Scant cues are provided indicating that the lines are "clickable". Unlike a typical email inbox, the top window lacks headers to suggest to the user that the displayed messages are only extracts meant to be clicked on. In a typical email inbox window, the main content of the inbox pane typically include the date of delivery, the sender, and the subject. It is possible that the absence of body text in inbox displays are an invitation to click: users know that the subject headers are not the actual message, and thus conclude that to see the actual message content they must select the appropriate message.

When asked whether this interface was used to access older messages, one user replied,

> *You know what? No. I have always had a problem with that. I have to, when I see that, and they have click on to your message, mine won't work when you say click on to the message to read, mine have never worked so what I do, I go to my email and read my old messages or go up to home, now if I hit home, I see, 'Hi Ruth. Hope things are going well this week'. That's for this week. Now okay? But if I click on*

> *that, on this yellow, okay it's highlighted in um yellow right now, if I click on that, nothin.*

This user ended up checking email, a more familiar interface, to look at old messages. Other users echoed this sentiment, saying they preferred to look at messages through email or to read the most recent message on the welcome screen.

3.2.3 Updating Action Plan Progress

Figure 2.3 shows the action plan screen. Users can communicate their action plan progress by incrementing the activity counts using the green arrows. All of the participants were unable to perform this task. A few participants were confused by the status legend at the bottom: the red, orange, and green squares are a guide for the progress text. In Figure 2.3, the progress notes are all set to "none"; they are displayed in red to reflect the "not started" status; if the users had entered a number lower than the stated goal, the text would appear in yellow. Some participants clicked on the colored squares, thinking they were buttons. This might have been an issue of affordances, as the legend icons have a slight three dimensional look to them. This three dimensional look may lend a sense of clickability to the legend markers where in fact they are not interactive elements.

Problematic affordances also affected the "accordion menu" interface, where clicking on the accordion expands one of the menu items while collapsing the others. In Figure 2.3, "Sleep Coaching" is expanded while "Physical Exercises" is collapsed—clicking on the Physical Exercise bar will expand that section and collapse the other sections. One of the participants was unfamiliar with the accordion interface and thus did not know to click on the header initially. The user did eventually click on the proper section, but the participant's interactions did not suggest that he understood the menu's interactivity. The larger arrows to the right of the menu seemed more clickable, and participants more readily clicked the buttons rather than the accordion headings.

Finally, the green arrow buttons used to increment action plan progress did not invite interaction from the participants without specific instruction from the facilitator to click. While participants thought the legend was interactive, the arrows embedded in the action plan did not prompt interaction. The arrows were positioned near the action plan progress cell and stood out from the other elements by virtue of its shape as well as its color. Nevertheless, the participants did not view the arrows as something to be clicked.

3.2.4 Reading Performance Graphs

The last two tasks called for the participant to read graphs displaying his or her performance (see Figure 3.3. The two similar windows, game performance and game scores, were meant to replace the badges window (see Figure 2.2).

