Comparison of Various Display Representation Formats for Older Adults Using Inlab and Remote Usability Testing.

Sajitha Narayan

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Industrial and Systems Engineering

Dr. Tonya L. Smith-Jackson, Chair
Dr. Maury Nussbaum, Member
Dr. Mary Beth Rosson, Member

March 29, 2004
Blacksburg, Virginia

Comparison of Various Display Representation Formats for Older Adults Using Inlab and Remote Usability Testing.

Sajitha Narayan

ABSTRACT

The population of seniors is growing and will continue to increase in the next decade. Computer technology holds the promise of enhancing the quality of life and independence of older people as it may increase their ability to perform a variety of tasks. This is true for elderly. By the year 2030, people age 65 or older will comprise 22% of the population in the United States. As the population shifts so that a greater percentage are middle-aged and older adults, and as dependence on computer technology increases, it becomes more crucial to understand how to design computer displays for these older age groups.

The research has compared various display representation formats to try to find out which is the best way to represent information to seniors in any form of display and the reason for the preferences. The formats compared include high and low density screens for abstract icon representation, concrete icon representation, tabular representation and graphical representation. This research also endeavored to study the effectiveness of remote usability testing as compared to inlab testing for seniors.

Results indicated that density of screen is a very important factor affecting the performance of older adults. Density effect showed statistical significance $F\ (1,112)=8.934$, $p< .05$ from further post-hoc analysis that was conducted. Although significant results were not obtained, different formats of display representations may still be an area worth pursuing. Also it was noted that remote usability testing is not as effective as inlab testing for seniors in terms of time taken to conduct the study and the number of user comments collected. Implications, as well as recommendations and conclusions, of the study are presented.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents, who have been very encouraging and always provided me moral and financial support, to overcome all the difficulties.

I would like to thank all of my committee members for their advice and support. I would like to thank my chair Dr. Tonya L. Smith-Jackson for her unwavering guidance and spirit of enthusiasm in assisting me throughout this time. She also provided the initial idea and motivation to pursue this research. We had numerous discussions and brainstorming sessions, which helped shape this work and provided me a clear idea of what was expected out of my thesis. Her ability to constantly help and encourage those around her has been truly inspiring. I like to thank Dr. Maury A. Nussbaum for his expertise and encouragement in his role on my thesis committee. His dedication to helping students remains unparalleled. I wish to thank Dr. Mary Beth-Rosson for her support and assistance. I am very grateful that I was blessed with such a wonderful committee.

I extend my gratitude to Dr. Misha W. Vaughan and Dr. Delia Grenville for their support and motivation, which helped me to get back on track when I had let my busy schedule get the best of me. Thanks to Dr. Joe. H Goldberg, for explaining statistic calculations. Thanks to my colleague, Miranda Capra, who helped me set up the lab server for my study and helped me with it whenever I needed any clarification. Thanks to Dr. Hartson and his student Narayanan for letting me use the Remote Evaluation tool which they developed and helping me make changes to it as when I had to. I thank Paul Smeal, member of the Blacksburg Senior Center, for allowing me to work with the center during the course of the study. I thank all my participants that helped me with this research and endured those long laboratory sessions. And, last but not least, I thank God for blessing me with a life filled with such nice people.

DEDICATION

I dedicate this work to my parents, my greatest source of endless support and inspiration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Growing Population of Elderly

The number of seniors in the population has been growing and will continue to increase in the next decade. By the year 2030, people age 65 or older will comprise 22% of the population in the United States. This proportional change represents a 10% increase since 1990 (Czaja, 1997). Seniors are one of the fastest growing demographics. The United States alone has an estimated 4.2 million Internet users over the age of 65.

1.2. Need for Independence

At the same time that the population is aging, technology is rapidly being integrated into most aspects of life and changing the nature of work, the form and scope of personal communication, education and healthcare delivery systems. Some form of computer technology is commonplace in most environments including the home. Many routine activities such as banking, information retrieval, bill paying, and shopping increasingly involve use of computers. In other words, it is highly likely that older people will need to interact with some form of computer technology to carry out routine activities (Czaja, 1997).

Further, computer technology holds the promise of enhancing the quality of life and independence of older people as it may augment their ability to perform a variety of tasks. This is true for elderly who are frail and homebound. Hendrix (2000) says that use of computers helps to augment self-esteem in older adults. It is evident that older people prefer to stay independent and have a good quality of life. White & Weatheral (2000) stated that computer use appears to promote positive relations with family and friends.

Computer applications such as electronic mail will allow this population to perform tasks they were previously unable to perform such as maintaining social interaction. Computer technology is coupled with a positive shift in computer attitudes while decreasing feelings of depression and isolation (Kelley, Morrell, Park, & Mayhorn, 1999). It is also seen that although they are typically retired, seniors lead very active lives and often have great interest in modern technologies such as the Internet, which gives them another method to communicate and stay informed (Nielson, 2002).

1.3. Improving Quality of Life

It is important to note that computerized systems are so common that older adults may not be able to avoid them and is more important to note that computer use has the potential to significantly improve the quality of life of older adults (Mead, Batsakes, Fisk, and Mykityshyn, 1999). Older adults gain many benefits as a result of interacting with computers (Hendrix, 2000; Morrell & Echt, 1997).

The dominant approach to design in ergonomics is to act behalf of the users. Specific users, tasks and environments often have to be studied before design conclusions appropriate to a particular setting can be drawn. Shackel (1984) stated that there is no manual that is complete and generally accepted on how to design good human factors into computer systems. There are many aspects to usability, which must be taken into account if a system is to be good.

There are four principles of system design in addressing each aspect of usability (Gould & Lewis, 1983; 1985). The first one is early- continual –focus on users. The second is early- and continual –user testing. The third is iterative design and the forth is integrated design. Therefore the first step in designing a system is to decide who the users will be and what they will be doing with the system. The user population may be broader or heterogeneous, in that case the system will have to be tailored and tested for the other groups as well, and design tradeoffs may be required.

The process described in Figure 1 can be applied to many issues, even information display design. In this strategy, the ergonomist acts as an active agent, supporting the user community as it tries to determine the way forward by injecting theories, knowledge, methods and data into the process where it is most appropriate. The two aims are to ensure that the users have the opportunity to take those decisions that are important to them and to make sure that they do so in as informed a way as possible.

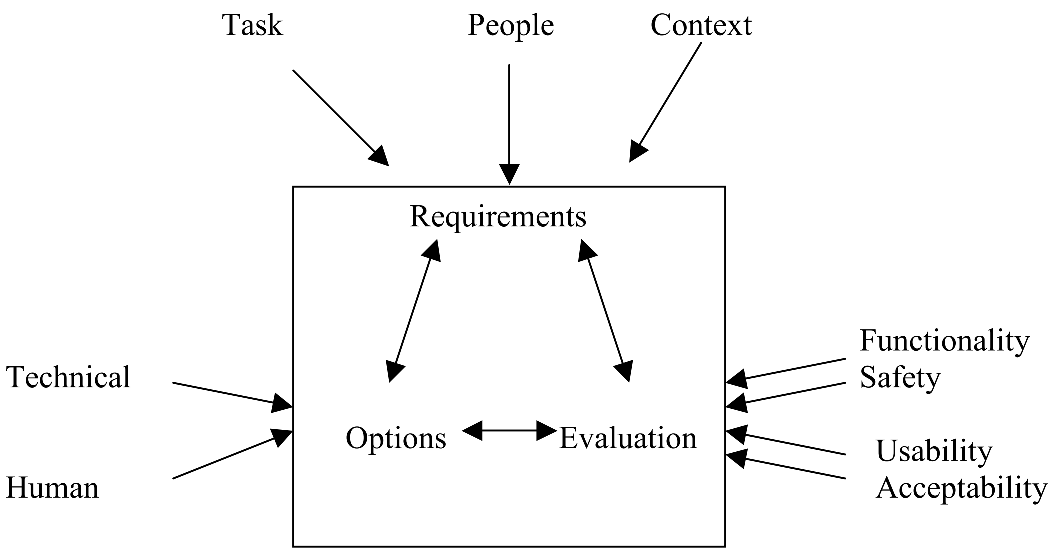


Figure 1: Design for the users by the users (Source: Eason, 1995, p. 1672. Ergonomics).

1.4 Universal Design

Universal Design is a design that recognizes, respects, values, and attempts to accommodate the broadest possible range of human abilities, skills, requirements, and preferences in the design of all computer based products and environments. This promotes a design perspective that eliminates the need for special features and fosters individualization and end-user acceptability (Stephanidis et al., 1998; Story, 1998).

Universal design is very broad, so it covers accessibility challenges that diversity poses in (a) the target user population (including people with special needs) and the individual and cultural differences, (b) the scope and nature of tasks (especially as related to the shift from business tasks to communication and collaboration intensive computer mediated human activities), and (c) the technological platforms and associated devices through which information is accessed (Stephanidis & Salvendy, 1999).

Jacob Nielson (2002) stated that websites tend to be produced by young designers, who often assume that all users have perfect vision and motor control, and know everything about the web. These assumptions rarely hold, when the users are seniors. It is clear from previous research that seniors are more affected by usability problems than younger users. Among the obvious physical attributes often affected by the human aging process are eyesight, precision of movement and memory which are very important features for processing information from displays. Human information input and processing operations depend on the sensory reception of relevant external stimuli. Display is a term that applies to any indirect method of presenting information such as highway traffic sign, charts, graphs, labels, and computer displays (Sanders and McCormick, 1993).

1.5. Special Needs for Older People.

As age increases, functional limitation also increases. At age 15-24 yrs, functional limitation is 5.3 % of which severe functional limitation is 1%. But at age 75+ functional limitation increases to 72.5 % of which severe functional limitation is 41 %. Functional limitation exists when a desired or required activity cannot be independently performed in a specific environment. The most common problem seniors face as they age is natural deterioration of the eyesight (Deborah, 1996). As they age and spend increasingly large amounts of time in front of a computer screen, it is no wonder that eyestrain and eye fatigue become reality for many.

Deborah, (1996) also states that by age 65 of most people have lost at least some of their ability to focus, resolve images, distinguish colors and adapt to changes in light. Deborah also found that in the United States alone more than 10 million Americans have significant vision impairments, and at least 3 million have partial sight loss and over 60 percent of those considered visually impaired are older persons. As part of the natural aging process and longevity, the need for contrast increases because of the discoloration in the eye fluids and lens. The amount of light that passes through the eye is reduced for the seniors by common impairments such as cataract or clouding of the lens. Most old people have a loss in color perception that accompanies their dimmed vision.

1.6. Research Support and Justification

Many previous researchers have stated that because technology is becoming so embedded in our daily living, if older adults are to continue working independently, our society should find usable computer interfaces (Charness, Kelley, Bosman, and Mottram, 2001). Some previous researchers noted that many companies actually prefer to have older adults working for them because they are more likely to possess characteristics like responsibility, honesty, and wisdom (Carter and Honeywell, 1991).

White and Weatherall (2000) indicate that limited research has been done investigating older adults and Information Technology. Despite the literature available that examines individual strategies and suggestions from researchers in the areas of technology for older adults, there have been few attempts to utilize all of this literature for a common purpose. Therefore the need for this research is commonly noted, but there is a gap in the research and empirical data in the area. Integrating the current literature is required to identify usable information display formats that are likely to be the most beneficial for older adults.

There is a risk of misjudging the capability of nearly half of future users if designs are only tested on younger users because the effects of age become noticeable from the mid-forties onward (Hawthorn, 2000). Walker, Philbin, & Fisk (1997) successfully undertook a theory-driven design of interfaces for older adults that greatly improved older adults' performance and simultaneously benefited younger adults' performance. This is consistent with Morrell and Echt (1997), who states that even if they focus specifically on older adults, it is likely that senior friendly guidelines will extend to computer users of other ages. According to Czaja (1997) in order to maximize the benefits of computer technology for older adults and optimize their interactions, older adults must be included in system design and evaluation efforts. Czaja (1997) also states that this participatory approach will not only serve to benefit older adults, but all potential users of computer systems.

This research will help in designing interfaces which are more easily understood by seniors and there by improving the quality of life and independent living of older adults. The benefits of this research will not only be applicable to one display format such as a smart house interface but also to other interfaces such as cell phones, PDA's, road signs and websites. The results of the study can also be applied to the design of information displays for the younger population. Finally in a more business-oriented view, this research will allow businesses to increase their consumer market.

2. LITERATURE REVIEW

2.1. Technology and Older Adults

Research has been done for several technologies for older adults such as Automatic Teller Machines, Smart houses and Accessible remote controls. There is some literature that suggests that older adults are less likely to use new technology, relative to younger adults (Gilly & Zeithaml, 1985; Rogers, Cabrera, Walker, Gilbert, and Fisk, 1994 in Rogers, Gilbert and Cabrera, 1994). Although older adults may use newer technologies less than young adults, there are certainly a number of older adults who are willing to use it. Visual impairment is a problem that disrupts the ability to use technologies among older adults.

Visual impairment is one of the most prevalent conditions in adults age 65 and over. One of the reasons why older adults find it difficult to use ATMs is difficulty in seeing the screen (Rogers, Gilbert, Cabrera, and Fisk, 1994 in Rogers, Gilbert and Cabrera, 1994).According to Berge and Cassells (1990), if you define poor vision as 20/50 acuity or worse, 11% of the adults aged 65 to 73 who wear glasses are impaired and also 26% of the adults aged 65 to 73 who do not wear glasses are also impaired. Lin, Williges and Beaudet (1995) conducted a study for accessible remote controls for older adults who tend to complain about small controls and lettering of the remote control unit. These researchers found that the readability of labeling should be improved with enlarged label and label background contrast should be enhanced with high contrast labels.

Scheiber and Kline (1994) conducted a study on age differences in the legibility of symbol highway signs as a function of luminance and glare level. It has been shown repeatedly that older subjects have shorter legibility distances for symbol and text signs than their young driver counterparts. Therefore they stated that legibility distances should be determined using a representative sample of older drivers, not only young drivers with visual acuity of 20/20.

According to a compendium of 26 studies, from 20-60 years of age, simple reaction time slows by 20 % (Birren, Woods, and Williams, 1980). Birren (1965, 1974) has stated that as a consequence of their slowness of behavior, older adults may be living in a qualitatively and functionally different environment than younger people. Research shows that most of the elderly strongly prefer to remain at home, but many will require some form of aid.

To improve the quality of life for elderly and disabled, a smart multi sensor system has been developed (Chan, Hariton, Ringeard, and Campo, 1995). This is based on advanced telecommunication and information technology. Such a system will help to monitor them and also to bring them more security and safety without disturbing their life. “The elderly are going to be an enormous slice of the total population, (Larson, 2001, p.1)”. Intel sponsors some smart-home research projects directed toward older people.

Kwahk, Smith-Jackson and Williges (2002) described a Smart House as a sensor- based intelligent home designed to monitor the health and well being of senior residents. This information can be made available to all family members and healthcare providers who stay away from the seniors by a variety of interfaces. Therefore the Smart House can be called a friendly and intelligent assistant to support the older resident’s daily activities. Kwahk et al., (2002) conducted a study to design an adaptive telemedical support system based on Smart House Technology to display tailored medical information to variety of user groups. These researchers also found that an interface has to be designed to display information about the activity level, eating pattern, nutritional intake to the senior residents and also to their remote family members and healthcare providers. Therefore the interface development of the Smart House equipped with a telemedical support system is very important recently and has been receiving a lot of consumer attention (Tang and Venables, 2000). It is imperative that consumer products are designed so that they are accessible to all the populations including seniors.

A three-phase methodology was created for designing consumer products for individuals with special needs using human factors design approaches (Elkerton, Williges, and Williges, 1987; Hartson and Hix, 1993). This method allowed designers to identify accessibility barriers on a product for users with special needs, select and implement design solutions for eliminating or mitigating these barriers and evaluate empirically the accessibility and usability of design alternatives for the target user population before product production. Lin, Williges, and Beaudet (1995) state that until products are designed for accessibility, individuals with visual impairments will continue to be excluded from enjoying any of the potential benefits such technology can provide. But to produce usable display formats for seniors, we must first understand display design and identify user requirements.

2.2. Concrete and Abstract Icon Representations

During the past decade, graphical user interfaces have become a common method of interaction between the user and the computer in end user application programs, such as word processors, spreadsheets, databases and electronic mail. Wiendenbeck (1999) stated that graphical user interfaces present information to the user by icons rather than textual descriptions. He defines icons as interface objects that represent in a simplified pictorial fashion larger, more complex and harder to grasp system objects. Icons are used in the interface because they are presumed to facilitate human computer interaction. There are many claims that icons are easier to process than text (Shepard, 1967). Some claims regarding icons made by previous researchers (Horton, 1994) are the following:

- Icons improve the productivity and reliability of work.
- Icons are better than words for representing subtle visual and spatial concepts.
- Well-designed icons save space.
- Icons decrease search time.
- Icons lead to immediate recognition.
- Icons lead to better recall.
- Icons reduce the necessity of reading.
- Icons make interfaces more international.

According to Gittens (1986), icons can represent not just the referent but its attributes, associations and state. Icons show self-contained presentation of process in underlying systems. Some of them serve as visual indication of the process and others change their form dynamically to reflect changes in process, thus showing the state of the system. Icons also show the attributes of the referent such as the color and shape. When the users look at the icon, they can infer what its attributes are as well, thus showing associations. For e.g., in an interface if email is represented as mail trays and arrows pointing into and out of the trays, the user can infer it as incoming mails and outgoing mails. This can be used without prior knowledge.

Detweiler and Familant (1993) make the point that all iconic depictions are abstractions that represent only a subset of features of the referent. There are several icon classifications in the literature (Blattner et al., 1989; Detweiler and Familant, 1993; Gittens, 1996; Rogers, 1989). Blattner et al. (1989) classify icons as representational (concrete), abstract and semi-abstract.

Representational icons are simplified images of familiar objects or operations. Representational icons are concrete icons and are meant to make the mapping icon and referent obvious. But Wiedenbeck (1999) states that there is no obvious pictorial representation or none that can reasonably be implemented. The abstract icons employ geometric shapes or graphic symbols instead of concrete images, for example an exclamation point to represent the operation of executing a program. Blattner et al. (1989) defines semi abstract icons as a combination of representational pictorial element with an abstract symbol, for example like a folder with an arrow that indicates placing items in it.

Many previous researchers (Benbasat and Todd, 1993; Gittens, 1986; Rogers, 1989; Shneiderman, 1997) have summarized many advantages and possible disadvantages of iconic representations over text in terms of human computer interaction. Shepard (1967) showed that recognizing visual images is easier than identifying words or sentences. Standing (1973) showed that humans have an indefinite capability to recognize images they have seen before. Previous researchers have found advantages of visual images over text in simple recognition tasks but yet it is not sure whether there are significant advantages of visual images over text in human computer interaction tasks of realistic complexity.

Benbasat and Todd (1993) propose that use of icons will help the users to devote more cognitive resources to the primary task they are trying to accomplish. Navon (1984) asserts that for different activities such as perception and cognition, there are different resources available. If the primary task such as problem solving requires cognitive resources, then icons which require perceptual resources are better than text because icons make more resources available for the primary task. On the other hand text based interface will use up all the cognitive resources.