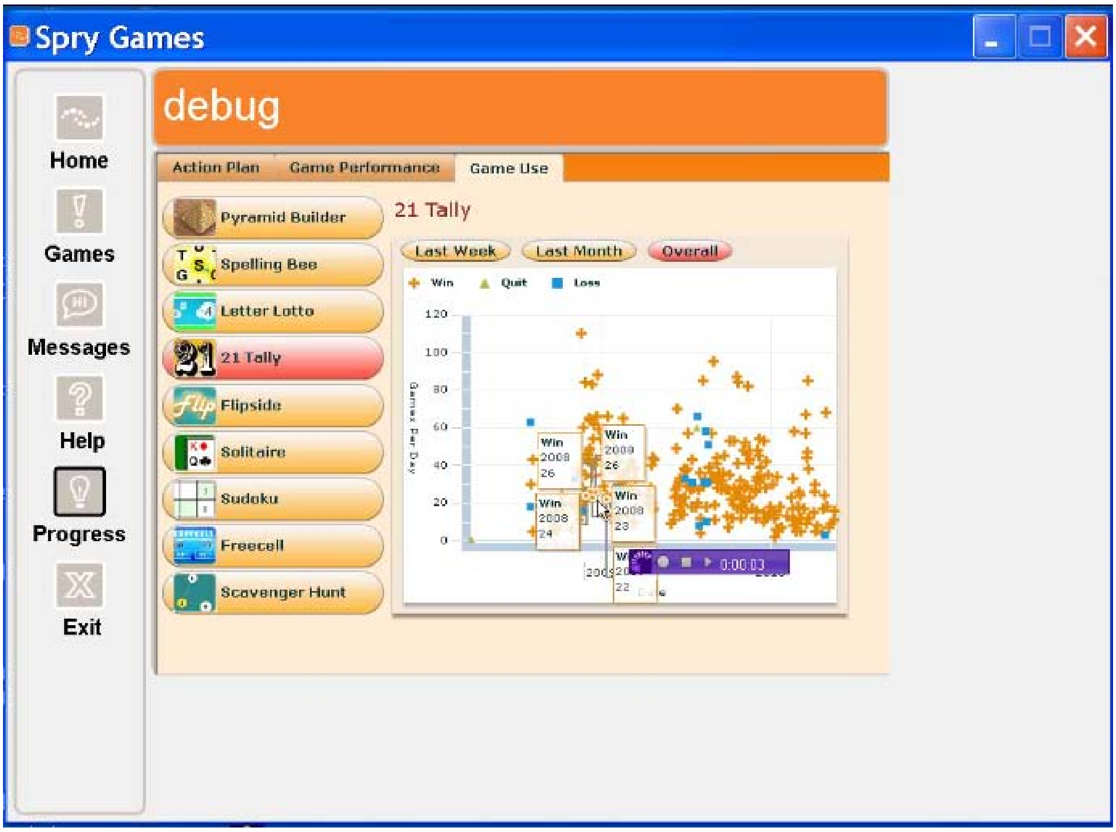

Figure 3.3: An example of a game metric graph. For game usage, each datapoint represents the number of games played on a given day. For game performance graphs, each data point represents a single game with the final score for the game.

As shown in the figure, windows with some more commentary popped-up when the mouse pointer hovered over a data point. These windows, game use and game performance, had an additional layer of complexity in that the graph display could be adjusted to show data from three increasing time windows. The user could choose to view metrics from the past week, past month, or their lifetime game performance.

No participant was able to navigate the screen without prompting or help from the test administrator. The participants had difficulty locating the tabs at the top of the window that initially led to the graphs page, a problem that affected the other tasks. One user, using his mouse pointer as a focus for attention, moved his cursor past the relevant tab, only to find it much later. The added level of complexity with the three different time window views proved to be a taxing visual search for the participants.

One of the participants was completely unable to use the graphs as a result of special screen magnification software (see Figure 3.4). The field of view afforded by the magnification window prevented the user from seeing all of the graph at once, significantly limiting its usefulness. Also, the pop-up windows frequently popped up outside of the magnifier's field of view; because the magnification window followed the cursor's motion, attempting to move the magnification window to see the message would result in the cursor moving off of the relevant data point and thus closing the window. This system behavior thus prevented the user from gaining any information through the pop-up window.

Another user became distracted by the pop-up windows. For some data points, many pop-up windows are invoked which can disorient as the user moves the pointer across the graph. In some cases, the number of pop-up windows exceeded the available screen real estate, resulting in some of the windows being obscured (see Figure 3.5

The popping up of several windows had a somewhat disorienting effect on the user. Additionally, the small size of the pop-up window did not permit sufficient explanatory text, leading the user to wonder the meaning of the pop-up's text labels:

> *In 2009, I played, this is, one fifteen... I don't know what they're talking about. I... okay the 2009, you know, the arrow... the line that goes straight up? Okay, I just see what 2009 they have fifteen, but what does that mean? I don't know...*