According to assimilation theory (Ausubel, 1968 in Wiedenbeck, 1999), an interface that represents an object in a concrete manner will help the individual in assimilating new computer concepts to a correlated base of concepts in the memory. Wiedenbeck (1999) states that this assimilation is possible because of the analogies of the new situation to the known situation. Carroll and Mack (1985) have argued that analogies are important in computer learning because they help users in assimilating new information to related existing information.

According to Larkin and Simon (1987), there may be fundamental reasons why visual representations are better than verbal descriptions. They have stated that icons group together information that describes an individual element and support the making of a large number of perceptual inferences. Detweiler and Familant (1993), state that icons should be immediately identified by the user population. Hutchins et al. (1985) defines articulatory distance of an icon as the difficulty of inferring the referent from the icon. He also mentions that icons that have greater representational content or are concrete and have more direct mapping between form and function are expected to be easier to infer. Abstract icons do not have direct mapping between the pictorial representation and referent.

Icons will increase acceptance of the system if they give an immediate impression that the system is easy to use to the first time user (Rogers, 1989). The perceptions about the ease of use of the system are affected by surface features of the system such as the use of icons in the interface, the color and information presentation formats (Davis, 1989; Davis et al. 1989). Davis also suggests that icons are important for initial acceptance and formation of positive perceptions about ease of use and usefulness of the interface.

Blankenberger and Hahn (1991) and Rogers (1989) state that icons depicting concrete objects tend to be most effective, which is consistent with Hutchin's et al. argument previously mentioned. Rogers (1989) has also stated that people can learn the arbitrary relationships of abstract icons, but only when few are being used, but not when they were embedded in a whole set of abstract icons. In some cases, abstract icons have been more successful than concrete icons (Arend et al.1987). This could be related to the findings of Bewley et al. (1983) that icon sets with more visual variety are easier to locate in a display.

Positional consistency (Green and Barnard, 1996; Lansdale, 1988) and visual simplicity (Byrne, 1993) of the icons on screen are very important and have a strong effect on usability. Several studies have been conducted comparing icons, text and combinations of both for software applications. Mutter and Mayson (1986) have conducted a study comparing text only menus to menus with text plus concrete graphics representing information categories in the system by exemplars. His study showed no time difference in answering questions, but an advantage in accuracy for the text plus graphics condition. Some researchers who also have done studies on icons have found that icons with text labels are easier for participants to learn to recognize correctly and are better than icon only and text only conditions (Bewley et al., 1983; Byrne, 1993; Egidio & Patterson, 1988). Studies conducted by Wiedenbeck (1999) also showed that there is a significant advantage for text only and icon-text interfaces compared to icon only interface for correctness, time and use of help.

Several researchers (Benbasat & Todd, 1993; Rohr & Keppel, 1984) have also produced opposite results. They conducted tests comparing text only condition to icon only condition. The results were such that there were no difference between the iconic version of the interface and the text version. This could be because of the system used in the experiment and could have been because the system is less complex and so less number of icons to interpret and less screen scanning required (Wiedenbeck, 1999). So for designing alphanumeric displays there are several aspects to be considered.

2.3. Screen Design Issues

Several guidelines focus on alphanumeric display formats (Brown, 1988; Galitz, 1985). These guidelines are summarized below.

2.3.1. Density

According to Sanders and McCormick (1993), overall density of information displayed on a screen is usually expressed in terms of the percentage of available character spaces being used. There are 1920 available spaces on an 80 character by 24-line screen. Tullis (1988) stated that the relationship of overall density of the screen and search performance is generally consistent, as long as necessary information is present, the search time and errors increase with increasing density. Tullis (1984) analyzed more than 600 data screens and found that mean overall density was about 25 percent of the data screen and that displays with densities higher than 40 or 50 percent of the data screen were relatively rare. Avoiding unnecessary information, using concise wording, and using tabular formats with column headings should minimize density. Tullis (1983), discusses a concept of local density, and proposes that starting with low local densities and then increasing density raises performances however at high local densities, increasing density will degrade performance. This means that if a small amount of information is spread all over the screen, search time to find one item will be long, but if it is little more dense, search time can be improved. Eventually, the information will become too crowded and performance will deteriorate.

2.3.2. Grouping

Grouping is defined as the extent to which data items form well-defined perceptual groups. Tullis (1986a) established that for displays where the average size of data groups subtended is less than 5 degrees of visual angle, search time was a function of the number of groups. The more the number of groups, the more time it takes to find a piece of information (Sanders and McCormick, 1993). For displays with average group sizes greater than 5 degrees of visual angle, search time was a function of the group sizes. Tullis recommends that it is a best approach to minimize the number of groups by making each one as close to 5 degrees in size as possible.

2.3.3. Complexity

Tullis (1983) defines layout complexity as the extent to which the arrangement of items on screen follows a predictable visual scheme. He also says that the best way to reduce the complexity of layout is to align the information in distinct columns. Previous research (Wolf, 1986, in Sanders and McCormick, 1993) says that the search times are faster for items which are arranged in columns of text rather than when material is arranged as a horizontal list of running text.

2.3.4. Highlighting

Using highlighting on a display to reduce search time is the common perception, but then its not always true (Fisher and Tan, 1989). According to these researchers highlighted displays may be no better than unhighlighted screens and could be even worse. Validity of highlighting, that is, the percentage of times the target being searched for is highlighted, is an important factor. The higher the validity of highlighting, the greater the advantage of highlighting (Fisher and Tan, 1989). Sanders and McCormick (1993) suggest that of all the ways of highlighting on a screen, blinking, except for only critical urgent information should be avoided. This is because blinking is hard to read, annoying and distracting. These researchers say that blinking should be used sparingly and only for urgent messages and warnings.

2.4. Graphical Representation

2.4.1. Graphical Representation of text

A graph is a paper or electronic representation of text or numeric data. Often because of language differences, instructions and procedure manuals use pictures or drawings such as graphs instead of words. Pictorial information is important for speed, but text is more important for accuracy (Booher, 1975; Fisk, Scerbo, and Kobylak, 1986 in Sanders and McCormick, 1993). Sanders and McCormick (1993), suggest that the general recommendation for any instructional material is to combine pictures with text for speed, accuracy and long-term retention.

2.4.2. Graphical Representation of data

The most common graphs include line graphs, pie charts and bar charts. The judgments people make in extracting information from graphs may be biased (Gillian, etal.1998). People tend to systematically overestimate or underestimate quantities relative to their true values. Some of these biases are associated with optical illusions that warp our sense of perception. These are some of the aspects of information display design, which have to be considered while designing displays or interfaces for seniors. For usability testing of the display formats the seniors can be involved without putting too many demands on them if they can sit at home in their own environment and perform the test. So a discussion on remote usability evaluation is imperative.

2.5. Usability Evaluation

There are generally three types of usability evaluation methods: Testing, Inspection, and Inquiry (Nielson, 1993). In Usability Testing approach, representative users work on typical tasks using the system (or the prototype) and the evaluators use the results to see how the user interface supports the users to do their tasks. In Usability Inspection approach, usability specialists -- and sometimes software developers, users and other professionals -- examine usability-related aspects of a user interface. In Inquiry, usability evaluators obtain information about users' likes, dislikes, needs, and understanding of the system by talking to them, observing them using the system in real work (not for the purpose of usability testing), or letting them answer questions verbally or in written form.

2.6. Remote Usability Evaluation

Remote evaluation is defined as usability evaluation where evaluators are separated in space and /or time from users (Hartson, Castillo, Kelso, Kamler and Neale, 1996). In traditional laboratory –based usability evaluation, users are observed directly by evaluators. In remote usability evaluation, the opportunity for direct observation of users is absent. Here the computer network serves as a bridge between users and evaluators, thereby taking the interface user evaluation to a broad range of users in their natural work settings or home (Castillo, Hartson and Hix, 1997).

Types of remote evaluation methods identified and distinguished by Castillo, Hartson and Hix (1997) are commercial usability services, remote questionnaire or survey, collaborative remote evaluation, video conferencing supported evaluation, instrumented or automated data collection for remote evaluation, and the user reported critical incident method. Software application can be used to display a user questionnaire and to gather subjective user preference data about the application and its interface. The appearance of the questionnaire, requesting feedback concerning usability is all triggered by an event. The User Partnering Module from Up Technology (Albelow, 1993) uses event- driven triggers to awaken dialogues that ask users questions about their usage. The advantage of these remote questionnaires is that it captures remote user reactions while it is fresh but they would be answers only to prewritten questions by developers or evaluators (Castillo, Hartson and Hix, 1997).

The distributed location of users and evaluators yields the prospect for collaborative usability evaluation via the network (Hammontree, Weiler and Nayak, 1994). The evaluators and users are connected through the Internet and or phone line and uses commercially available software to help them during remote evaluation of the user interface. This method supports synchronous and asynchronous collaboration (Castillo, Hartson and Hix, 1997). Inlab testing uses video/audio cables to record it, but when remote testing is conducted, the evaluators and users have to be connected using the network and video conferencing software as an extension of video audio cables (Hartson et al., 1996).

The user reported critical incident method applies selective data collection triggered directly by users while performing tasks in their normal work context (Hartson et al., 1996). Users will be trained to identify critical incidents and report specific information about these events. The reports would be transmitted to developers along with context information about the user task and the screen where the problem occurred. Castillo, Hartson, and Hix (1997) state that the user-reported critical incident method for remote evaluation would be more cost-effective, since the user gives the data and evaluators look only at the data that relate to usability problems.

Johansen (1988) has found several different situations that can occur during task performance with remote user evaluation along the dimensions of time and user location.

Figure 2 characterizes user location and time of evaluation.

TIME OF EVALUATION

USER LOCATION	Different (Asynchronous)	Same (Synchronous)
User's own working environment	**User-reported critical Incident method** Instrumented or automated data collection Remote questionnaire or survey Collaborative remote Evaluation	Collaborative remote evaluation Video- conferencing supported evaluation
Controlled environment	N/A	Third-party laboratory evaluation Traditional lab Based Usability evaluation

Figure 2: Characterization of user location and time of evaluation. (Source: Castillo, Hartson, and Hix, 1997)

In the user reported critical incident method, all data collection, remote questionnaires and remote evaluations occur within the user's normal work setting. Evaluators participate at a different time than users. The user has to both identify and report critical incidents during task performance and the evaluator analyzes these reports to find out the usability problems (Castillo, Hartson, and Hix, 1997).

No added equipment is required for remote usability evaluation while inlab testing requires a high level of equipment and each provides a large quantity of data (Castillo, Hartson and Hix, 1997). Remote questionnaires and surveys require no added equipment but produce a small amount of data. The user reported critical incident method also requires less equipment and produces somewhat less data than traditional inlab testing.

When a cost analysis to collect data and analyze it was conducted by Castillo, Hartson and Hix (1997), the inlab testing method was clearly the one with a high cost to analyze and collect data. The user-reported critical incident method, the cost for collecting data are significantly lower than inlab testing and is approximately the same as the inlab testing for analysis because that is very similar. The cost to collect data and analyze it is very low for remote questionnaire or survey, since data is a small amount.

Castillo, Hartson, and Hix (1997) found that only the user reported critical incident method meets all the following criteria:

- Data are centered around critical incidents that occur during task performance,
- Tasks are performed by real users,
- Users are located in normal working environment,
- Users self report own critical incidents
- Data are captured in day to day task situations
- No direct interaction is needed between user and evaluator during an evaluation session,
- There is a cost-effective way to capture data,
- Data are high quality and therefore relatively easy to convert into usability problems

Training will be required for the users to recognize and report critical incidents as discussed by Hartson, Castillo, Kelso, Kamler, and Neale (1996). Some previous researchers have obtained information about pilot-error experiences in reading and interpreting aircraft instruments from people not trained in the critical incident technique. Castillo, Hartson, and Hix, (1997) state that traditional laboratory based formative evaluation has critical incident detection done by an evaluator who is trained in human–computer interaction, and this might lead to skepticism about casting a user in that role. But critical incident technique has been tailored for human computer interaction so that untrained users identify critical incidents during their own task performance (del Galdo, Nixon, Williges R.C., Williges B.H., 1986).

Bergel, Cianchette, Fleischman, McNulty, Tullis (2001), conducted an empirical comparison of lab and remote usability testing of websites. The study indicated that there is not much behavioral difference between test users indicated by task rates and task times. The remote users provided rich comments. They also stated that more reliable subjective data is provided in remote testing since more users can be included than lab tests. It was evident from their study that lab tests are more better to capture certain kinds of usability issues such as excessive scrolling and failure to see certain elements on screen at first, but the most significant usability issues for a website can be captured by both the methods.

Waterson, Landay and Matthews (2002), conducted a study on comparing inlab testing and remote testing for mobile devices. The study indicated that the browser related and device related issues were hard to find with remote testing. The study also revealed that it is difficult to capture the user reactions and curiosity to explore other areas of a site with remote usability testing.

2.7. Ultimate and Actual Criteria for Usability Evaluation Method (UEM) Effectiveness:

The realness attribute plays a pivotal role in several of the UEM measures. A usability problem is real if it is a predictor of a problem that users will encounter in real work context usage and that will have an impact on usability (Andre, Hartson and Williges, 2001). The ultimate criterion will be the real usability problems encountered. The actual criteria will be the usability problems encountered in each of the usability testing method. The realness of usability problems can also be determined by review and judgments of experts in which each candidate usability problem is examined by one or more usability experts and determined by some guideline to be real or not. Bastien and Scapien (1995) have stated three measures for examining an evaluation method. They are Thoroughness, Validity and Reliability. Thoroughness is achieved if the results obtained from each evaluation method such as inlab testing and remote usability testing bring as many existing problems as possible. Validity is achieved if the results obtained from the testing are real. Reliability is achieved when the results obtained by each method is consistent, independent of the individual performing the usability test. Below given is a figure 3, which shows the ultimate and actual criteria

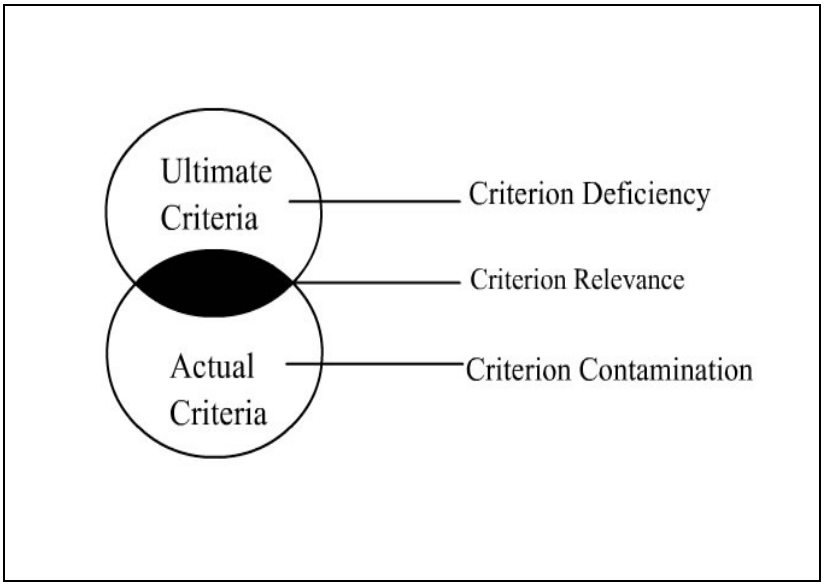


Figure 3: Ultimate and Actual criteria. (Source: Andre, Hartson, & Williges, 2001)

As discussed earlier several researches has been done in the field of display representation formats such as use of icons, graphical representation and different aspects such as highlighting, density, grouping and complexity. Research has been conducted on remote usability testing and remote critical incident reporting as well. But only few research studies have been conducted for interface designs for seniors. So there was a need to conduct research on comparison of display representation formats for seniors. Since many seniors prefer staying at home and working at home, it is also imperative that we study the effectiveness of remote usability testing in comparison with inlab testing. By assessment of past research, it was apparent that apart from individual motives, comparing information display formats and coming up with an appropriate display concept is beneficial for older adults and can be considered a positive ambition. If the reason to conduct research in this field was decided by financial aspects as usually the case when it comes to business, it still supported the claim that studies should be conducted in this area. Having discussed the pertinent points of prior research for the study, the next section will include an overview of the important aspects of the research study.

3. RESEARCH OVERVIEW

3.1. Research Purpose

3.1.1. Primary

The primary purposes of this research were to:

1) Compare information display formats for seniors by using an abstract, concrete, graphical and tabular interface in high and low density screens to evaluate the usability of display formats.
2) To understand the basis of user- preferences for display design.

3.1.2. Secondary

The secondary purpose was to examine the effectiveness of remote usability testing and compare it with inlab testing.

3.2. Research Questions

1. What is the difference in performance among older adults when the information is displayed in concrete form compared to abstract form?
2. What is the difference in performance among older adults when the information is displayed in graphical form compared to tabular form?
3. What is the difference in performance among older adults when the screen density is high than from when it is low?
4. What is the reason for user- preferences of each of these display formats and which is preferred more?

5. Which mode of usability testing is more effective in terms of time, data collected and preference when comparing remote usability testing to inlab testing?

3.3. Research Hypotheses

1. The concrete representation, when compared to the abstract representation, will be a more usable format of representation, as measured by performance, for older adults within the given interfaces. This will be in agreement with the argument previously mentioned of Blankenberger and Hahn (1991), Hutchin's et al and Rogers (1989) who stated that icons depicting concrete objects tend to be most effective. Results supporting this hypothesis will indicate that older adults perform better in an interface, which displays information in a more precise and concrete manner than in an abstract manner, which will take more time to process.
2. The tabular representation, when compared to the pictorial representation, will be a more usable format of representation, as measured by performance for older adults within the given display representation formats. This is supported by Tullis (1984), who stated that avoiding unnecessary information, using concise wording, and using tabular formats with column headings should minimize density and minimal density will reduce search time. The support for this hypothesis will imply that older adults perform better if information is displayed in a tabular form which is clearer than information displayed in graphs. This could be because graphical information takes more time to process.
3. The low-density interfaces, when compared to the high-density interfaces, will be a more usable interface, as measured by performance, for older adults within the given interfaces. As previously mentioned Tullis (1988) stated that the relationship of overall

density of the screen and search performance are generally consistent, as long as necessary information is present, the search time and errors increase with increasing density. Support for this hypothesis will indicate that older adults perform better in low-density interfaces than high-density interfaces for any form of information display format. This can be generalized to any form of information display format.

4. The reasons for user-preference of these display formats will be the usability of the displays, the ease of comprehension, the ease of use and ease of mapping as measured by older adults with the given interfaces.
5. The remote usability testing will be as effective as inlab testing as measured by time, for older adults within the given interfaces. As previously mentioned Bergel, Cianchette, Fleischman, McNulty, Tullis (2001) stated that there is not much behavioral difference between test users indicated by task rates and task times. Support for this hypothesis will indicate that older adults will perform remotely as well as they will when they are in a lab.

The variable "usable" indicates that the information display representation is easier to comprehend, meets expectation, and is easier to use (Nielson, 2002).

4. METHODOLOGY

4.1. Experimental Design

A 2 x 2 x 4 mixed factor design was used in the study. The main factors were the display representation formats, screen density, and usability testing methods. The display representation formats is a within subject factor and screen density is a between subject factor since the screens that were used for testing display representations had different content while the screens used for testing screen density had the same content. The order of presentation of displays was balanced using counterbalancing. Four presentation orders were used, consisting of the 4 display representations and the 32 participants were exposed to each order in such a way that every 8 participants were exposed to one particular order. Several iterations were conducted to evaluate and redesign the method based on results of the pilot study. Further information about the levels and types of these factors is provided in Table 1. The data matrix illustrating the design is given in Table 2.