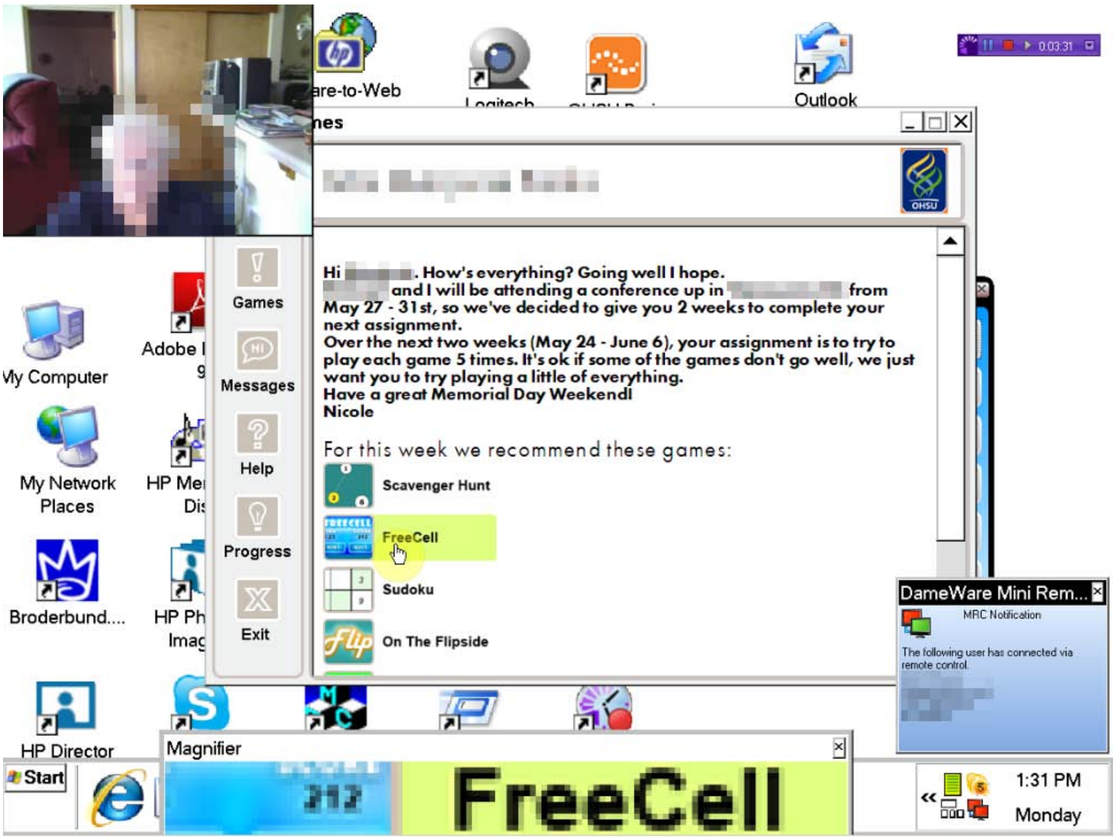

Figure 3.4: One of the participants used special screen magnification software. During the course of the usability test, it was clear that she attended almost exclusively to the magnfied window at the bottom of the screen.

One of the participants did not find the scatterplot formate informative. She was looking at a particularly sparse set of data, and did not recognize the display as a graph. This problem was repeated by another participant when viewing a sparse scatterplot. Instead of reading off the few data points that were in the main portion of the graph, she mistakenly read the axis labels. The particular graph that she was reading from was displaying metrics from a game she played frequently. Because the the boundaries of the axis were set by overall game data, the maximum value of the axis was 2.2×10^9. This caused the data points on the screen to be close to the x-axis, nearly at the bottom. In turn, she failed to take note of the tick marks, mistaking the axis labels as her high scores.

One of the users seemed to miss the purpose of the screen altogether. When asked to find a previous high score, she attempted to recall a score from memory in lieu of read a previous score from the screen. Another participant commented on the difficulty of

Figure 3.5: Some of the data points invoked an excessive number of pop-up windows, so that some windows could not be fully displayed

the screen: "...I probably could figure this out if I sat here and really looked at it, you can kind of ...uh you have to work at it. Put your brain to work here."

Ultimately, the users received most of their performance information from a high scores display that was presented to the right of the the game selection screen (Figure 2.1). When asked how they gauged their performance, a number of participants would read off their last score from this display. Other participants were indifferent to their scores, noting that they played primarily for fun.

Chapter 4

Discussion

4.1 Technical Barriers

One of the aims of this study was to explore the feasibility of remote usability testing among the living lab participants, particularly with respect to the software and hardware capabilities of the testing apparatus. In spite of having standardized hardware and the same base software (including software that enabled remote desktop control), external factors presented barriers to smooth usability testing. The remote desktop client, combined with network performance, posed the greatest hindrance to usability testing. On several occasions, the remote desktop software was unable to establish a connection with the tester's computer. Unable to connect, the research assistant was forced to abandon the usability test session on a number of occasions. When a connection was established, the the performance of the remote desktop management occasionally left much to be desired, making the preparation of the participant's computer more laborious than expected. Because the participant's webcam was not activated at that time, it was difficult to gauge if the participant was frustrated or not. Regardless, the technical difficulties associated with the set up caused some of the test sessions to go less smoothly.

Once the software was installed, configuration of the software itself proved challenging. While Skype had been configured by health coaches previously, the conferencing software had not been tested with some participants. Consequently, audio settings needed to be diagnosed with the participant. This might have been another source of frustration, as the testing was tedious ("Can you hear me now? How about now?"). It is possible that persistent audio problems arose from a conflict between Skype and Gotomeeting, but this theory could not be verified.

Software configurations also affected the installation procedure. Some of the participants had installed many computer programs on their own, which might have contributed to the performance gaps. One computer in particular had many extra programs, possibly contributing to the slowness of the remote desktop program. It underscored the unpredictability of remote testing computers; even with standard hardware and base set of software, there was still considerable variability in how participants configured their computers.

Another participant had turned on an accessibility feature designed for people with impaired vision. The desktop was magnified to the point that not all of the desktop was visible at one time; the screen panned to show the edges of the screen. Also, the user had a window that magnified the area around the pointer, obscuring a portion of the bottom screen (see Figure 3.4). Usability problems associated specifically with the magnifier included difficulty in using the graph as mentioned previously. In spite of the seemingly small field of vision afforded by the magnifier, the participant stated that it did not hinder gameplay.

In spite of the variety of unique configurations, once the internet conference software was installed and configured, the remote desktop was no longer necessary. As soon as the conferencing application and screen recording software was fully configured, the network or the other software did not impede the usability testing in any way. The video conferencing software turned out to be quite robust and we were pleased with the results.

A software issue common to all users was the interface's reset behavior; after a period of inactivity, the game would revert back to the home screen. This proved to impede the usability testing, especially where the seniors tended to take longer on certain tasks or get distracted. On a number of occasions, the software would revert to the home screen and the participant had to retrace his or her steps back to the previous location, interrupting their train of thought.

4.2 Cognitive Barriers

A common theme among the participants was insecurity with the computer. In exploring strategies to instruct seniors to use the world wide web, Dickinson and her colleagues observed greater anxiety and a less confident usage pattern (Dickinson et al., 2005). Marquié also documented a disparity in self-reported computer knowledge and general cultural knowledge between young and elderly subjects; for computer-related knowledge, elderly users scored their own retrospective and prospective knowledge significantly lower than general knowledge, while younger participants showed no difference in self-assessment scores in the two knowledge domains (Marquié et al., 2002).

The users in the present study showed a similar pattern of anxiety and lack of confidence. In particular, some users reported a fear of causing harm to the computer:

> *I'm always afraid that I'm going to smash this piece of equipment or that I'm something's going to break and I get real real upset and I don't know how, but I have encouragement... I'm slowly slowly getting to it but, you know, I'm afraid everything is going to smash and all the wires are gonna stop. . .*

The reluctance to explore on the computer seems to result in a reinforcing pattern of anxiety: "I don't surf around on the computer, I'm always I'll get in trouble so I don't do it... I just don't want to do something that I screw up the computer... I'm not a real brave person I guess."

These examples of anxiety illustrate some of the challenges in conducting think aloud studies in an elderly population. Dickinson et al. (2007) remarks that the combined cognitive effort of exploring a new interface combined with the novel task of thinking aloud can impede task completing. We found that while the tasks proved challenging for the participants, the hesitant use pattern also hindered completing the tasks successfully. For think aloud challenge tasks to be completed successfully, a certain amount of exploratory behavior is necessary for the participant to discover the solution. However, test performance anxiety may have rendered the participants more reluctant to search for the solution: “I little bit [unintelligible] trying to follow through with the lines but uh I mean, it’s got to be I mean I guess I’m a little bit nervous”.

The low level of exploratory behavior also explains why all of the participants had never seen the test interfaces. The interfaces had been deployed for at least one and a half months prior to the usability test. When asked, all of the respondents noted that they had never seen the test interfaces. This finding suggests that none of the participants explored the program beyond typical gameplay functions.

4.3 Methodological Issues

In conducting the usability tests, we attempted to follow guidelines outlined by Boren and Ramey (2000). In line with the paradigm established by Ericsson and Simon (1998), key among their recommendations was to demonstrate the think aloud task, and to minimize interaction with the participants to obtain a narrative as close to “raw” cognition as possible. At the same time, Nielsen (1993) distinguishes the use of think aloud in usability testing as primarily oriented toward discovering usability problems, while more traditional verbal protocol analysis is aimed towards uncovering cognitive mechanisms.

While efforts were made to assure the participant that the system was being tested and not the user, and think aloud demonstrated by the test administrator prior to the usability test, the novelty of the task proved challenging. Participants were instructed to speak as if speaking to oneself, and not as an explanation to an interlocutor. However, the patients ended up engaging the test administrator in conversation. Difficulties encountered during task performance prompted normal conversation, as opposed to thinking aloud.

Dickinson and her colleagues propose that alternative methods of usability testing may prove more efficacious, such as focus groups or presentation of a multiple protocols

(Dickinson et al., 2007). The investigators suggested this alternative to offset any hesitancy at criticizing the designer's work. If several alternatives were presented, the participants might be more forthcoming in comparing the different potential interfaces and present criticism more readily. Finally, if think aloud were a skill that could be practiced, several practice sessions with different test software could be conducted prior to the actual usability test to ensure the participants' familiarity with the task.