Table 1: Factor Levels and Types

Factor Name	Levels	Type
Display representation format (A)	Abstract, Concrete, Graphical, Tabular	Within-Subject, Fixed Effects
Screen density (B)	High and Low	Between-Subject, Fixed Effects
Usability testing method (C)	Remote testing and Inlab testing	Between- Subject, Fixed effects
Subjects (S)	$S_1 \ldots S_{32}$	Between-Subject, Random Effects

Table 2: Data Matrix

	Usability Testing method			
	Remote testing		Inlab testing	
	Screen Density			
Display Representation Format	*Low Density*	*High Density*	*Low Density*	*High Density*
Abstract	S_1 S_2 S_3 … S_8	S_9 S_{10} S_{11} … S_{16}	S_{17} S_{18} S_{19} … S_{24}	S_{25} S_{26} S_{27} … S_{32}
Concrete	S_1 S_2 S_3 … S_8	S_9 S_{10} S_{11} … S_{16}	S_{17} S_{18} S_{19} … S_{24}	S_{25} S_{26} S_{27} … S_{32}
Pictorial	S_1 S_2 S_3 … S_8	S_9 S_{10} S_{11} … S_{16}	S_{17} S_{18} S_{19} … S_{24}	S_{25} S_{26} S_{27} … S_{32}
Tabular	S_1 S_2 S_3 … S_8	S_9 S_{10} S_{11} … S_{16}	S_{17} S_{18} S_{19} … S_{24}	S_{25} S_{26} S_{27} … S_{32}

4.2. Independent Variables

4.2.1. Display Representation Format

The four levels of the variable were abstract, concrete, graphical and tabular. Figures 4, 5, 6 and 7 show examples of the various low density representations. In abstract information representation, there were abstract symbols to depict particular information. In concrete information representation, the information was represented in a concrete manner by using concrete symbols and text. In both cases, the information used was nutritional summary information for a resident in a smart house. The nutritional summary information gathered by Kwahk, Smith-Jackson and Williges (2002) for the Senior Health watch project was used for the computer displays so that the information used for the interfaces for the study was pertinent to the seniors.

Graphical representation used graphs to show particular data. The tabular representation depicted the same information using tables. In both these cases the information used were eating pattern summary of a resident in a smart house. The eating pattern summary information gathered from sensors in a smart house prototype located at University of Virginia for the Senior Healthwatch project was used so that the information used is authentic.

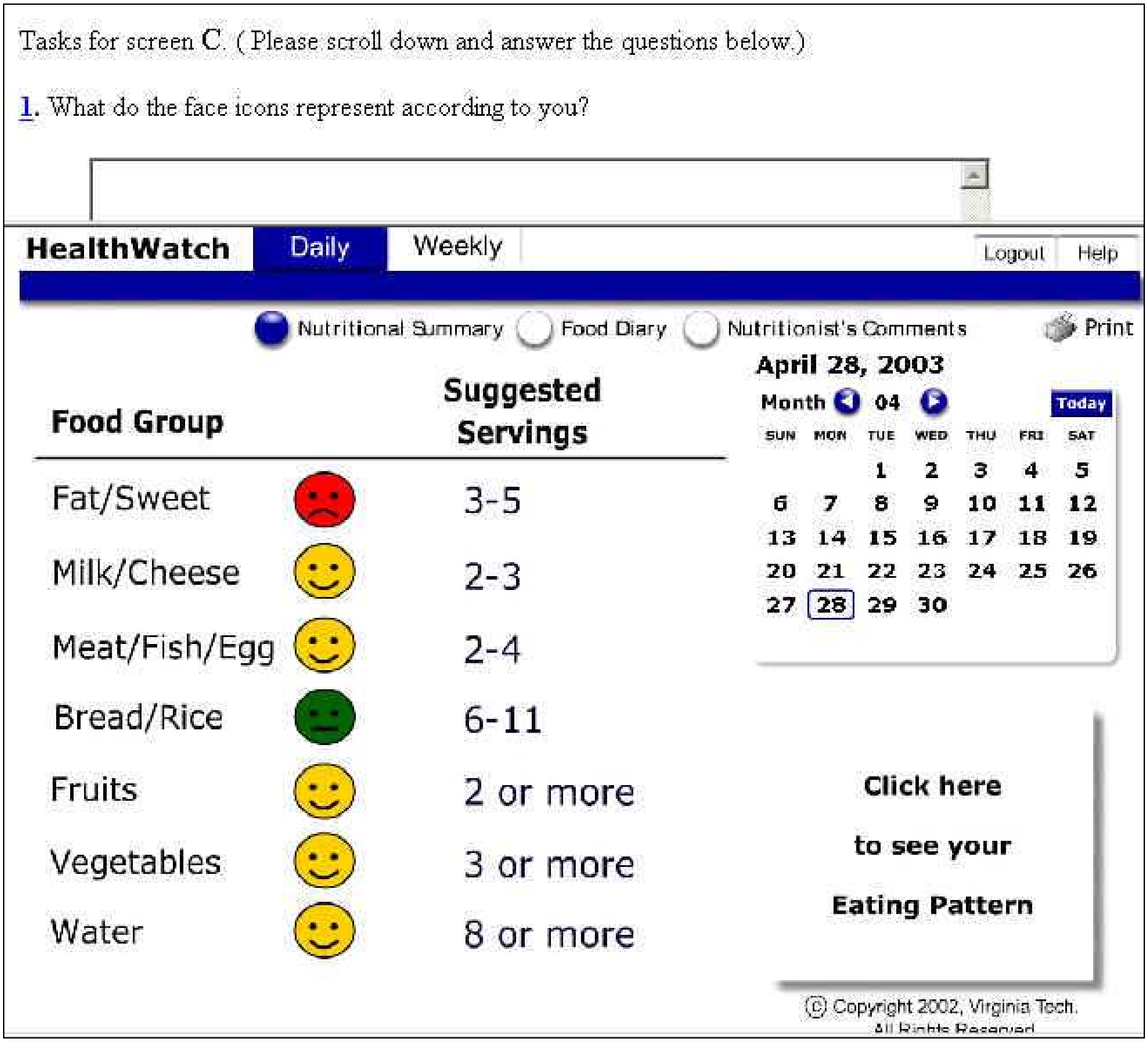


Figure 4: Example of a Concrete representation

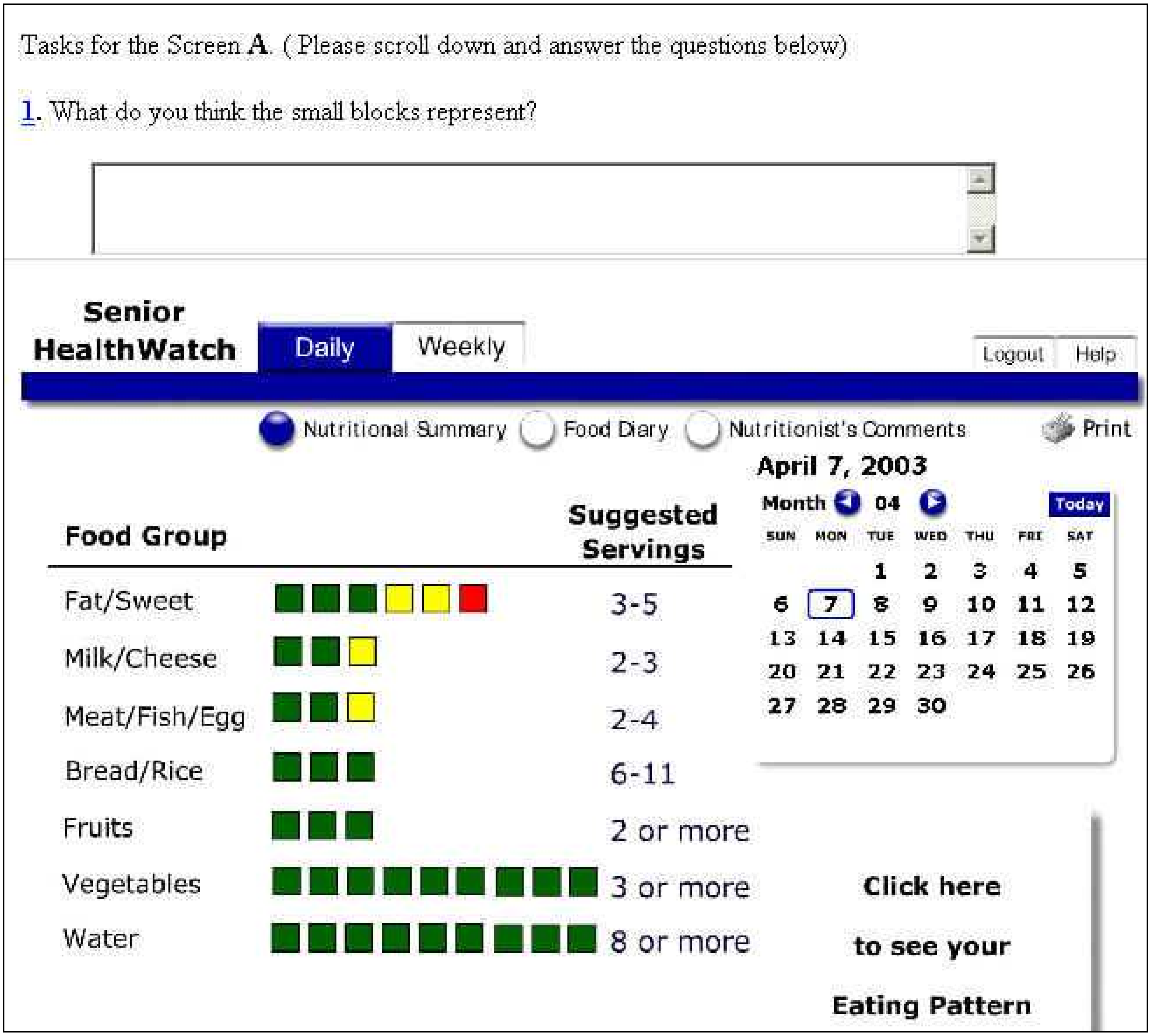


Figure 5: Example of an Abstract representation

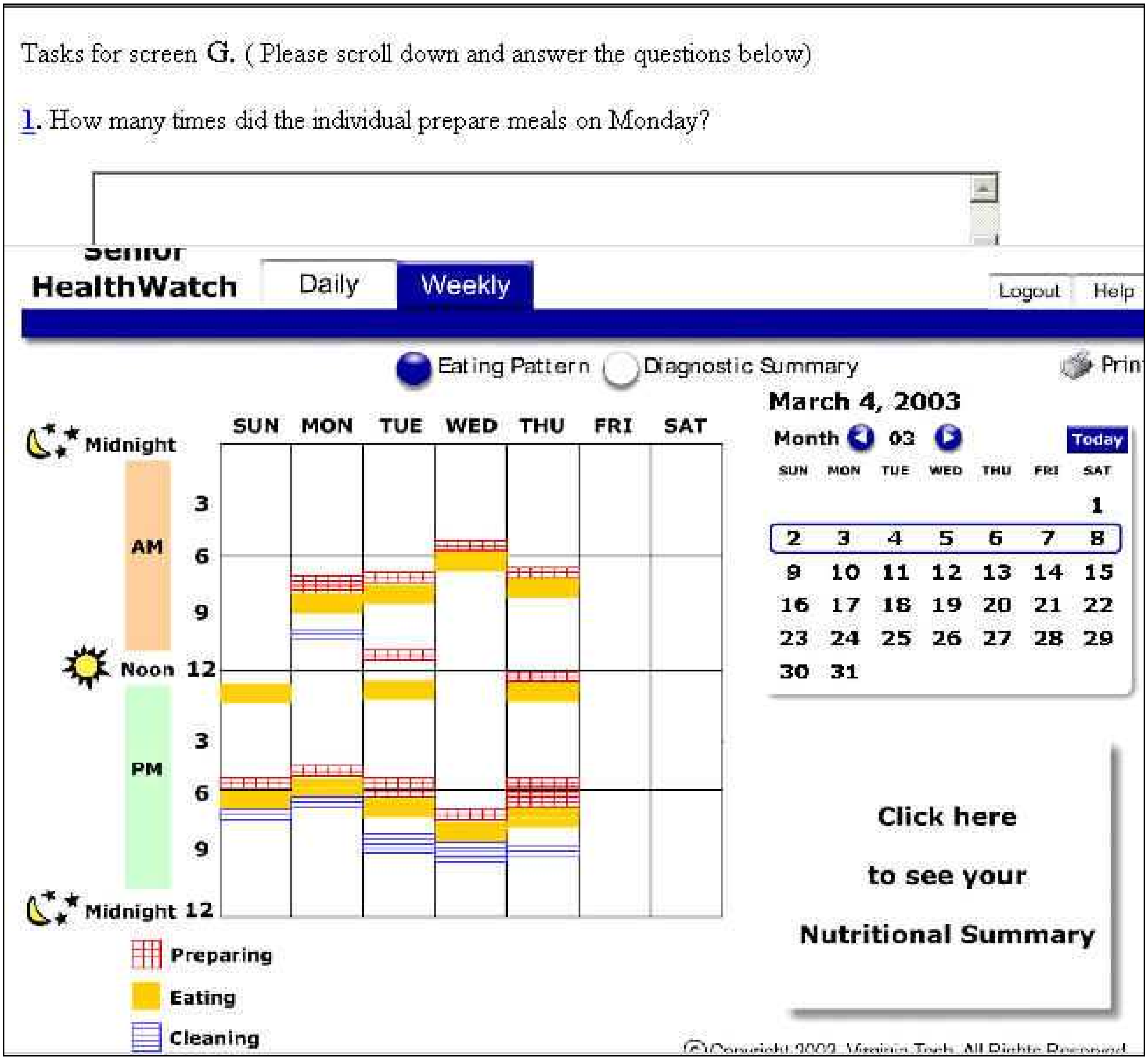


Figure 6: Example of a Graphical representation

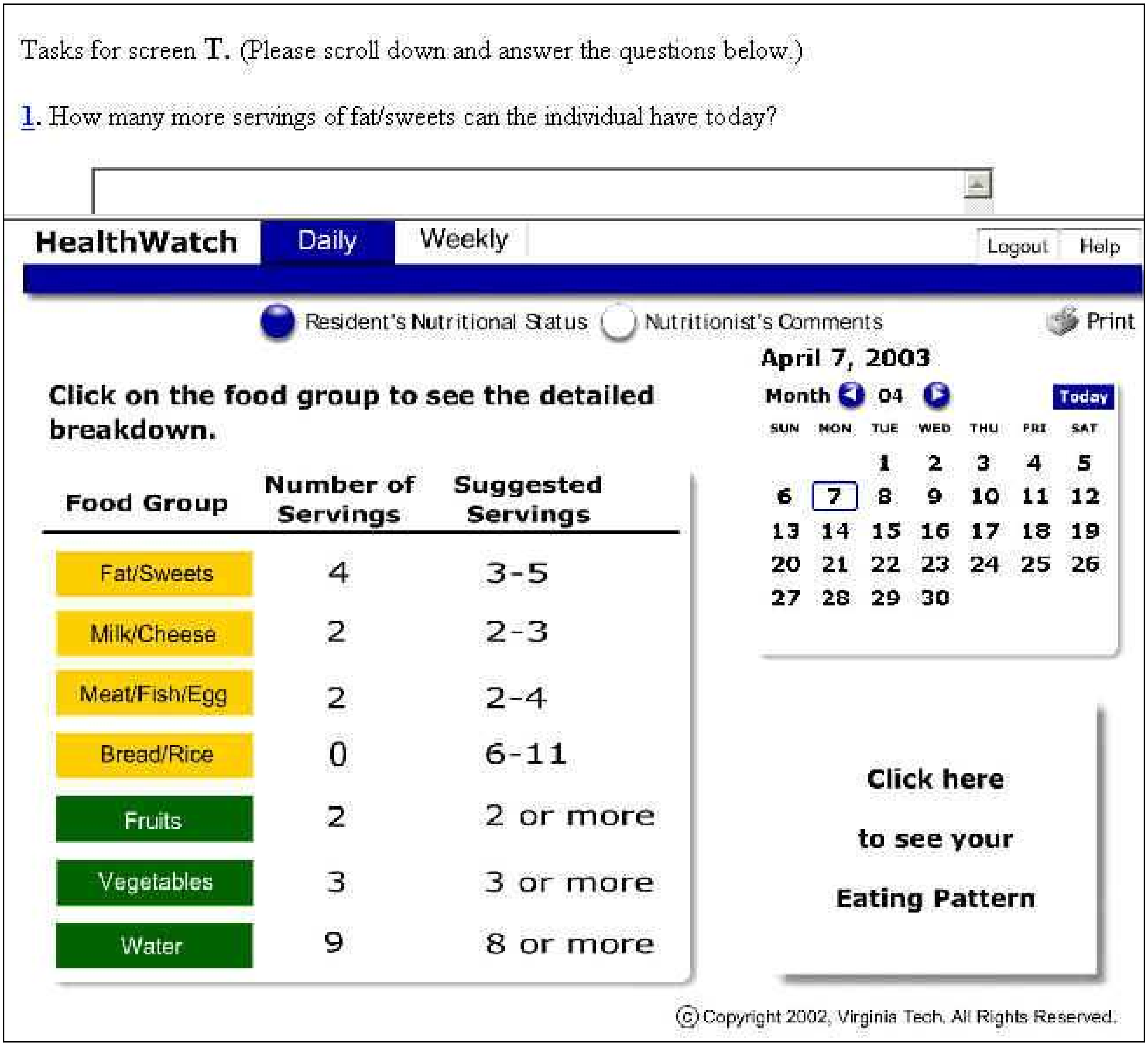


Figure 7: Example of a Tabular representation

4.2.2. Screen Density

Density was defined to be the number of characters displayed, expressed as a percentage of the total spaces available based on the definition by Tullis (1988). The definition used in this study is a different operationalization of Tullis's definition. In this study, perceived density has been adjusted not by changing the number of characters displayed in the space but by changing the total space available to show the information. The two levels of this effect are high density and low density. In this case the number of characters remained the same in both high density and low-density screen, for controlling internal validity, but number of spaces available changed. In high density all the information was cluttered in less space. Where as in low density, the information was arranged over more amount of space on the screen. So, in effect, the whole screen became smaller in size for higher density when compared to lower density. This is a between subject factor. In the high-density screen, the information was cluttered together in small space with a lot of empty space, while in low density the same content was displayed in more organized manner with less empty space. The two levels of screen density were tested with 4 levels of the display representation formats.

4.2.3. Usability Evaluation Method

The two levels of the effect were remote usability testing and inlab usability testing. Usability Testing Method was a between subject factor. For the inlab testing 16 participants were brought to the lab and the usability evaluation of the interface was done in the lab. The participant reported user comments by themselves using the Remote Evaluation Tool. For the remote usability testing, 16 participants evaluated the interface at their respective homes using their home computers and reported user comments on their own using the same tool. The study was conducted using a web interface.