4.4 Design Considerations

In reviewing the usability problems identified by the participants, some design changes could be implemented to improve usability. Regarding the first task, sending a message to the health coaches, the observed low completion rate could be attributable to two reasons. First, the screen's location within the program was in the help section, which users might not associate with communication. A possible solution is to move the message screen in the "Messages" section of the program, which would centralize all of the functions related to communication in a single location. Secondly, the screen is located in a tab titled "comments", which may be ambiguous. By changing the title to something akin to "Send a message to the health coach", the function will be clearer to the end user.

Another usability issue related to communication concerns the usability of the message screen (Figure 3.2. A number of participants did not realize that the bottom portion of the message window was for the display of the messages previewed in the upper pane. One user mistook it for a text-entry field. A simple way to make its purpose clearer is to include some grayed out explanation in the field explaining that selection of a message in the upper pane will cause the full text of the message to appear in the bottom pane. Another way of reinforcing the window's function is to make the messages more explicitly "clickable" to ensure that the user knows the message list does not contain the full message, but just an extract. One way is to make the header include a subject instead of a preview of the message text. This could indicate to the user that the message line needs to be clicked before the entire message can be viewed.

A pop-up message might also be effective, with text that clarifies that the message line needs to be clicked, though pop-ups carry some risk of distracting elderly users.

The action plan is the most visually complex window in the system. The nested tabs, accordion menus, and interactive buttons were not usable for the participants. Tabs were difficult for them to use, as they spent much time locating them even when they were located in salient locations, such as near the top of the window. Also, some users seemed to exhibit difficulty distinguishing between figure and ground, occasionally confusing background application windows with the open application. Unfortunately, external constraints of the program limit the screen real estate available to the system designers, necessitating the use of space-saving interaction elements, such as tabs and collapsable menus. The author recommends visually simplifying the layout of the screen, at the cost of making the users step through more windows sequentially. Given that visual search may prove challenging to some users, reducing the dependence on tabs and collapsing menus could make the system easier to use. Instead, a sequence of simply laid out menus such with a clear "breadcrumb trail" that allows users to step back in the sequence may allow for the same level of interactivity but easing the cognitive burden of search a complex screen.

Specific to the display and updating of action plans, the user can be stepped through the process of incrementing progress indicators to the action plan. The initial start up screen can include reminders for the participant to update his or her progress, with links to the action plan window. New windows can allow the user to step through each domain of the action plan (physical exercise, game play, and social activities), which will permit the user to focus on a single task without being distracted by an excess of visual elements.

With respect to the progress graphs, it should be noted that the information needs of the users were not collected prior to their design; the interface designer made *a priori* conclusions regarding the usefulness of these graphs. While some exceptional subjects would be interested in seeing their performance over certain intervals, the usability sessions revealed that many users were not as concerned about high scores or game usage beyond the simple metrics offered by the side score display (Figure 2.1). It is suggested that the graphs be limited to a single time window view, combined with a simple high score list. Pop-up windows ought to be suppressed to simplify the display, since they may be more distracting to an elderly population. While the discrete scatter-plot layout was difficult for some users to read (especially when the points were very sparse), it may be preferable to displaying trend lines. In the case of sparse data, connecting the data points may add information where it is not warranted.

In discussions with the users, a common theme that arose was the reliance and appreciation for paper based documentation. Many users attested to keeping the paper manual near the computer for easy reference. One of the reasons given for the preference toward paper documentation was the physical affordance; the user could hold it up to the screen and refer to it while interacting with the system. Another possible reason is the familiarity with paper-based documents. If further modifications are made to the interface, a hard-copy manual sent to the users may increase the acceptability of the system enhancements.

4.5 Conclusion

Despite the promise of greater convenience for both tester and participant, technical and cognitive barriers arose multiple times during remote testing. Despite the uniformity of the hardware given to the test participants, some users heavily customized their computing environment to the point that unexpected software configurations hindered the smooth set up of the testing software. Although usability testing was conducted in surroundings familiar to the participants, the novel testing procedure and laborious set up may have elicited nervousness or other negative emotions. While efforts were made to reassure the participants that they were the system was under review and not the participants' performance, they still felt test anxiety which may have affected task completion and willingness to explore the interface.