4.3. Dependent Variables

4.3.1. Time

The time taken to answer the questions for all the display representation formats were measured. The time taken to report user comments were subtracted from the time taken to answer questions on each display.

4.3.2. Accuracy

Accuracy was an important dependent measure to determine the effect of the fixed effects. To determine how clear each format of display representation is in high density and low-density screens, the individual had to answer particular questions related to the screen display. The performance was evaluated and the accuracy was measured for each display representation in high and low-density screens.

4.3.3. Subjective Ratings

A rating scale was used to rate the information display formats. The ease of comprehension, ease of use, color usage, and preference were all noted after the participant had finished with each display format. The rating scale, which was used, is shown in Appendix B. A Likert-type scale was used to rate all variables/ constructs. The rating scale was completed after each format of display representation was tested.

4.3.4. Subjective Questionnaire

Beyond accuracy and ratings, it is important to know which display representation format was preferred. The subjective questionnaire is shown in Appendix B along with the rating scale. The subjective questionnaire consisted of open-ended questions regarding each display representation. A questionnaire for eliciting user perception about each method of usability testing was also used. This questionnaire can be found in the Appendix E and F.

4.3.5. Number of User Comments

The number of user comments collected from the remote testing and inlab testing was noted. The concept of critical incidents was proposed by Flanagan (Flanagan, 1954) for steadily collecting observations of critical human behavior in a field environment. A critical incident is defined as an interaction with a system feature leading to extremely good or extremely poor performance (delGaldo, Williges, Williges, and Wixon, 1987). Critical incidents are well suited for remote usability evaluation (Smith-Jackson and Williges, 2001). Therefore it was very important to determine effectiveness of remote usability testing compared to traditional inlab testing. For the research study in consideration critical incidents were reported in the form of user comments from participants and this included comments, which they reported on occasions when the participant finds it difficult to understand an icon, or finds it difficult to answer a question, or likes or dislikes a particular icon etc. User comments were also collected from experts as well. Experts were selected from students who have been trained in usability engineering. The experts each had 2 years experience in usability engineering. The experts were selected on the basis of knowledge of usability and availability to perform the analysis. Both the experts did the analysis of the display separately at their own convenience. The number of user comments collected from participants in inlab testing and remote testing were compared with the number of user comments collected from experts.

4.4. Participants

Thirty-two participants were recruited for the study. The participants were recruited using the age definition of older adults as given by previous researchers (Echt et al., 1998; Morrell et al., 2000), which is at least 60 years of age. The mean age of the participants was 71.22 (*SD* = 6.11). The recruited participants consisted of 16 males and 16 females. Participants were obtained through a newspaper advertisement and through an email to the senior center in Blacksburg. All participants were required to own a computer. Seniors were randomly assigned between the two levels of inlab testing and remote testing such that there were 16 participants in each level. In both inlab testing and remote testing, participants were randomly assigned to low density and high-density screens within each in such a way that 8 individuals were exposed to low-density screens and 8 were exposed to high-density screens. All 8 participants in each high density and low-density condition were exposed to all four levels of display representation format. There was no gender constraint, since the individuals were recruited via randomized selection for the study.

4.5. Equipment and Apparatus

4.5.1. Computer Hardware and Software

Remote data collection occurred in respective homes of the participants on a Home PC. The participants used a desktop computer, keyboard and a mouse. A Remote Evaluation tool was used to collect user comments. This tool is described in the next section. An example of a user comment was given to participants as an occasion when the participant finds it difficult to understand an icon, or finds it difficult to answer a question, or likes or dislikes a particular icon etc. For the inlab testing, a lab PC was used. A desktop computer, keyboard and mouse were also used. A remote evaluation tool, which is discussed later in the paper, was also used to collect user comments.

4.5.2. Observation and Data Recording

For the remote testing, the information display formats were available online as web pages for the participants. The questionnaire and the rating scales were also available online. The individuals reported user comments online by using the Remote Evaluation Tool developed by Dr.Hartson and Narayanan Kodiyalam of the Computer Science Department’s Center for Human-Computer Interaction (HCI) at Virginia Tech. Figure 8 shows the Login screen of the remote evaluation tool .

For the inlab testing, the information display formats were available online as well. The questionnaire and rating scales were also available online. The Remote Evaluation Tool was provided to participants and they were asked to report their own user comments.

The Remote Evaluation Tool describes a user comment and how to report it. The participants had to click on the button “Begin Session” to open the required webpage. The button “Report Comment” appeared on top of the computer screen, which they had to click when they wanted to report a user comment. The window with the button could be minimized while the user is working on the interfaces. When the participant wanted to report user comment they maximized the “Report Comment” button and clicked on the button. The action brought up a pop-up, which asked the participant to give comments about the user comment. The pop-up had three questions which are as follows: “Please give a name to the usability problem you are reporting, a name that you can use to remember or identify this problem later, if necessary?”, “Please enter the location of the screen where you found the problem.”, “Please give a brief description of the comment you have .” The tool was used to record user comments and the comment file was saved on the participant’s computer in a folder called Remote Evaluation in the c directory.

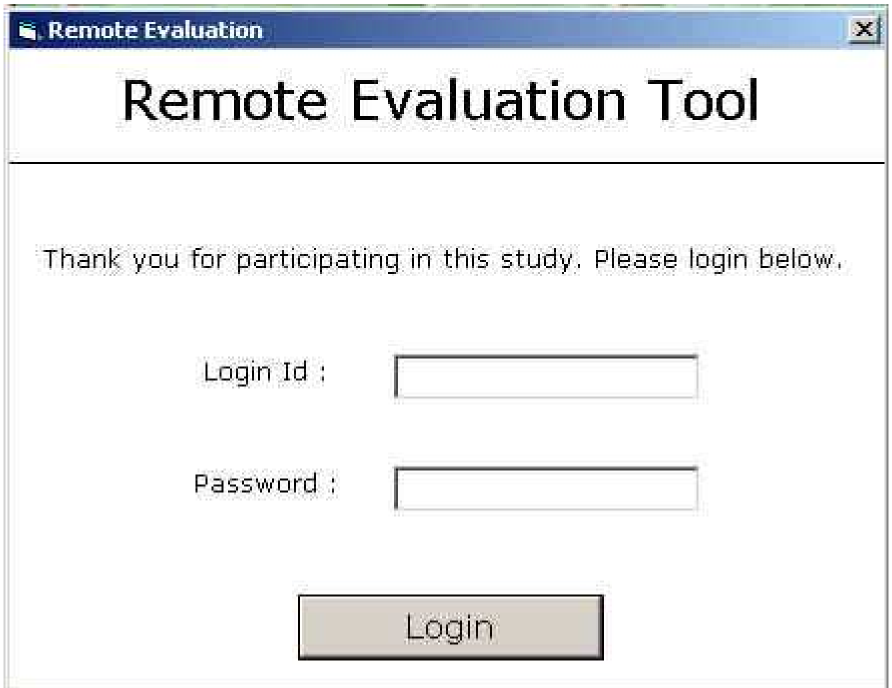


Figure 8: Login screen for the Remote Evaluation tool

4.6. Procedure

For the remote testing, the participants were sent an email about the study. The participants were given instructions to review. The participants were asked to complete a demographic questionnaire given in Appendix D. The questionnaire included computer experience-based questions as well as information about each participant's age, gender and occupation.

The instructions introduced the participants to the study and also gave directions on how to perform the test. The participants were also given a step-by-step instruction sheet on how to perform the test (Appendix A). Once the participants had completed reading the instructions, they had to open the Remote Evaluation Tool and login. Then they had to visit the first interface provided to them and answer a set of online task questions related to the information on the display (Appendix H). While they were viewing each screen they had the opportunity to report a user comment. If the participants opted to report a user comment, they maximized the Remote Evaluation Tool window and reported it as per the procedure described above.

Once participants reported all user comments related to the interface, they were asked to complete an online questionnaire. The questionnaire consisted of rating scales and open-ended questions. Participants had to rate the ease of use, ease of comprehension and color usage of the interfaces. The open-ended questions were regarding preferences for the interface and reasons for the preferences. Participants were then provided with the next interface and had to complete the questionnaire for the same. After the participants were tested with the four display representation formats one by one, they were presented with the same four display representation formats in the end to decide which they preferred the best. For noting their preference they had to complete two more questionnaires. The first questionnaire consisted of open-ended questions that helped determine which display representation format was most preferred by most of the individuals and the second questionnaire was about the remote testing method and how they preferred it. These questionnaires were also available online. This way the test was conducted remotely and also user comments and the test data was collected remotely. Figure 14 shows a flowchart describing the procedure for the testing procedure.

For the inlab testing, the data collection occurred in the HCI/ACE lab at Whittemore Hall in Virginia Tech. The participants were asked to read instructions. The participants completed a demographic questionnaire same as in the remote testing. Once filling the demographic questionnaire was completed, the participants had to open the Remote Evaluation Tool and login. The participants were presented with the first display representation format. After viewing the display, they had to complete a task questionnaire with a few questions regarding the interface. While they were viewing each screen they had the opportunity to report user comments using the

Remote Evaluation Tool. When the participants opted to report a user comment they maximized the Remote Evaluation Tool window and reported it as per the procedure described above.

Once participants reported all the user comments related to the interface, they were asked to complete an online questionnaire. The questionnaire consisted of rating scales and open-ended questions. Participants had to rate the ease of use, ease of comprehension and color usage of the interfaces. The open-ended questions were regarding preferences for the interface and reasons for the preferences. Then the individual was presented with the next display. The same procedure followed for all four-display representation formats. After the participants were tested with the four display representation formats one by one, they had to complete two more questionnaires. The first questionnaire consisted of open-ended questions that helped determine which display representation format was most preferred by the participants. In order to answer the questionnaire, the participants were presented with all four displays once again. Once the questionnaire was completed, they were presented with a questionnaire with questions about the inlab testing method and their preference. A debriefing of the study followed. Figure 9 shows a flowchart describing the procedure for the testing.

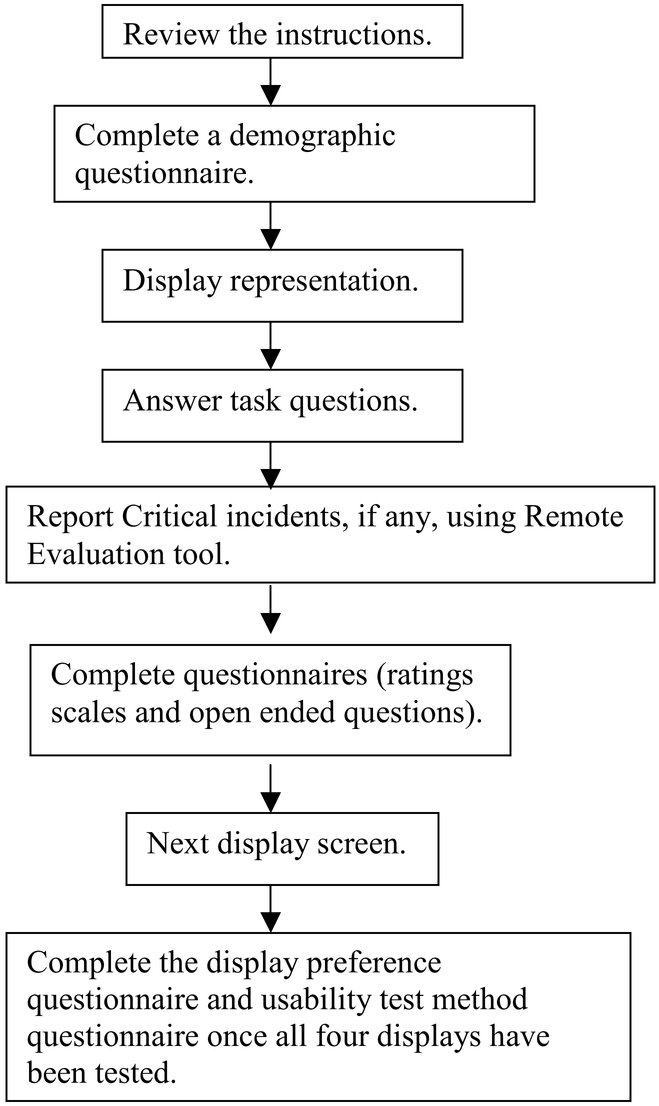


Figure 9: Flowchart representing inlab/remote testing method.

5. RESULTS

5.1. Data Analysis

Statistical Analysis System (SAS) was used to conduct all analyses. All Analyses of Variance (ANOVAs) used the General Linear Model (GLM) procedure. All post hoc analyses used Tukey HSD Analysis. Alpha was set at a level of .05. The dependent measures tested for significance are given in Table 3.

Table 3: Dependent Variable and Description

Dependent Variable	Description
Time	The length of time taken on each display screen measured from the time the display screen appears until the time the screen was exited excluding the time taken to report user comments.
Accuracy	Determined as the ratio of number of tasks answered correctly out of total number of tasks as defined by the answer key by four tasks per screen.
Subjective ratings	Mean of the Likert scale responses to the 11 questions on the subjective questionnaire referring to the ease of comprehension, ease of use, color usage, preference.
Subjective open ended responses	Open ended responses regarding what the user preferences about the display.
# of User Comments	Total # of user comments per display.

5.2 Analysis of Dependent Variable Time

A 3 way ANOVA was used to examine the time dependant variable as a function of display, density and mode. Time taken to complete tasks for abstract (TA), concrete (TC), tabular (TT), and graphical (TG) displays were analyzed. Figure 10 illustrates the plot of the mean time taken for all displays.

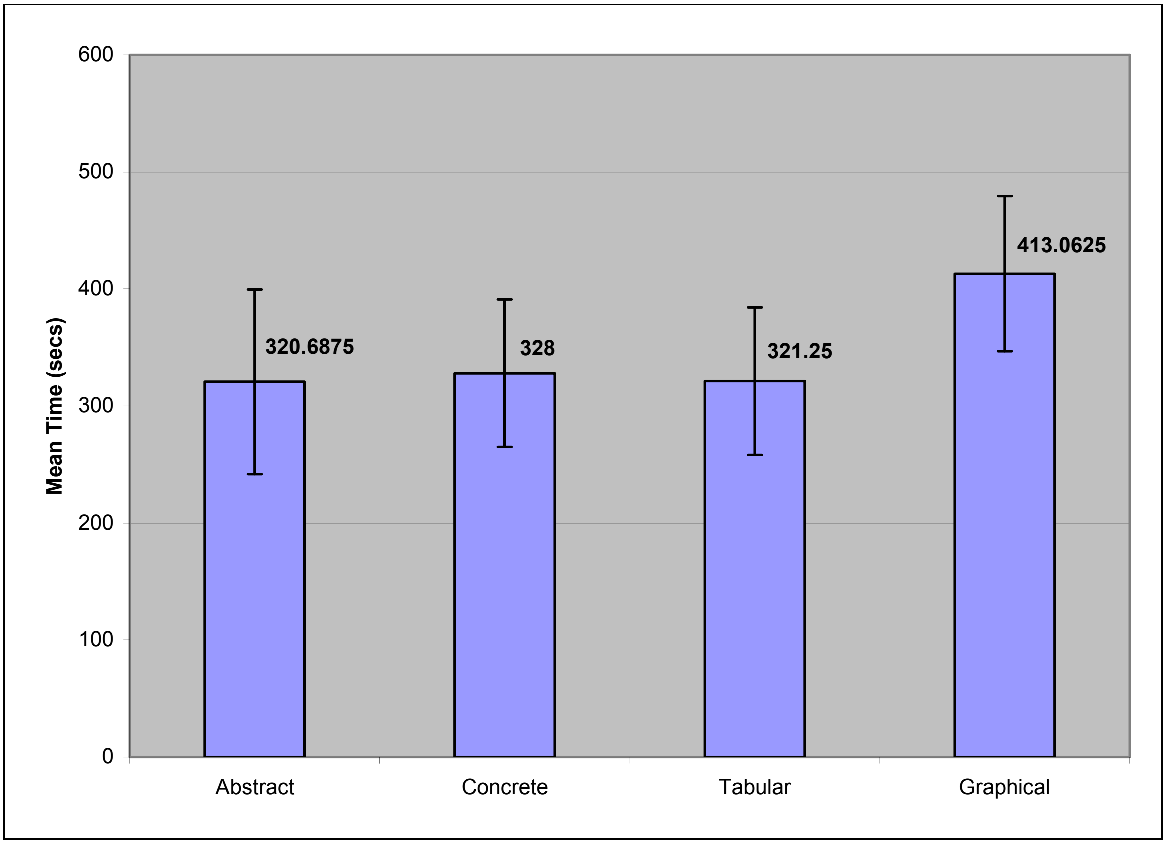


Figure 10: Task time for different display representations

Though there is an increase in task time for graphical displays no significant difference was noticed when ANOVA proc glm analysis was conducted, $F\ (3,112) = 1.460$, $p > .05$.

Though there was no significance in displays effect, mode effect was significant, $F\ (1,112) = 5.031$, $p < .05$. Density effect also showed statistical significance $F\ (1,112) = 8.934$, $p < .05$. From further post-hoc analysis it was determined that participants using higher density screens ($\mu = 401.47$, $SD = 259.84$) took significantly higher time to perform tasks than participants in lower density screens ($\mu = 290.09$, $SD = 139.78$). Participants in remote lab testing ($\mu = 387.52$, $SD = 232.69$) took a significantly longer time to perform tasks as compared to participants in inlab testing ($\mu = 303.99$, $SD = 188.73$). Figures 11 and 12 illustrate the difference in time taken to perform tasks by participants in high density and low-density screens and in remote testing and inlab testing respectively.

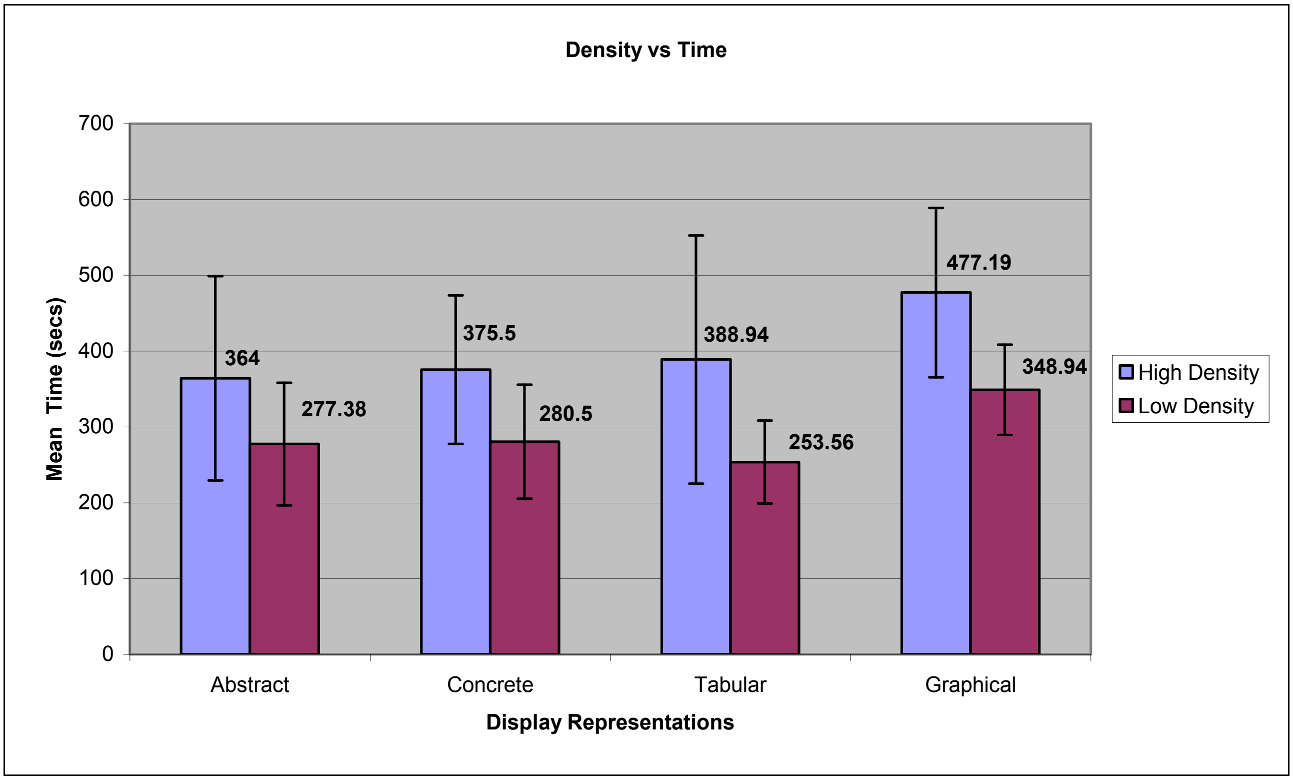


Figure 11: Density vs. time for different display representations.