In testing the usability of seniors, the difficulties the participants encounter are the result of a number of factors: the baseline usability of the system; the physical and cognitive limitations of the participants; and the lack of familiarity with computing metaphors. With respect to the second factor, the usability of the system was clearly impacted in the case of the participant with the screen magnifier. Other participants may also have visual attention deficits as evidenced by the difficulty in distinguishing figure and ground. Regarding the third factor, the lack of cultural familiarity may be a considered a cohort effect(Dickinson et al., 2005). As baby boomers age and familiarity with computer interface metaphors become more common among the elderly, this effect may fade with time. However, the design of computer systems today ought to meet the accessibility needs of the current cohort of elderly computer users. While remote usability offers the potential to involve greater participation in the design process from this less mobile user base, a more robust system for testing needs be implemented to minimize participant anxiety and encourage active engagement.

Bibliography

Alemayehu, B. and Warner, K. E. (2004). The lifetime distribution of health care costs. *Health Services Research*, 39(3):627–42.

Andreasen, M. S., Nielsen, H. V., Schrøder, S. O., and Stage, J. (2007). What happened to remote usability testing?: an empirical study of three methods. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1405–1414, New York, NY, USA. ACM.

Bandura, A. (2004). Health promotion by social cognitive means. *Health Education & Behavior*, 31(2):143–64.

Becker, S. (2004). A study of web usability for older adults seeking online health resources. *ACM transactions on computer-human interaction*, 11(4):387–406.

Bennett, J., Perrin, N., Hanson, G., Bennett, D., Gaynor, W., Flaherty-Robb, M., Joseph, C., Butterworth, S., and Potempa, K. (2005). Healthy aging demonstration project: Nurse coaching for behavior change in older adults. *Research in nursing & health*, 28(3):187–197.

Bodenheimer, T. and Handley, M. A. (2009). Goal-setting for behavior change in primary care: an exploration and status report. *Patient Educ Couns*, 76(2):174–80.

Bodenheimer, T. and Laing, B. Y. (2007). The teamlet model of primary care. *Annals of Family Medicine*, 5(5):457–61.

Bodenheimer, T., Lorig, K., Holman, H., and Grumbach, K. (2002). Patient self-management of chronic disease in primary care. *Journal of the American Medical Association*, 288:2469–2475.

Boren, T. and Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3):261–278.

Brush, A. B., Ames, M., and Davis, J. (2004). A comparison of synchronous remote and local usability studies for an expert interface. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 1179–1182, New York, NY, USA. ACM.

Butterworth, S., Linden, A., McClay, W., and Leo, M. C. (2006). Effect of motivational interviewing-based health coaching on employees' physical and mental health status. *Journal of occupational health psychology*, 11(4):358–65.

Calfas, K. J., Sallis, J. F., Zabinski, M. F., Wilfley, D. E., Rupp, J., Prochaska, J. J., Thompson, S., Pratt, M., and Patrick, K. (2002). Preliminary evaluation of a multicomponent program for nutrition and physical activity change in primary care: Pace+ for adults. *Prev Med*, 34(2):153–61.

Castillo Jose, C., Hartson, H., and Hix, D. (1997). Remote usability at a glance. Technical report, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0106.

Chadwick-Dias, A., McNulty, M., and Tullis, T. (2002). *ACM SIGCAPH Computers and the Physically Handicapped*, chapter Web usability and age: how design changes can improve performance, pages 30–37. Number 73-74 in ACM SIGCAPH Proceedings. ACM, New York, NY, USA.

Chernichovsky, D. and Markowitz, S. (2004). Aging and aggregate costs of medical care: conceptual and policy issues. *Health Economics*, 13:543–62.

Committee on Quality of Health Care in America: Institute of Medicine (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington D.C.: National Academies Press.

Craik, F. (2000). *Cognitive aging: A primer*, chapter Age related changes in human memory, pages 75–92. Philadelphia: Psychology Press.

Czaja, S., Charness, N., Fisk, A., Hertzog, C., Nair, S., Rogers, W., and Sharit, J. (2006). Factors predicting the use of technology: Findings from the Center for Research and

Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging*, 21(2):333–52.

Czaja, S. and Lee, C. (2007). The impact of aging on access to technology. *Universal Access in the Information Society*, 5:341–9.

Czaja, S. J. and Hiltz, S. R. (2005). Digital aids for an aging society. *Communications of the ACM*, 48(10):43–4.

Demiris, G., Finkelstein, S., and Speedie, S. (2001). Considerations for the design of a web-based clinical monitoring and educational system for elderly patients. *Journal of the American Medical Informatics Association*, 8:468–72.

Dickinson, A., Arnott, J., and Prior, S. (2007). Methods for human–computer interaction research with older people. *Behaviour & Information Technology*, 26(4):343–352.

Dickinson, A., Eisma, R., Gregor, P., Syme, A., and Milne, S. (2005). Strategies for teaching older people to use the world wide web. *Universal Access in the Information Society*, 4(1):3–15.