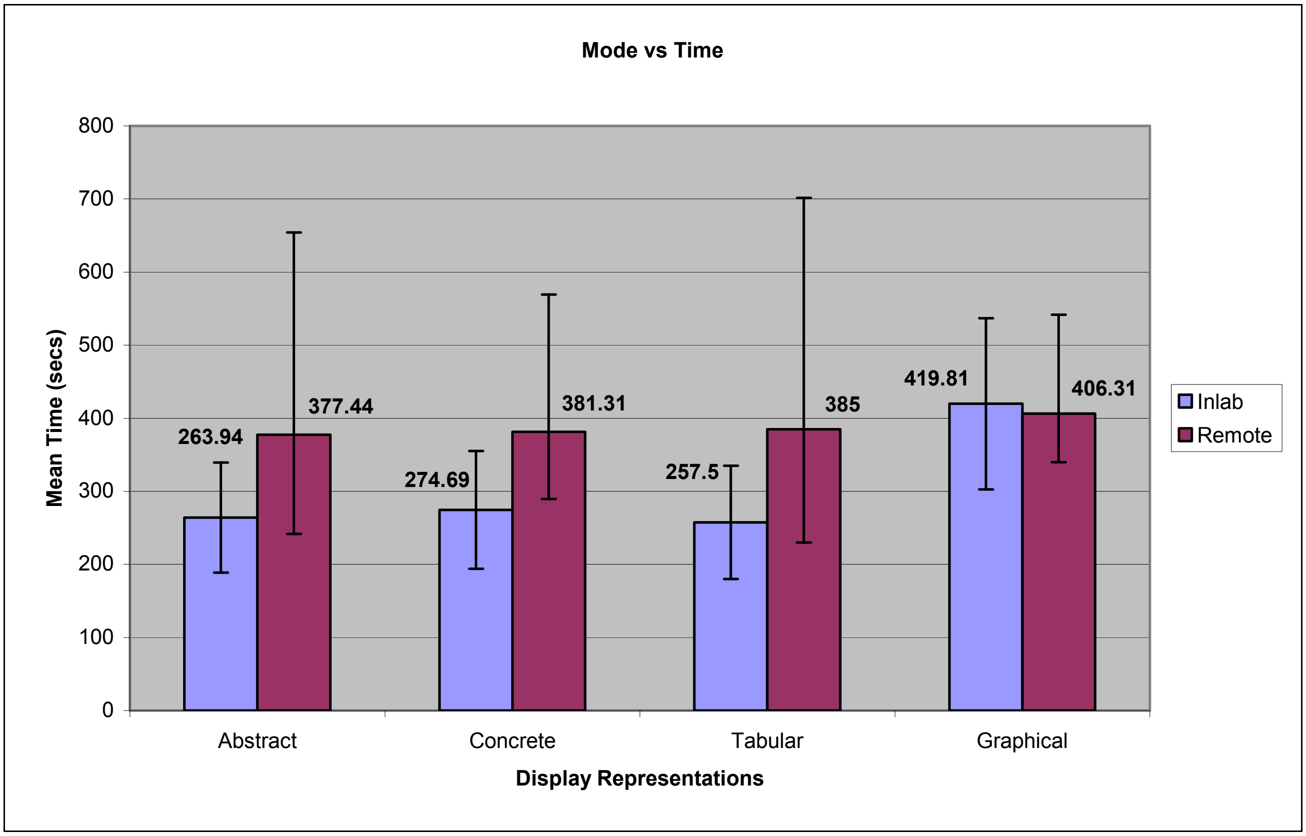


Figure 12: Mode vs time for different display representations

Though there are no significant interactions, from the Figure 11 we can see that participants take more time to perform tasks in high-density screens as compared to low-density screens. It is also noted that participants take more time to complete tasks when they are using graphical representation as compared to the other display representation formats in both high-density and low-density screens. From Figure 12 it can be noted that both in inlab testing and remote testing participants performed slower in graphical displays. Order effect did not show any significance.

5.3. Analysis of Dependent Variable Accuracy

A three way ANOVA was also conducted to analyze the dependent variable accuracy as a function of display, density and mode. Accuracy for abstract tasks (AA), accuracy for concrete tasks (CC), accuracy for tabular tasks (AT), and accuracy for graphical tasks (AG) were analyzed. Figure 13 illustrates the accuracies for different displays.

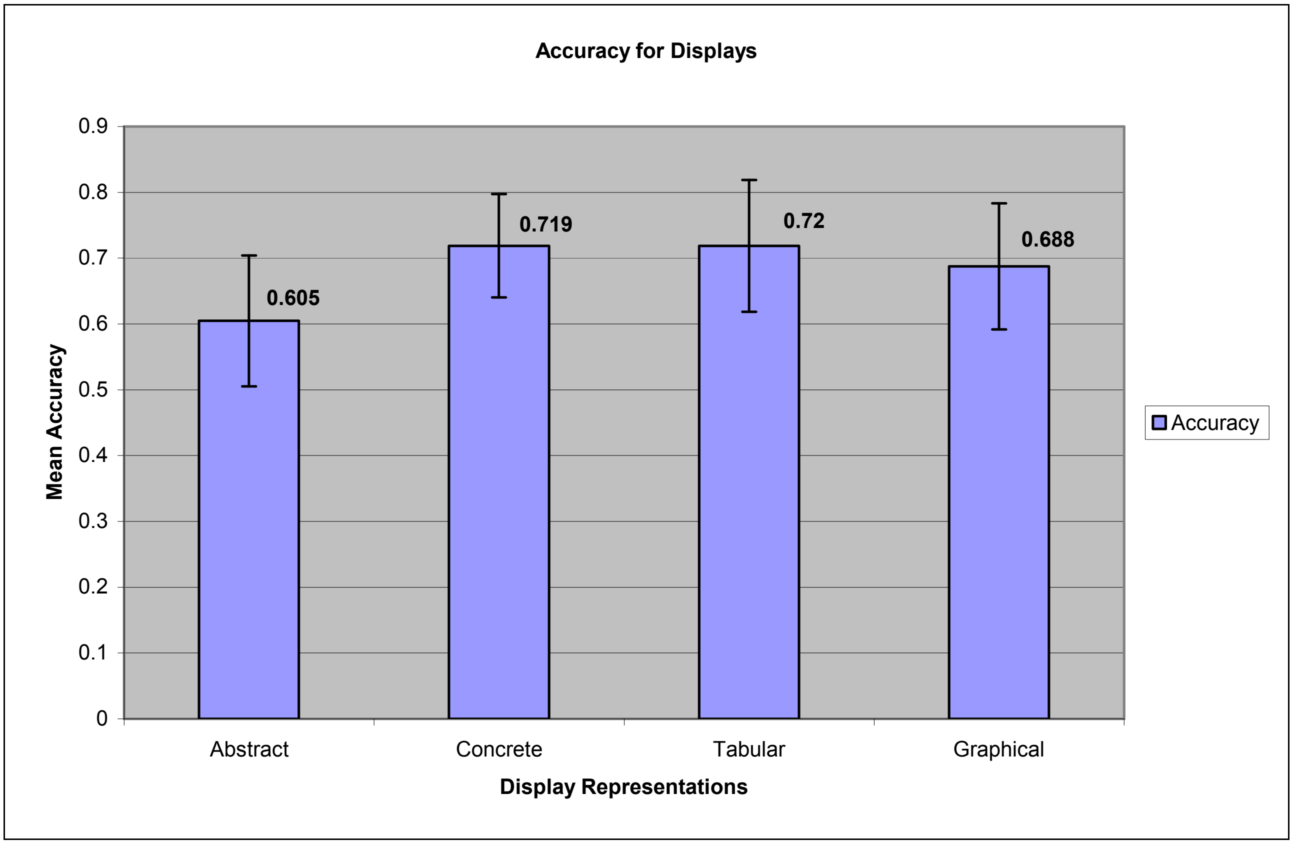


Figure 13: Accuracy for different display representations.

Though there is an increase in accuracy for concrete and tabular displays no significant difference (p>.05) was noticed when ANOVA proc glm analysis was conducted. There was no significant difference in density effect nor in mode effect. Figure 14 illustrates density vs accuracy for different displays.

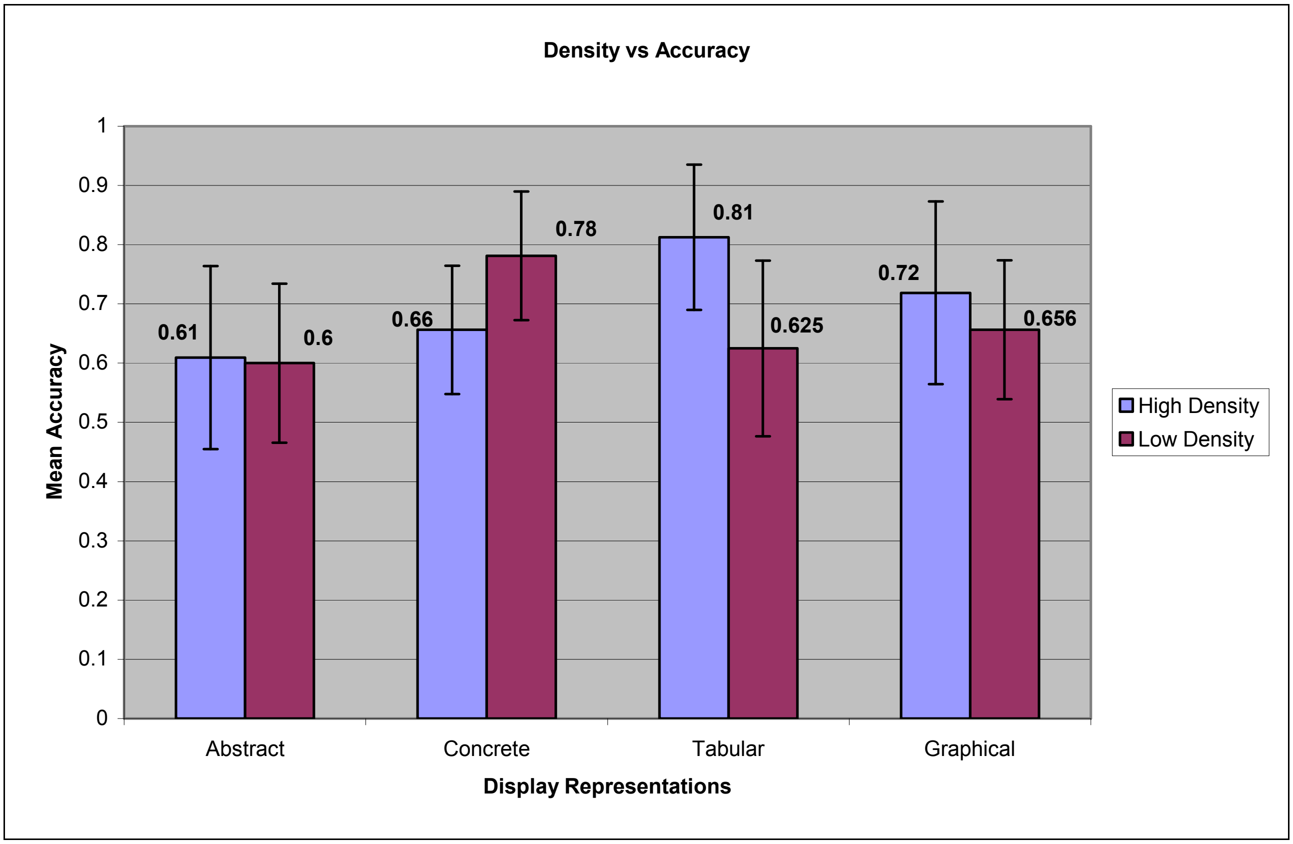


Figure 14: Density vs accuracy for different display representations

Accuracy for tabular screens was higher in high-density screens. From Density vs Accuracy plots, participants showed greater accuracy for tabular and graphical displays in high density while for concrete display participants showed better accuracy in a low-density screen. Participants showed almost similar amount of accuracy in high density and low-density screens for abstract displays.

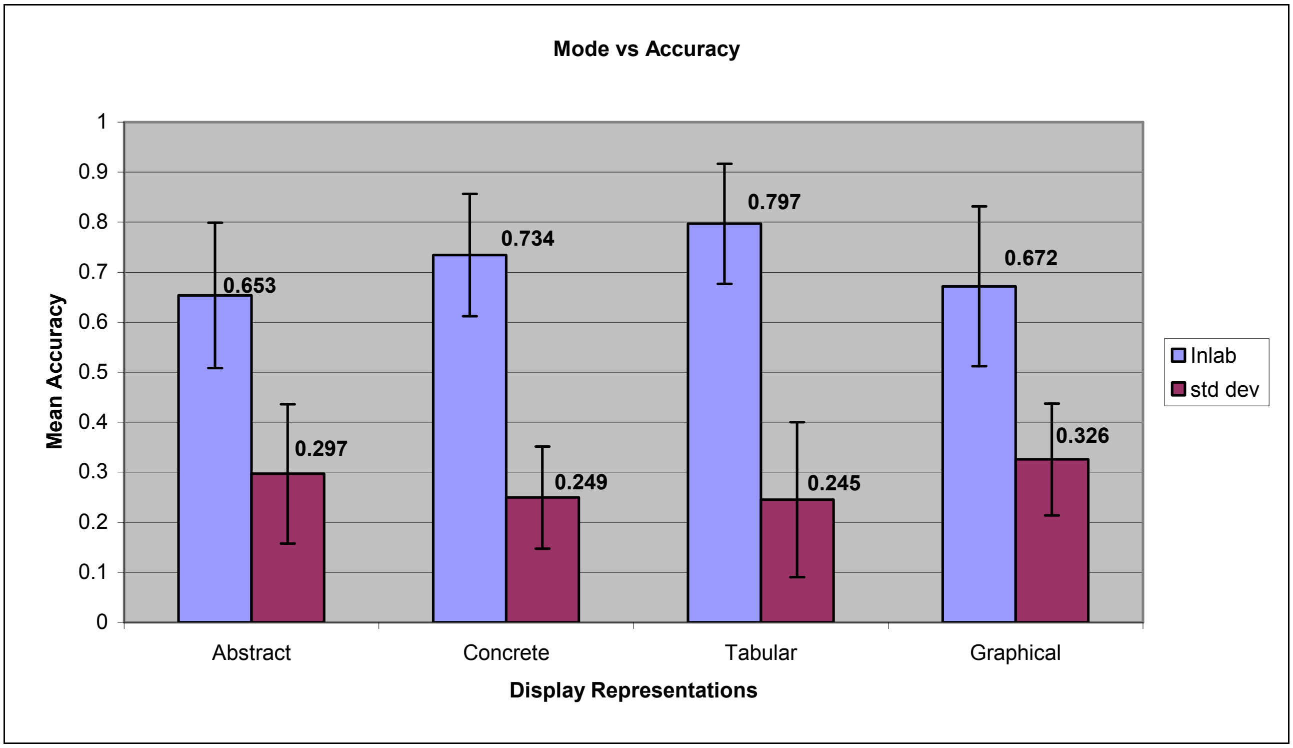


Figure 15: Mode vs accuracy for different display representations

In Figure 15 we observe that remote participants have less accuracy compared to inlab except for in graphical representation though there is no statistical significance shown. Order effect did not show any significance.

5.4 Analysis of Subjective Preference of Displays

The preferences of the participants were analyzed using Chi Square ($\chi^2 = 40.75$, df=4, $\alpha = 0.05$) and significance was noted in preferences. Results showed that 37.5% preferred abstract display, 25 % preferred concrete display, 18.75 % preferred tabular and 15.625 % preferred graphical and 3.125 % preferred all the displays. Figure 16 shows the preferences of the different display formats.

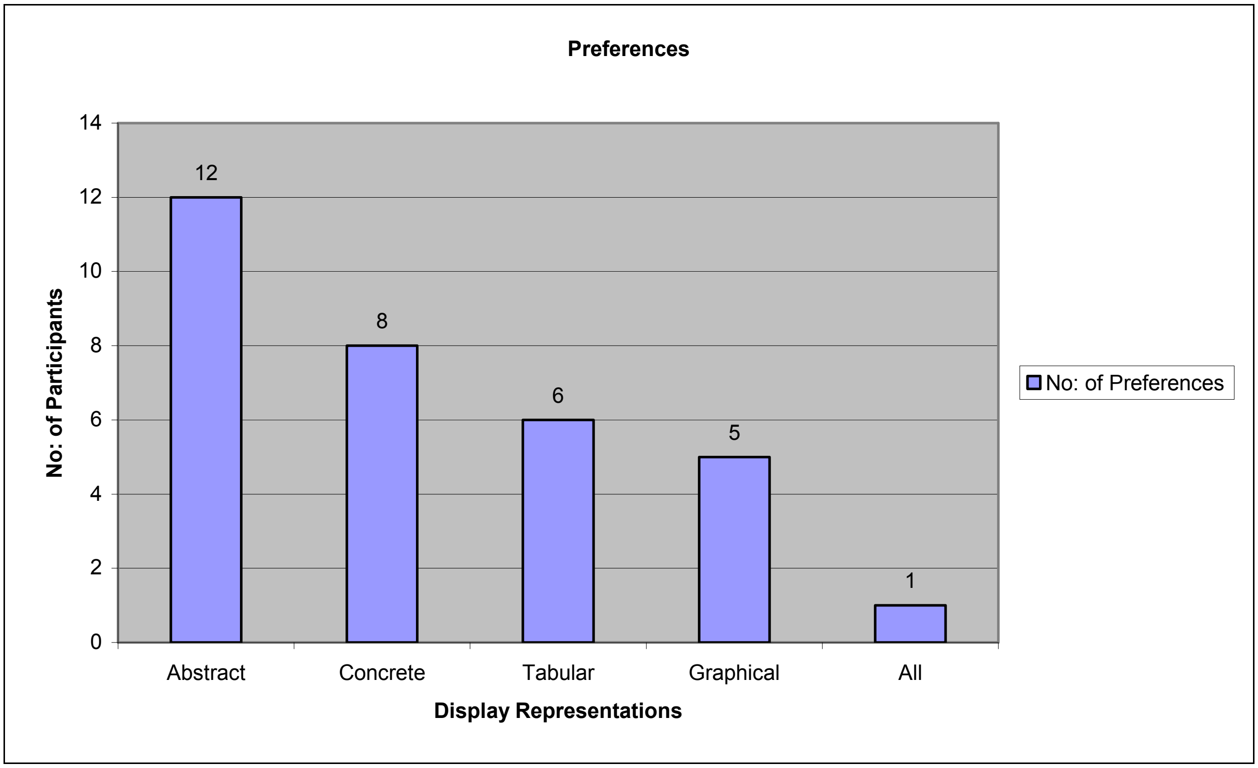


Figure 16: Preference data for different display representations

5.5. Subjective Questionnaire Data Analysis

Two questionnaires were administered to participants using an on-line form-- a questionnaire for every display representation and a post-test questionnaire related to testing preferences (inlab and remote). The purpose of these questionnaires was to solicit subjective ratings regarding the usability of a particular interface and how the participants liked the testing procedure. An analysis and discussion of ratings gathered from each questionnaire are presented below.