Dray, S. and Siegel, D. (2004). Remote possibilities?: international usability testing at a distance. *interactions*, 11(2):17.

Ericsson, K. and Simon, H. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3):178–186.

Estabrooks, P. A., Nelson, C. C., Xu, S., King, D., Bayliss, E. A., Gaglio, B., Nutting, P. A., and Glasgow, R. E. (2005). The frequency and behavioral outcomes of goal choices in the self-management of diabetes. *Diabetes Educ*, 31(3):391–400.

Fisk, A., Rogers, W., Charness, N., Czaja, S., and Sharit, J. (2009). *Designing for Older Adults: Principles and Creative Human Factors Approaches*. Human Factors and Aging Series. Boca Raton: CRC Press, 2nd edition.

Fox, S. (2006). Are "wired" seniors sitting ducks? Retrieved March 8, 2009 from `http://www.pewinternet.org/PPF/r/180/report_display.asp`.

Goldberg, H., Lessler, D., Mertens, K., Eytan, T. A., and Cheadle, A. (2004). Self-management support in a web-based medical record: A pilot randomized controlled trial. *Joint Commision Journal on Quality and Safety*, 30(11):629–635.

Gray, A. (2005). Population ageing and health care expenditure. *Ageing Horizons*, 2:15–20.

Handley, M., MacGregor, K., Schillinger, D., Sharifi, C., Wong, S., and Bodenheimer, T. (2006). Using action plans to help primary care patients adopt healthy behaviors: A descriptive study. *journal of the American Board of Family Medicine*, 19(3):224–31.

Hendrix, C. (2000). Computer use among elderly people. *Computers in Nursing*, 18(2):62–8.

Jastrzembski, T., Charness, N., Holley, P., and Feddon, J. (2005). Input devices for web browsing: age and hand effects. *Universal Access in the Information Society*, 4:39–45.

Jenson, J. (2007). Health care spending and the aging of the population. Technical Report (RS22619), Congressional Research Service, Washington, DC.

Jimison, H., Pavel, M., Bissell, P., and McKanna, J. (2007a). A framework for cognitive monitoring using computer game interactions. In Kuhn, K., Warren, J., and Leong, T.-Y., editors, *Proceedings of the 12th World Congress on Health (Medical) Informatics*, volume 129, pages 1073–7.

Jimison, H., Pavel, M., Larimer, N., and Mullen, P. (2007b). A general architecture for computer based health coaching. In *Proceedings of the International Conference on Technology & Aging, Toronto*.

Jimison, H., Pavel, M., McKanna, J., and Pavel, J. (2004). Unobtrusive monitoring of computer interactions to detect cognitive status in elders. *IEEE transactions on information technology in biomedicine*, 8(3):248–252.

Jones, S. and Fox, S. (2009). Generations online in 2009. Retrieved May 15, 2010 from `http://www.pewinternet.org/PPF/r/275/report_display.asp`.

Locke, E. A. and Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. a 35-year odyssey. *Am Psychol*, 57(9):705–17.

MacGregor, K., Handley, M., Wong, S., Sharifi, C., Gjeltema, K., Schillinger, D., and Bodenheimer, T. (2006). Behavior-change action plans in primary care: A feasibility study of clinicians. *journal of the American Board of Family Medicine*, 19(3):215–23.

MacGregor, K., Wong, S., Sharifi, C., Handley, M., and Bodenheimer, T. (2005). The action plan project: discussing behavior change in the primary care visit. *Annals of Family Medicine*, 3 Suppl 2:S39–40.

Marquié, J., Jourdan-Boddaert, L., and Huet, N. (2002). Do older adults underestimate their actual computer knowledge? *Behaviour and Information Technology*, 21(4):273–80.

McCray, A. T. (2005). Promoting health literacy. *J Am Med Inform Assoc*, 12(2):152–63.

McFadden, E., Hager, D., Elie, C., and Blackwell, J. (2002). Remote usability evaluation: Overview and case studies. *International journal of human-computer interaction*, 14(3):489–502.

Mead, S., Batsakes, P., Fisk, A. D., and Mykityshyn, A. (1999). Application of cognitive theory to training and design solutions for age-related computer use. *International Journal of Behavioral Development*, 23:553–573.

Nielsen, J. (1993). *Usability Engineering*. Academic Press, Inc., Cambridge, MA.

Nielsen, J. (1995). Applying discount usability engineering. *IEEE Software*, 12(1):98–100.

Nielsen, J. (2002). Usability for senior citizens. Retrieved March 8, 2009 from `http://www.useit.com/alertbox/seniors.html`.

Nielsen, J. and Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *CHI '93: Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, pages 206–213, New York, NY, USA. ACM.