5.5.1 Display Representation Questionnaire Results

The display representation questionnaire was administered to all participants that went through the test. It was comprised of a series of questions that required the participant to rate some aspect of the interfaces. Ratings allocated by each participant is summarized in Table 5. Though 16 of the participants were exposed to low-density screens and 16 of the participants were exposed to high-density screens, the questionnaire data is analyzed with respect to density and display representations since all participants received the all four displays. A first iteration for evaluation and redesign of the questionnaire was conducted based on results of the pilot study.

Table 4: Questionnaire results

# of Question	Question	Abstract	Concrete	Tabular	Graphical
1	Over all I found the interface appealing.	2.38 (.83)	2.28 (.89)	2.28 (.81)	2.81 (.78)
2	I liked the colors used in the screen	1.93 (.56)	2 (.72)	2.06 (.56)	2.55 (.75)
3	I was able to answer the questions easily.	2.97 (1.23)	2.34 (1.12)	2.40 (1.10)	3.09 (.88)
4	I had no difficulty in answering the questions.	2.97 (1.28)	2.40 (1.12)	2.34 (0.9)	2.90 (1.05)
5	The fonts were clear to read	2.21 (.79)	1.81 (.64)	2.15 (.72)	2.27 (.56)
6	I liked the interface	2.38 (1.03)	2.09 (.89)	2.25 (.98)	2.90 (1.14)
7	The interface was clear	2.53 (1.07)	2.34 (1.03)	2.31 (1.12)	2.72 (1.14)
8	I am confident I can use this interface easily	2.62 (1.16)	2.22 (.79)	2.19 (.97)	2.64 (.95)
9	I will feel comfortable trying to use this interface again	2.59 (1.07)	2.16 (0.81)	2.34 (0.97)	2.55 (.88)
10	The interface meets my expectations	2.84 (0.81)	2.5 (0.88)	2.47 (0.88)	2.73 (0.90)
11	The icons were easy to learn	2.88 (1.00)	2.31 (0.86)	2.22 (0.87)	2.64 (0.84)

A 2way ANOVA (Factor A= Display Representations and Factor B=Density) was carried out for each question. A parametric analysis was considered a valid approach since 5-point Likert scales were used, which approximate interval scale data. Significant effects were found in 5 out of 11 questions. Table 5 summarizes the constructs and the significant results. Figures 17-22 illustrate the questionnaire ratings, which had significance.

Table 5: Subjective questionnaire significant results

Construct	Abstract	Concrete	Tabular	Graphical	Low density	High density
No Difficulty			T-G	T-G		
Liking		C-G		C-G	L-H	L-H
Confidence					L-H	L-H
Comfort Level					L-H	L-H
Easy to Learn	C-A, T-A	C-A	T-A			

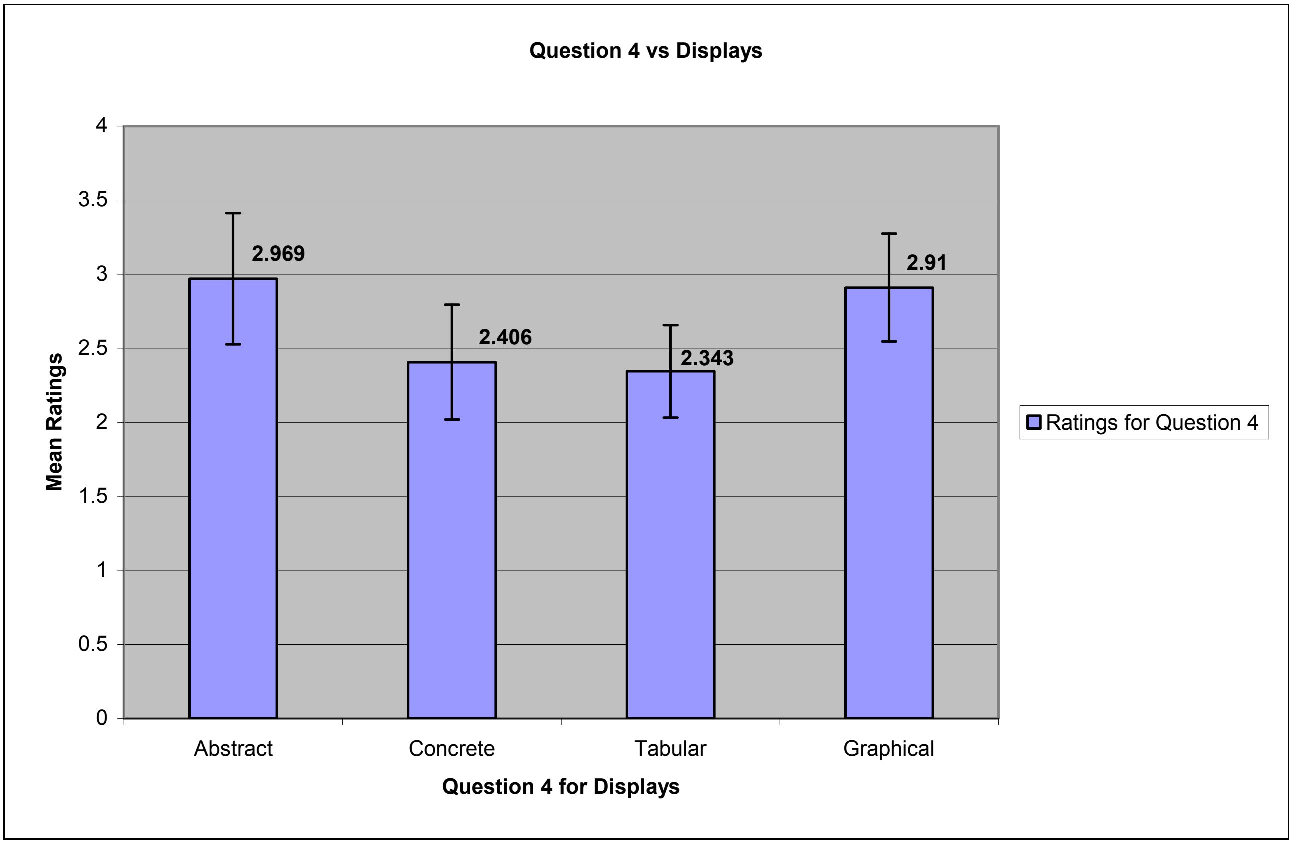


Figure 17: Question4. I had no difficulty in answering the questions.

Figure 17 shows that question 4 had a significant display effect, F (3,112) = 3.681, p< .05. From post hoc comparisons it was clear that there is a significant difference between graphical and tabular ratings.

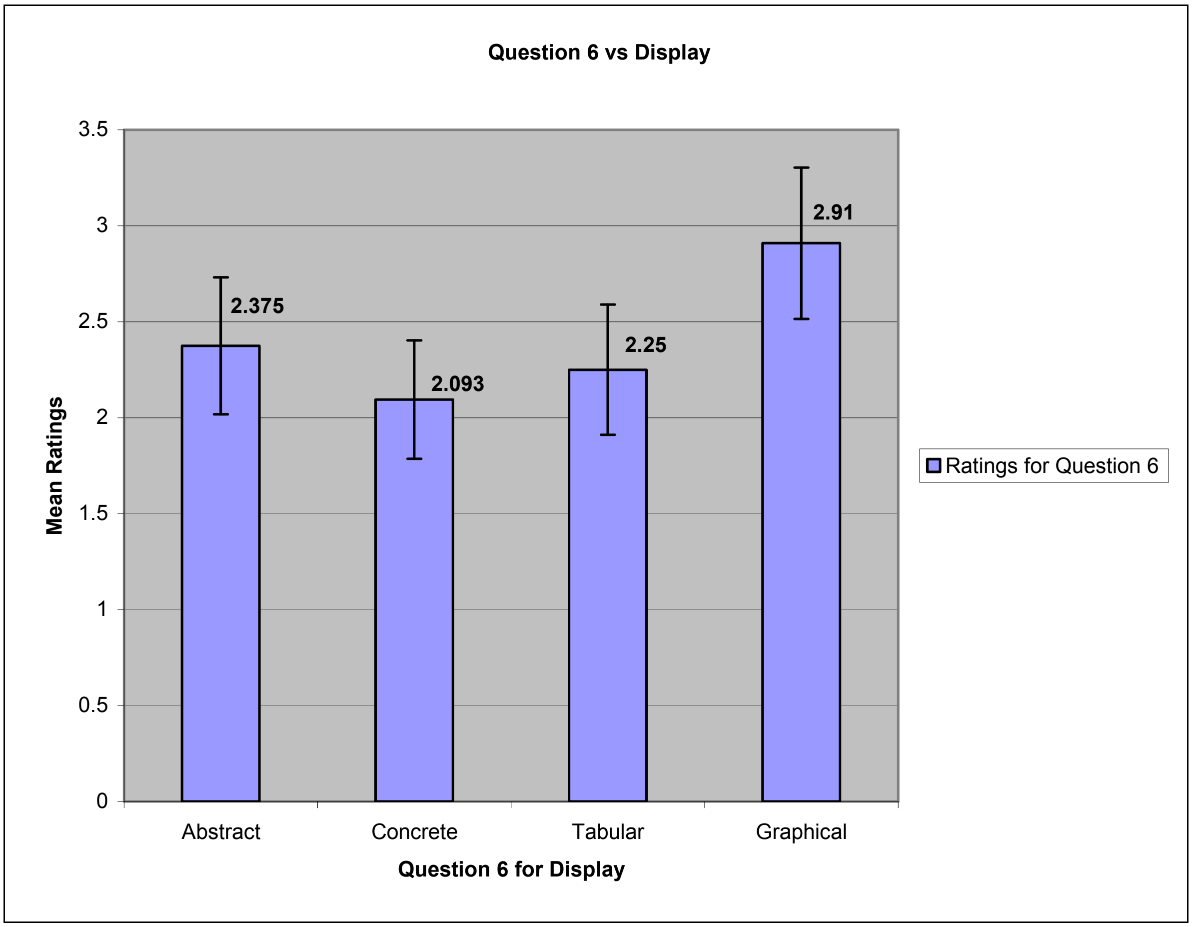


Figure 18: Question 6. I liked the interface

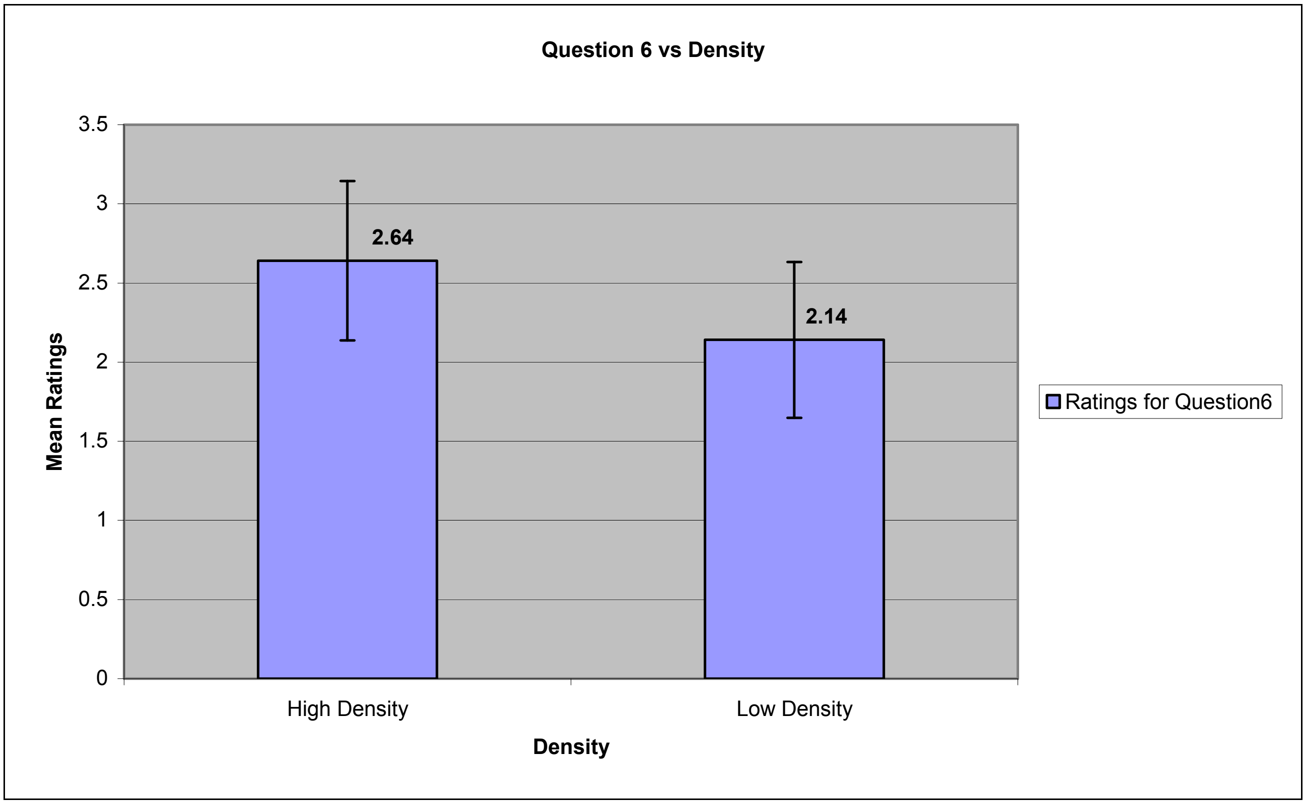


Figure 19: Question 6. I liked the interface

Figure 18 and Figure 19 show that question 6 had a significant display effect F (3,112) = 3.152, $p < .05$, and significant density effect, F (1,112) = 8.493, $p < .05$. From post hoc comparisons it was clear that there is a significant difference between graphical and concrete ratings. Also there was a significant difference between low density and high density ratings.

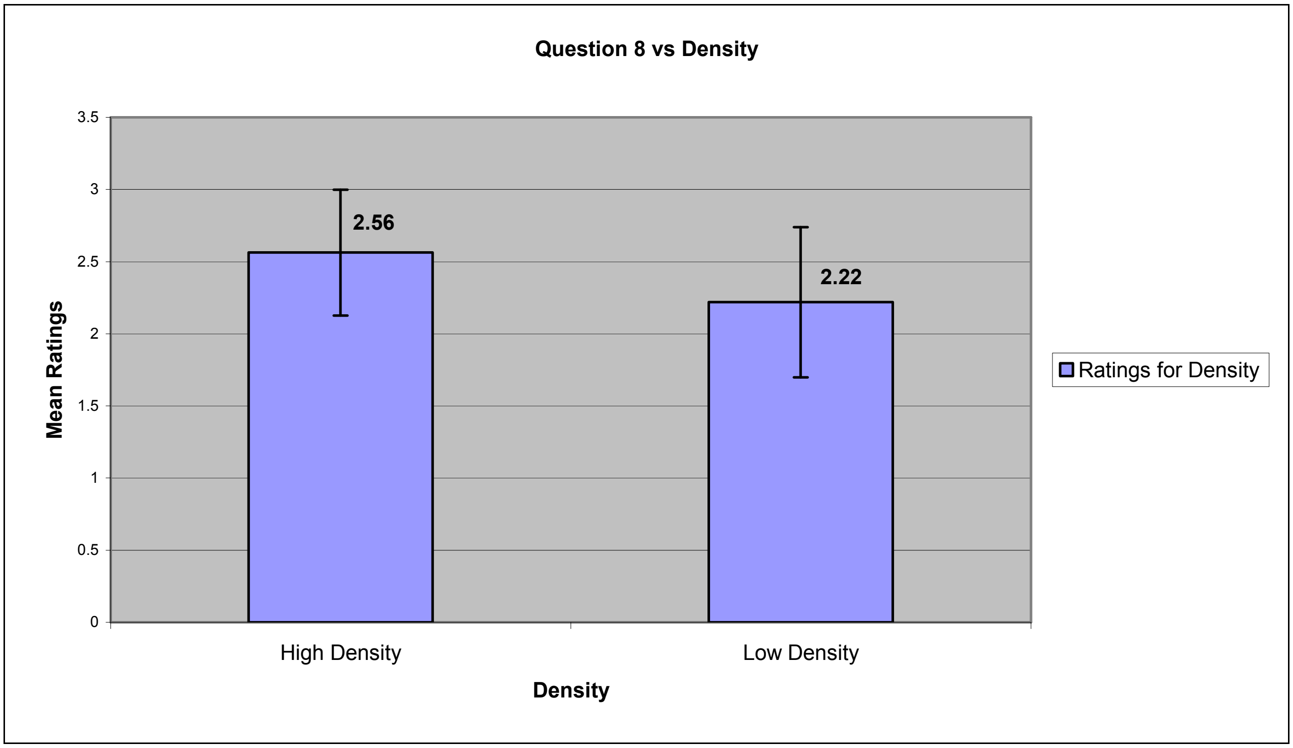


Figure 20: Question 8. I am confident I can use this interface easily

Figure 20 show that question 8 had a significant density effect F (1,112) =4.053, p< .05. So there is a significant difference between low density and high density ratings.

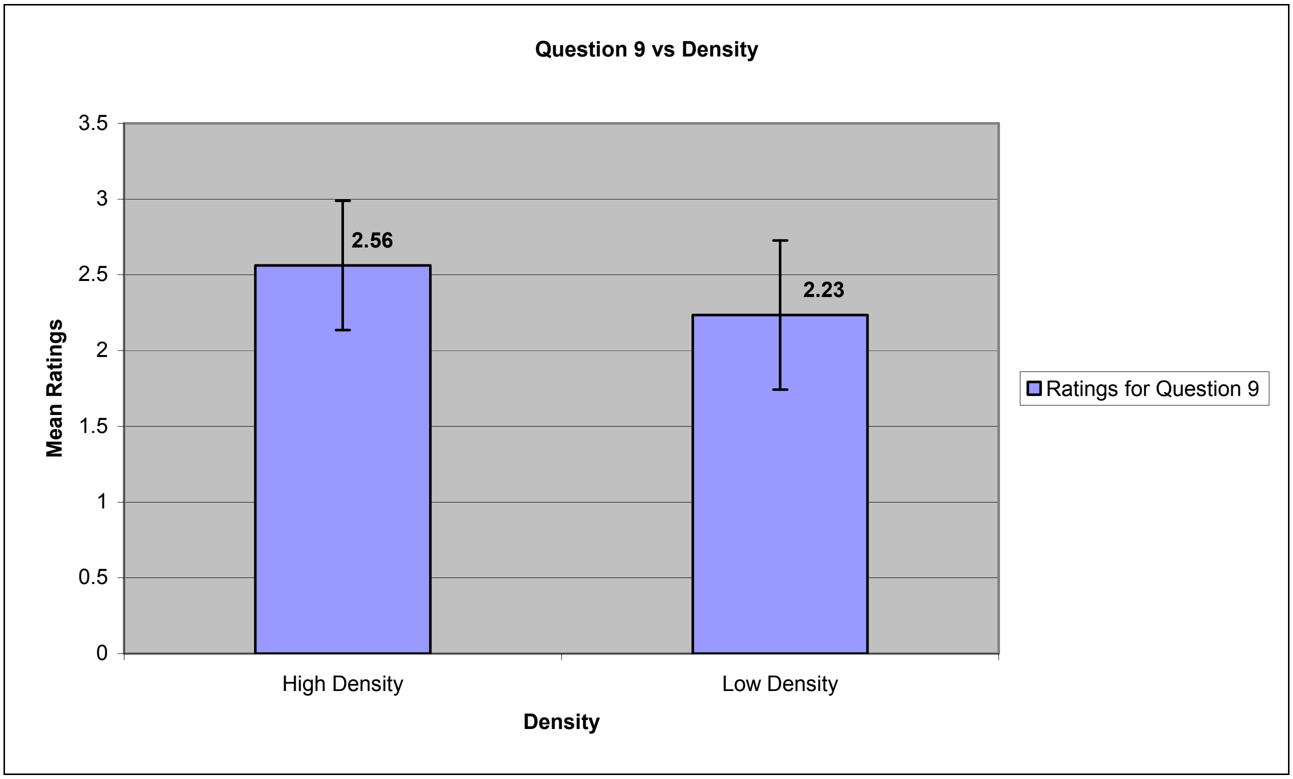


Figure 21: Question 9. I will feel comfortable trying to use this interface again

Figure 21 show that question 9 had a significant density effect F (1,112) = 4.025, p< .05. So there is a significant difference between low density and high-density ratings.

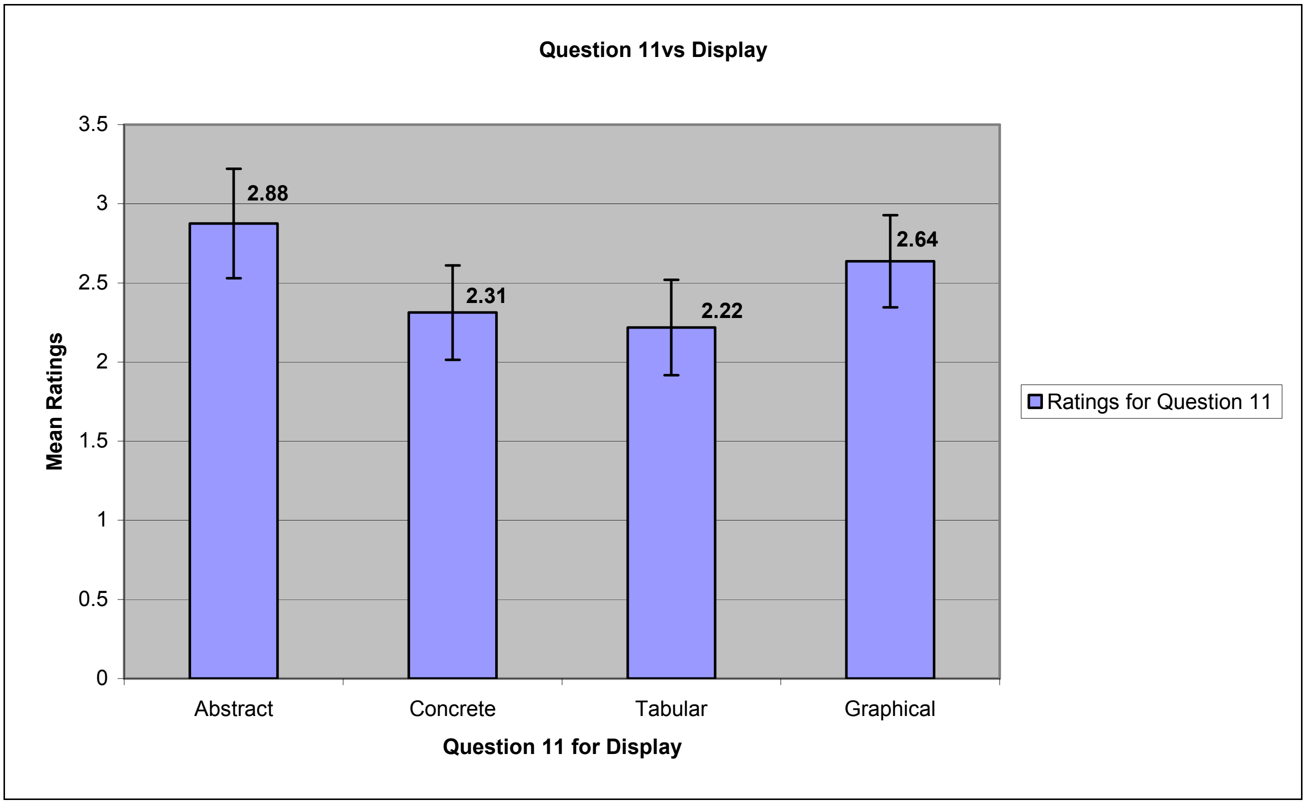


Figure 22: Question 11. The icons were easy to learn

Figure 22 shows that question 11 had a significant display effect $F(3,112) = 3.826$, $p < .05$. From post hoc comparisons it was clear that there is a significant difference between abstract and concrete ratings, and abstract and tabular ratings.

5.5.2 Inlab /Remote Testing Questionnaire Results

Comparing the subjective ratings of remote vs inlab testing, there was no significant difference in ratings between modes. Remote testing was liked as much by the participants as inlab testing. Figure 23 illustrates the questionnaire ratings for the inlab testing vs remote testing.

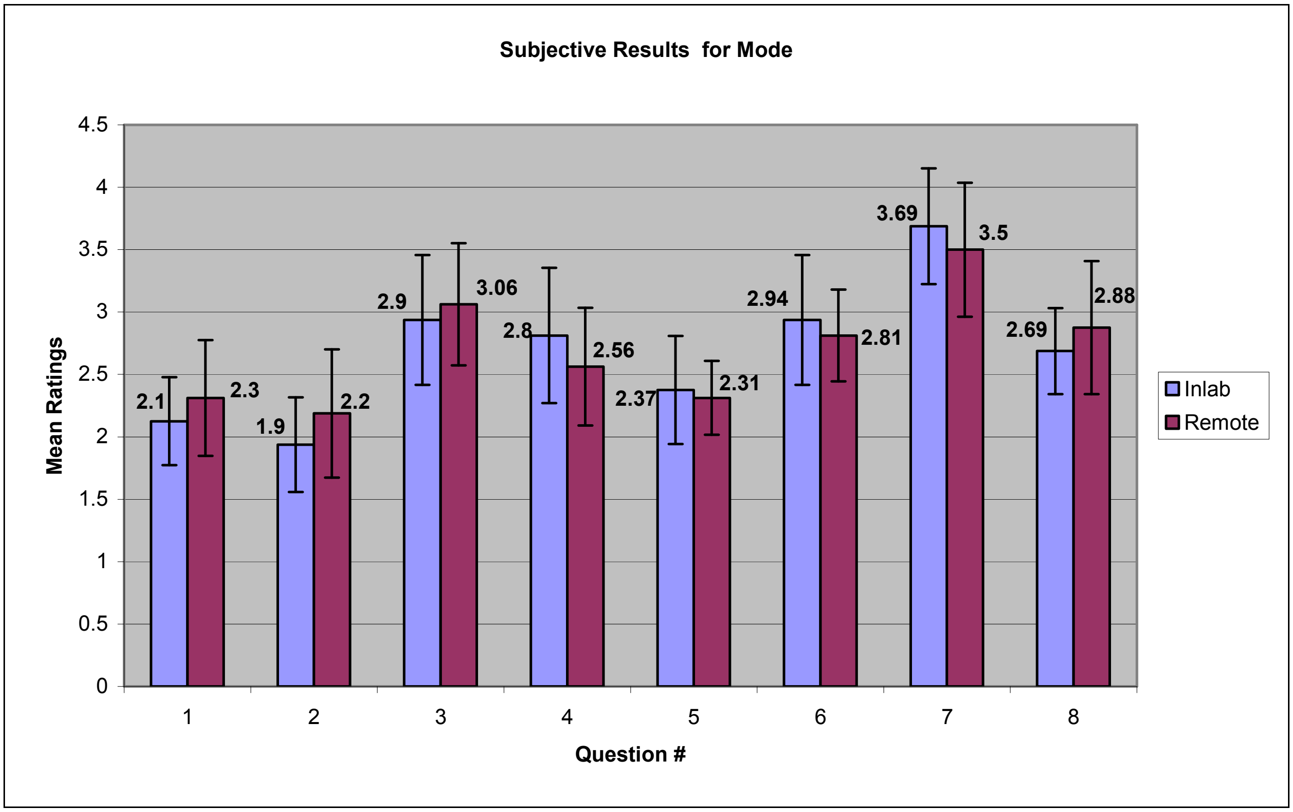


Figure 23: Ratings for Inlab vs Remote testing for Lab Questionnaire

5.5.3. Analysis of Number Of User Comments.

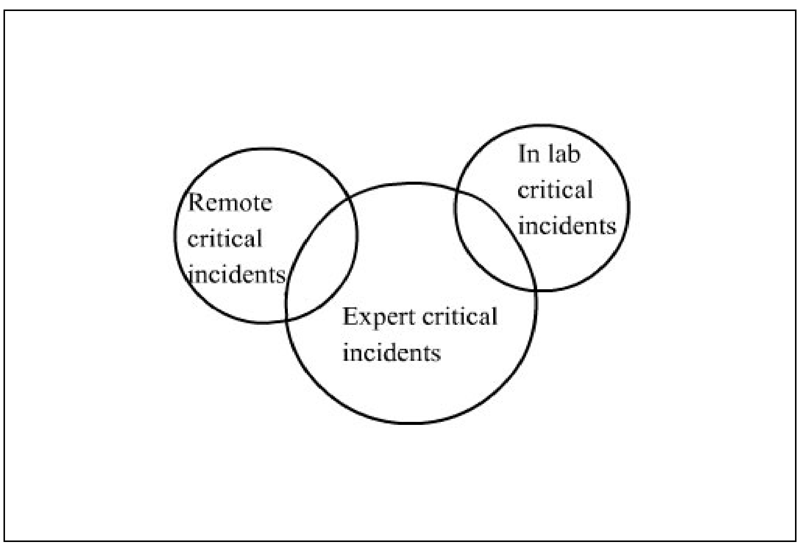


Figure 24: Ultimate criteria and Actual criteria.

Figure 24 shows the ultimate criteria and actual criteria. The number of user comments collected was analyzed using the ultimate and actual criteria. Two experts were asked to report user comments and the mean number of user comments collected per participant in both modes was compared to the mean number of user comments reported by experts. Experts did not come to an agreement on the number of user comments. One of the experts reported two extra user comments.

The unique set of user comments reported by experts included 20. Example of a user comment is " It was presented in a clear manner". A t test was conducted to see if the number of user comments collected from lab is significantly different from number of user comments collected from remote testing. It was noted that there was no significant difference between the number of user comments collected from inlab testing to the number of user comments collected from remote testing. Table 6 shows the mean user comments collected from inlab testing, remote testing and experts. Table 7 and 8 show the Summary of User Comments collected from inlab testing and remote testing respectively. Table 9 shows the user comments collected from Experts.

Table 6: Mean Number of user comments collected.

Mean number of user comments from inlab testing	Mean number of user comments from remote testing	Mean number of user comments from experts
16 (1.22)	12 (1.46)	20 (1.41)

Table 7: Summary of User Comments Collected in Inlab Testing

Abstract		*Concrete*		*Tabular*		*Graphical*	
Positive	*Negative*	*Positive*	*Negative*	*Positive*	*Negative*	*Positive*	*Negative*
I liked the icons.	I had difficulty in understanding the icons because of no explanations	It was easy to interpret what foods were being consumed and whether they were good or not.	Display needs more explanation on what it is.	This screen was also very clear in food allowed and what the individual had eaten.	The significance of the colors on the bars should be explained.		This was difficult to interpret - lines indicating hours would have helped.
The colors were attractive.	No real problem, but the calendar detracts from the main purpose of the screen.	Easier to understand than the other screens.	Uncertain at first whether the faces were “required intake” or “reported intake” of an individual.		Buttons are not uniform and therefore not clear.		Need larger fonts.
The colors were easy to understand.	Relationship between various icons and fields are unclear.		Faces show only ranges and don't give specific data.				It would seem reasonable to start the chart at breakfast instead of midnight.
	Need larger fonts.		All displays need to be a little larger.				
			Screen was somewhat confusing.				

Table 8: User Comments Collected from Remote Testing

Abstract		*Concrete*		*Tabular*		*Graphical*	
Positive	*Negative*	*Positive*	*Negative*	*Positive*	*Negative*	*Positive*	*Negative*
It is clear.	Print a legend.	I thought the face expressions and colors were more obvious than the squares.	There was no explanatio n of the meaning of the colors.	Good layout.	The screen should define the values of the colors to remove room for individual interpretat ion and thus error.	This seems a good way to present the information involved.	Time increment s were not clear ***(negative)***
Very attractive.	Need larger fonts.	Very colorful.	Larger fonts required.	With the number of servings given, there is no need to interpolate .			I find graphs harder to use than tables. ***(negative)***
							Make time intervals every hour or even every half hour. ***(negative)***
							Font should be larger on this screen. ***(negative)***

Table 9: User Comments Collected from Experts

Abstract	*Concrete*	*Tabular*	*Graphical*
Stoplight formatting is very effective.	Faces are very cheerful.	It is tough to get what the colors mean, since in this case there is no red, only yellow and green squares.	Tough to understand the screen.
Legends will be more helpful.	Adding colors to faces is a good idea with redundant coding.	Very easy to answer questions.	Its tough to make out the actual time on the graph without any demarcation.
More place to answer task questions needed.	Explanation of color-coding required.	Easy to understand.	Very tough to view on high density screen.
A column header for the boxes is needed.	Also telling the user what the faces stand for is required.		Wording is not clear, with the starting time being midnight on the graph.
Fonts need to be bigger on high density screen.	Font size is too small in high density screen.		Colors are nice and pleasing.
Too much clutter on high density screen.	Faces don't give any specific data.		Extra information on the screen which is not being tested.

The total number of user comments collected by inlab testing is more than the total number of user comments collected by remote testing. t test was conducted to analyze the number of user comments collected. No significant difference was noted between the number of user comments collected from inlab testing and experts or between remote testing and experts. Also the user comments were characterized as negative and positive comments. It was noted that there was no significant difference between the number of positive comments and negative comments collected between inlab testing and remote testing. The mean number of user comments collected from a participant in inlab testing and remote testing is given above. Appendix I shows the expert comments and the design recommendations.

6. DISCUSSION

6.1. Answering the Research Questions

6.1.1. Research Question 1

The first research question explored the difference in performance between older adults when the information is displayed in concrete form compared to abstract form?

The hypothesis for this research question was that the concrete representation, when compared to the abstract representation, will be a more usable format of representation, as measured by performance, for older adults within the given interfaces.

Results did not show a significance level of .05. Therefore, the hypothesis was not supported. From the results it was observed that the participants took almost the same time to complete tasks in a concrete display and abstract display. Also it was seen that participants were more accurate while performing tasks using concrete display when compared to abstract display, though no significant difference. There was a clear trend that was observed.

6.1.2. Research Question 2

The second research question explored the difference in performance between older adults when the information is displayed in graphical form compared to tabular form?

The hypothesis for this question was that the tabular representation, when compared to the pictorial representation, will be a more usable format of representation, as measured by performance for older adults within the given display representation format.

A significance level of .05 was not obtained for this hypothesis. Therefore, the hypothesis was not supported. From the results it was observed that the participants took more time to complete tasks in a graphical display than in tabular display though there was no significant difference. Also it was seen that participants were more accurate while performing tasks using tabular display when compared to graphical display, though no significant difference. There was a clear trend that was observed.

6.1.3. Research Question 3

The third research question asked what is the difference in performance between older adults when the screen density is high than from when it is low?

The hypothesis for this question was that the low-density interfaces, when compared to the high-density interfaces, will be a more usable interface, as measured by performance, for older adults within the given interfaces.

The results indicated that older adults perform significantly faster when working on a low-density screen as compared to high density screen, but there is no significant difference in accuracy between low density and high density screens. In this study density was operationalized as having the same amount of information in both high density and low density, but more information was cluttered in small amount of space in high density and more spread over the screen in low density.

6.1.4. Research Question 4

The fourth research question asked what are the reasons for user- preferences of each of these display formats and which is preferred more?

The hypothesis for his question was that reasons for user-preference of these display formats will be the usability of the displays, the ease of comprehension, the ease of use and ease of mapping as measured by older adults with the given interfaces.

The results indicated that the ease of comprehension, the ease of use and ease of mapping was the reason for user preferences and abstract display was most preferred by older adults.

6.1.5. Research Question 5

The fifth research question asked to which mode of usability testing is more effective in terms of time, data collected and preference when comparing remote usability testing to inlab testing?

The hypothesis was that the remote usability testing will be as effective as inlab testing as measured by time, data collected and preference for older adults within the given interfaces

The results indicated that older adults perform significantly faster in inlab testing when compared to remote testing. Also there is no significant difference in number of user comments collected from remote testing and inlab testing. Subjective ratings did not show any significant difference in preference between inlab testing and remote testing. It was noted that quality of user comments was the same in both cases.

6.2 Implications of this Research

To summarize, significant results were found in relation to the idea that participants performed faster in lower density screens than in higher density screens. Upon further examination one sees that age and computer experience were found to be significant indicators of performance, with decreasing age ($r = -0.46$) and increasing experience ($r = 0.65$) both being associated with increased performance.

It was also observed that participants performed significantly faster in inlab testing as compared with remote testing. This could be associated to the fact that participants when in the lab are more focused to performing the test and are less distracted.

Participants did not have any significant difference in the time they took to complete the tasks with respect to different displays though participants took longer time to complete tasks using a graphical display as compared to the other displays. Although the participants in higher density took more time to do the tasks than the participants in lower density, no significant difference was noted in their accuracy. This led to the conclusion that there was no change in participant's accuracy between high density and low density nor was there a change in accuracy if the study was conducted inlab or remote, nevertheless interactions indicate that further investigation may be warranted. Also, though results showed that participants performed tasks more accurately in concrete and tabular displays, no statistical significance was observed. This may indicate that performance of participants is significantly affected by density and location of testing more than the way information is presented to them.