Oregon Center for Aging and Technology (2010). Oregon Health & Science University's Oregon Center for Aging & Technology. `http://www.orcatech.org/`.

Ownby, R., Czaja, S., Loewenstein, D., and Rubert, M. (2008). Cognitive abilities that predict success in a computer-based training program. *The Gerontologist*, 48(2):170–80.

Petrie, H., Hamilton, F., King, N., and Pavan, P. (2006). Remote usability evaluations with disabled people. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1133–1141, New York, NY, USA. ACM.

Prochaska, J. and Velicer, W. (1997). Behavior change: The transtheoretical model of health behavior change. *American Journal of Health ...*, 12(1):38–48.

Prochaska, J. O., Butterworth, S., Redding, C. A., Burden, V., Perrin, N., Leo, M., Flaherty-Robb, M., and Prochaska, J. M. (2008). Initial efficacy of mi, ttm tailoring and hri's with multiple behaviors for employee health promotion. *Prev Med*, 46(3):226–31.

Rainie, L. (2010). Internet, broadband, and cell phone statistic. Retrieved May 15, 2010 from `http://www.pewinternet.org/Reports/2010/Internet-broadband-and-cell-phone-statistics.aspx`.

Reinhardt, U. (2003). Does the aging of the population really drive the demand for health care? *Health Affairs*, 22(6):27.

Shapiro, M. (2009). Can an automated tailored health messaging system be an effective tool for health coaching elders? Master's thesis, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, OR, 97239.

Spillman, B. and Lubitz, J. (2000). The effect of longevity on spending for acute and long-term care. *The New England Journal of Medicine*, 342(19):1409.

Stone, D., Jarrett, C., Woodroffe, M., and Minocha, S. (2005). *User Interface Design and Evaluation*. Morgan Kaufmann, San Francisco, CA.

Strecher, V. J., Seijts, G. H., Kok, G. J., Latham, G. P., Glasgow, R., DeVellis, B., Meertens, R. M., and Bulger, D. W. (1995). Goal setting as a strategy for health behavior change. *Health Educ Q*, 22(2):190–200.

The Institute of Medicine (2008). *Retooling for an aging America: Building the health care workforce*. The National Academies Press, Washington, DC.

The National Institute on Aging (2007). Why population aging matters: A global perspective. Technical Report 07-6134, National Institute on Aging, National Institutes of Health, The U.S. Dpartment of Health and Human Services, and the U.S. Department of State, Washington, DC.

The National Institute on Aging and the National Library of Medicine (2002). Making your website senior friendly: A checklist. Retrieved March 9, 2010 from `http://www.nlm.nih.gov/pubs/checklist.pdf`.

The World Wide Web Consortium (2008). Web content accessibility guidelines (wcag) 2.0. Retrieved May 10, 2010 from `http://www.w3.org/TR/WCAG20/`.

Thompson, K. E., Rozanski, E. P., and Haake, A. R. (2004). Here, there, anywhere: remote usability testing that works. In *CITC5 '04: Proceedings of the 5th conference on Information technology education*, pages 132–137, New York, NY, USA. ACM.

Trickett, S. B. and Trafton, J. G. (2009). *The PSI Handbook of Virtual Environments for Training and Education*, volume 1, chapter A Primer on Verbal Protocol Analysis, pages 332–346. Praeger Security International, Westport, CT.

Tullis, T., Fleischman, S., McNulty, M., Cianchette, C., and Bergel, M. (2002). An empirical comparison of lab and remote usability testing of web sites. In *Usability Professionals Association Conference*.

Walker, N., Philbin, D. A., and Fisk, A. D. (1997). Age-related differences in movement control: Adjusting submovement structure to optimize performance. *Journal of Gerontology: Psychological Sciences*, 52B:40–52.

Yu, C.-H., Chan, M., Hatt, W. J., and Jimison, H. B. (In Press). Usability testing applied to a health coaching interface for older adults. In *International Society for Gerontechnology 7th World Conference*. International Society for Gerontechnology.

Zaphiris, P., Kurniawan, S., and Ghiawadwala, M. (2007). A systematic approach to the development of research-based web design guidelines for older people. *Universal Access in the Information Society*, 6(1):59–75.

Zaphiris, P., Pfeil, U., and Xhixho, D. (2009). *Universal Access in Human-Computer Interaction. Addressing Diversity*, chapter User Evaluation of Age-Centred Web Design Guidelines, pages 677–686. Berlin: Springer-Verlag.

Zweifel, P., Felder, S., and Meiers, M. (1999). Ageing of population and health care expenditure: a red herring? *Health Economics*, 8(6):485–496.