It was observed that participants in inlab testing gave more user comments than participants in remote testing. This may be because of the experimenter's presence in the lab with the participant and the fact that in remote testing the participant is not observed. Also it was observed that feedback and user comments collected depend on individual differences. Some participants tend to give more feedback than the rest.

Significant results were obtained for displays for subjective ratings of preference. It was observed that 37.5 % of participants preferred abstract displays, 25 % preferred concrete displays, 18. 75% preferred tabular displays, 15.625% preferred graphical displays and 3.125 % preferred all the displays. It is believed, by the author, that it is possible because the user liked the colored blocks used in the abstract display to represent the information. All the participants who marked abstract display as their most preferred display noted that they felt the counting of the number of blocks, which had colors ranging from green to red did give them a good idea of how the interface worked in a few minutes. Participants preferred the graphical display the least, and this could be possible due to the fact that participants had to take time and decipher the graph before they could answer the questions. There is a possibility that ratings might have changed if the graphical display had more accurate demarcations and precise numbering. Participants who preferred concrete displays indicated that they preferred it due to the face icons. This is supported by literature review that according to assimilation theory (Ausubel, 1968 in Wiedenbeck, 1999), an interface that represents an object in a concrete manner will help the individual in assimilating new computer concepts to a correlated base of concepts in the memory. Wiedenbeck (1999) states that this assimilation is possible because of the analogies of the new situation to the known situation. It also seems possible that these older adults prefer lively colorful displays. Participants who preferred the tabular display commented on the ease of interpreting information from tables. Participants liked the fact that they just had to read a table to decipher the information while in other displays they had to interpret the data using other coding as well like color-coding. It was noted by the author, that participants' reactions did not appear to vary according to their accuracy in performing the tasks, but on the time they took to interpret the data as well as the colors used in it and ease of interpretation of data. No significant results were found using correlation test but subjective data is used as supporting evidence.

The subjective ratings indicated that participants found it easiest to answer questions in tabular and concrete displays, which also match the accuracy ratings. From an information processing point of view, this means that it is easier for the user to process information displayed in tables and with concrete icons. It is believed, by the author, that this is possible because the participants were sure of the answers for the task questions and so the ratings for no difficulty in answering questions. Another interesting result that had to do with subjective ratings was that which referred to the liking of the interface and icons used in the interface. Significant results were observed between ratings for graphical and concrete displays. Participants seemed to like the face icons more than the graphical display. This could be attributed to the fact that though participants preferred color, they liked viewing pleasing displays. This is supported by Rowe (1999) who stated that color is a powerful tool to improve user interfaces for older adults. It was also noted that there was a significance difference between the subjective ratings of how much the participants liked the low-density screens versus high-density screens. It is possible that older adults preferred less cluttered screens with bigger fonts, which are easier to read and interpret. This is similar to findings by Nielson (2002). Subjective results related to confidence and comfort in using the interfaces gave significant results. Results indicated that the participants were confident and comfortable with low density interfaces more than high density interfaces. This is similar to findings we noted earlier where participants preferred low density screen displays more than high density screens. Another interesting result that had to do with subjective ratings was that which referred to the ease of learning icons and the interface. Results indicated that participants felt that tabular and concrete displays were easier to learn than abstract displays and graphical displays. This could be possible because tabular display did not require much thought processing and just required reading the table values. It is believed by the author that concrete displays were preferred in this case more because of the color combination and the love of older adults for lively expressive displays.

Subjective ratings, which did not show any significance, also had some interesting find. The results showed that participants found the concrete displays most appealing and graphical displays least appealing. This could be associated with the fact that participants were not comfortable using the graphical displays due to the difficulty in interpreting the time differences. If the graphs had more distinct time demarcations, the results might have been different.

This result did not show any significant differences, but it's still noteworthy. There are few other things to be noted with reference to the subjective rating results. Most of the participants noted that they would prefer a legend with all the color-coding explained for all the four displays.

A further area to reiterate with particular reference to the results of the time measure, there may have been a sort of generational effect, which the researcher observed, was that most participants did not appear to see the value in just playing with a computer to learn it. There were only a dozen participants who gave the impression that they had good experience with technology, and the mean age was 65. It seems that the younger generations, who grew up when computers became fairly commonplace, might have expressed different notions. However, many of the participants expressed fears to the researcher about pressing the wrong key because they remembered losing a great deal of work that way in the 1980's. So, perhaps the results might be related to the computer experience the participants had. In the 1980s, surveys taken amongst business executives concluded that older managers resisted the introduction of computers on their desktops (Hakala, 1996).

The author noted from pilot studies that participants made the verbal comment that they wanted to see everything they were asked to do in the test written down step-by step from how to login to the tutorial to how to note user comments to how to answer the task questions and questionnaires. But it was also observed that after the first display, the participants would not refer to the written instructions and tried to perform tasks based on how they performed it for the first display.

One last and important point to understand about this research is that it only measured performance of older adults using displays designed by the author. There was no testing on different kinds of concrete displays or different kinds of abstract displays etc. One might hypothesize that, if the screens are designed in a different manner, then different results may be obtained when testing in conditions similar to this study that measure performance or preference. Of course, future research would have to be conducted supporting this idea for it to obtain substantiation.

6.2.1 Recommendations

Aspects of the displays that the author would recommend changing mainly have to do with giving the user more scaffolding or training wheels to reassure the user (Echt, Morrell & Park, 1998; Rogers, Meyer, Walker, & Fisk, 1998). Some studies have indicated that today's older adults appear to have higher levels of computer anxiety than younger adults (Languna and Babcock, 1997). This idea is also supported by the author's experience. During the study, few older adults indicated that they were not very familiar with computers and are skeptical at their abilities to perform tasks on a computer. Therefore, the design changes suggested involve giving them extra reassurance so that they are less likely to experience an overabundance of stress, which could result in either not finding the experience as enjoyable or not remembering as much of what they have finished. For remote testing sessions extra assurance can be given via phone to the user.

Figures 25- 28 show the redesigns recommended for the display designs. If an individual wishes to utilize displays principles with older adults in the future, the following changes are recommended:

1) Provide good legends explaining the button, icons or color-coding used in the display. From subjective data collected it was very clear that legends would help the participant better understand the display.

2) Provide colorful displays, pictures seem to do well for this, to get the user's attention. Users indicated that they like colorful displays. Also the study has indicated that user preferences of the displays are dependent on its color and pleasing nature.

3) If designing web applications for seniors, give feedback at intervals while working on real tasks rather than waiting until the individual has completed a task in its totality. This way the participant will not lose interest and also this will help in getting more feedback from the participant. This could be a visual feedback that appears when the user performs a task letting them know if they have completed it correctly or not. In the study it was observed by the author that the participant was not clear when he/she had answered a task question whether it was correct or not and how to proceed further.

4) Use more tables than graphs while displaying information to seniors. From the results of the study, it was indicated that the tabular display is preferred more by seniors than graphical displays.

5) Also use low-density screens when displaying information to older adults. From the results of the study, it was clear that older adults performed significantly better in low-density screens as compared to high-density screen.

There are also two general things to note that the author has not seen in previous literature. The following are general points for all individuals who wish to design applications for older adults:

1) Older adults would do similar in inlab testing and in remote testing if the test involves just viewing the screens and performing tasks. But if they have to report user comments themselves, then there needs to be good amount of online instructions frequently to report it. Also giving positive feedback after each user comment is reported would be ideal.

2) It was also observed that older adults reported more user comments when observed rather than when left by themselves to perform tasks. So a good way to conduct remote testing would be to have an online software to capture the participant and have the experimenter observe the participant from his or her facility. This way we can increase the number of user comments collected from seniors during remote testing.
3) If using computer applications, quick reference and overview sheets are not likely to help older adults to supplement their working memory because, once immersed in the new software environment, they are likely to have forgotten about them. Nevertheless it will help them to get started and overcome the initial fear. So, for older adults especially, it will be recommended to make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

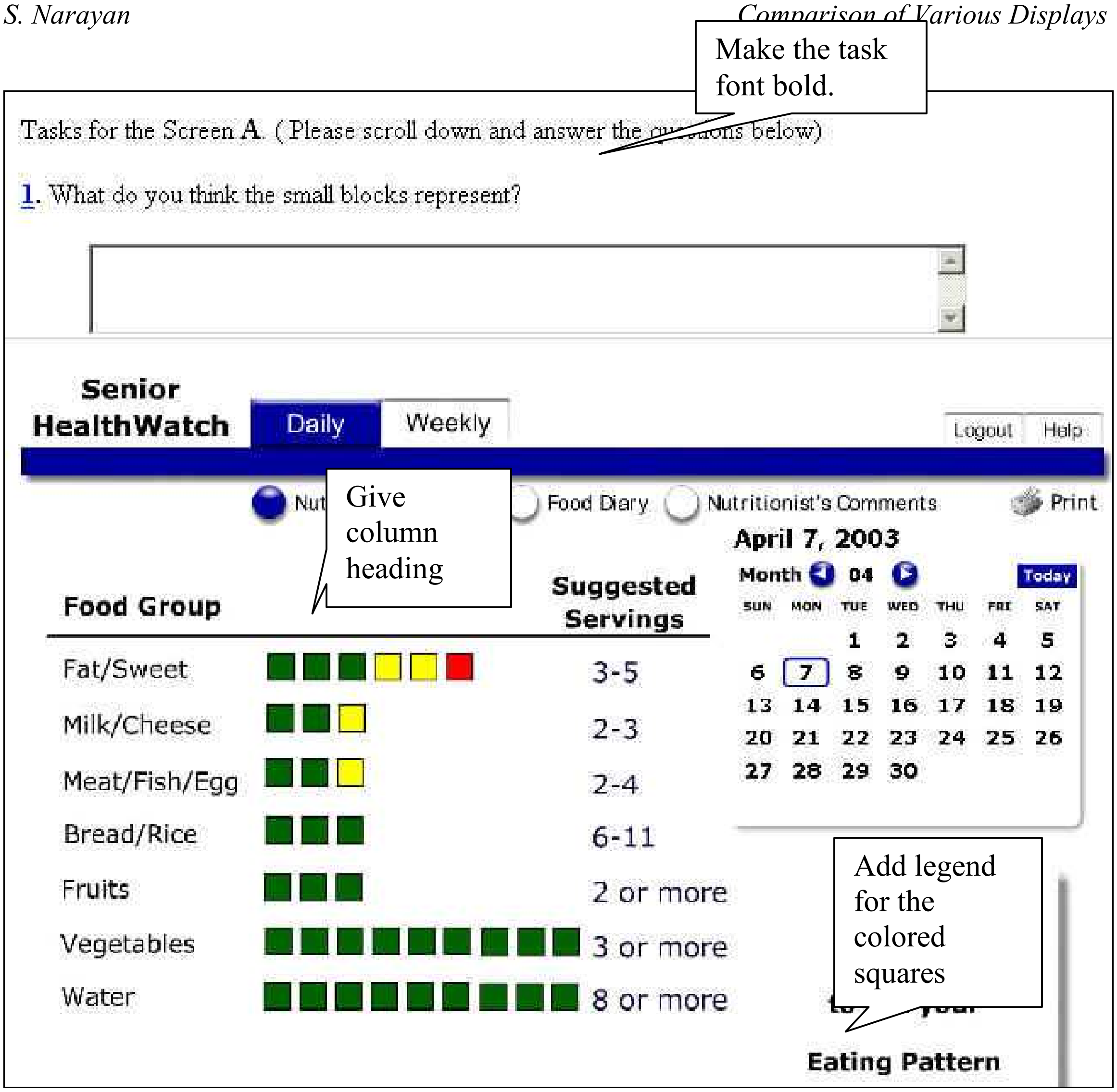


Figure 25: Redesigns for Abstract display

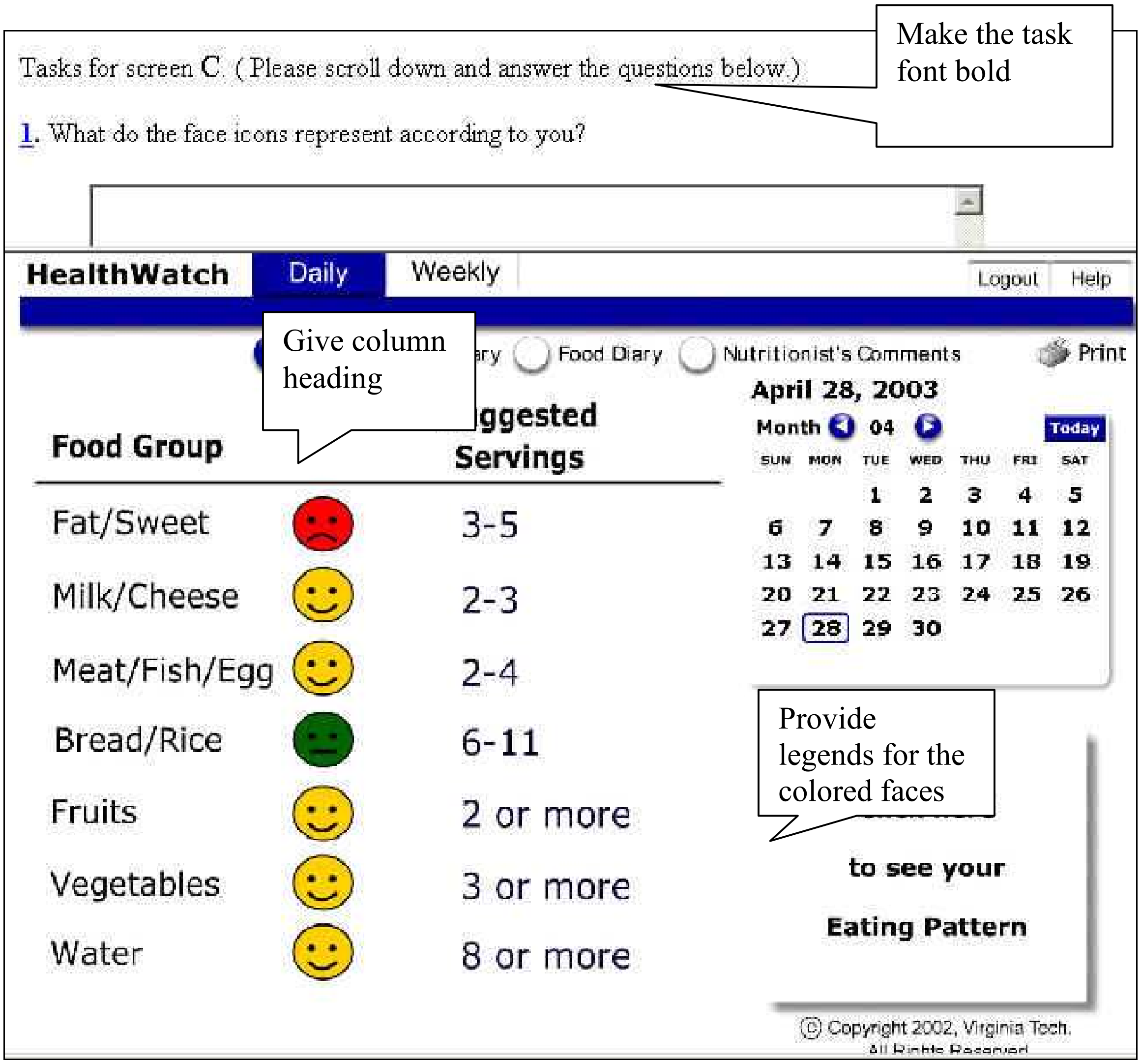


Figure 26: Redesigns for Concrete display

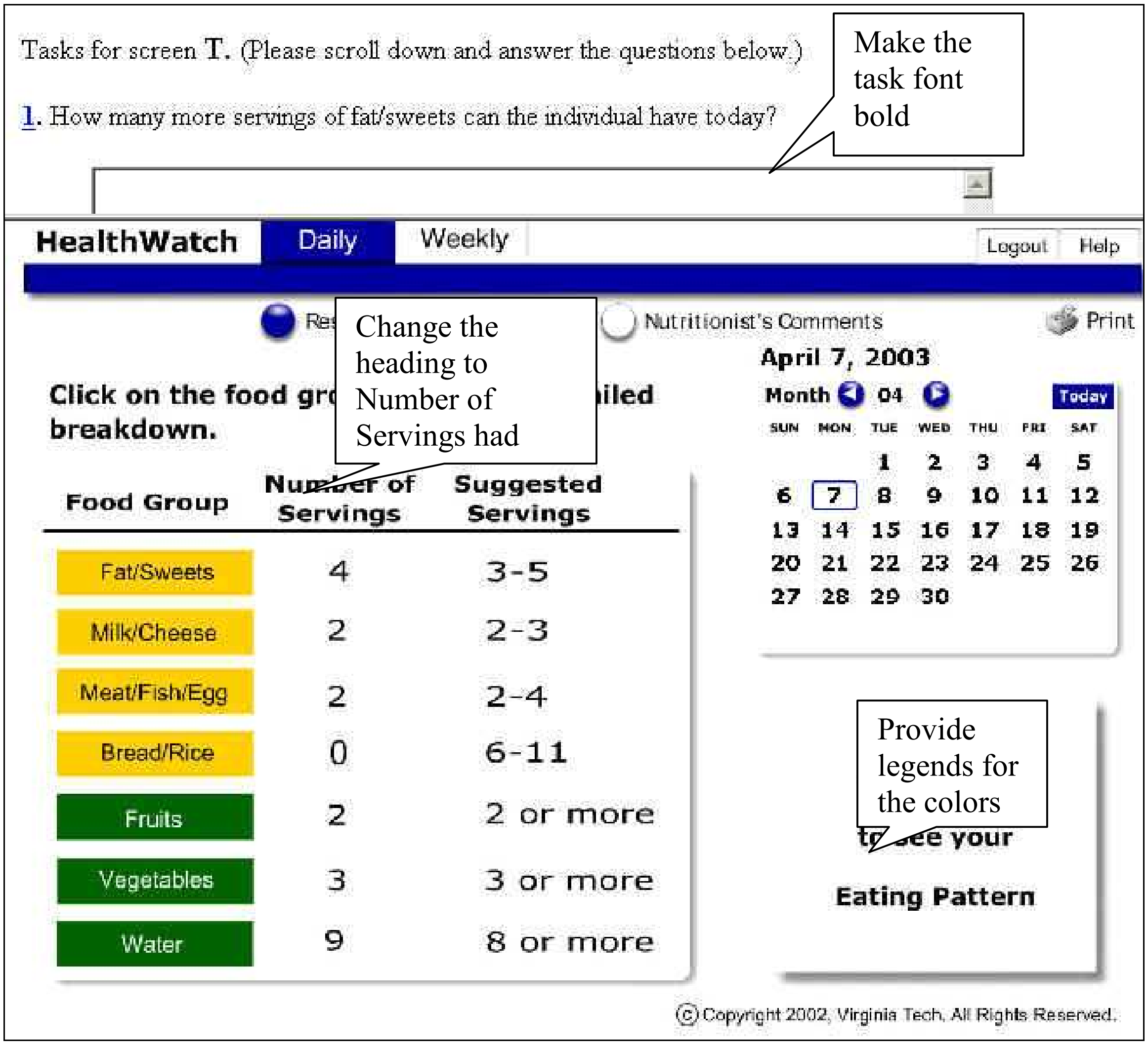


Figure 27: Redesigns for Tabular display

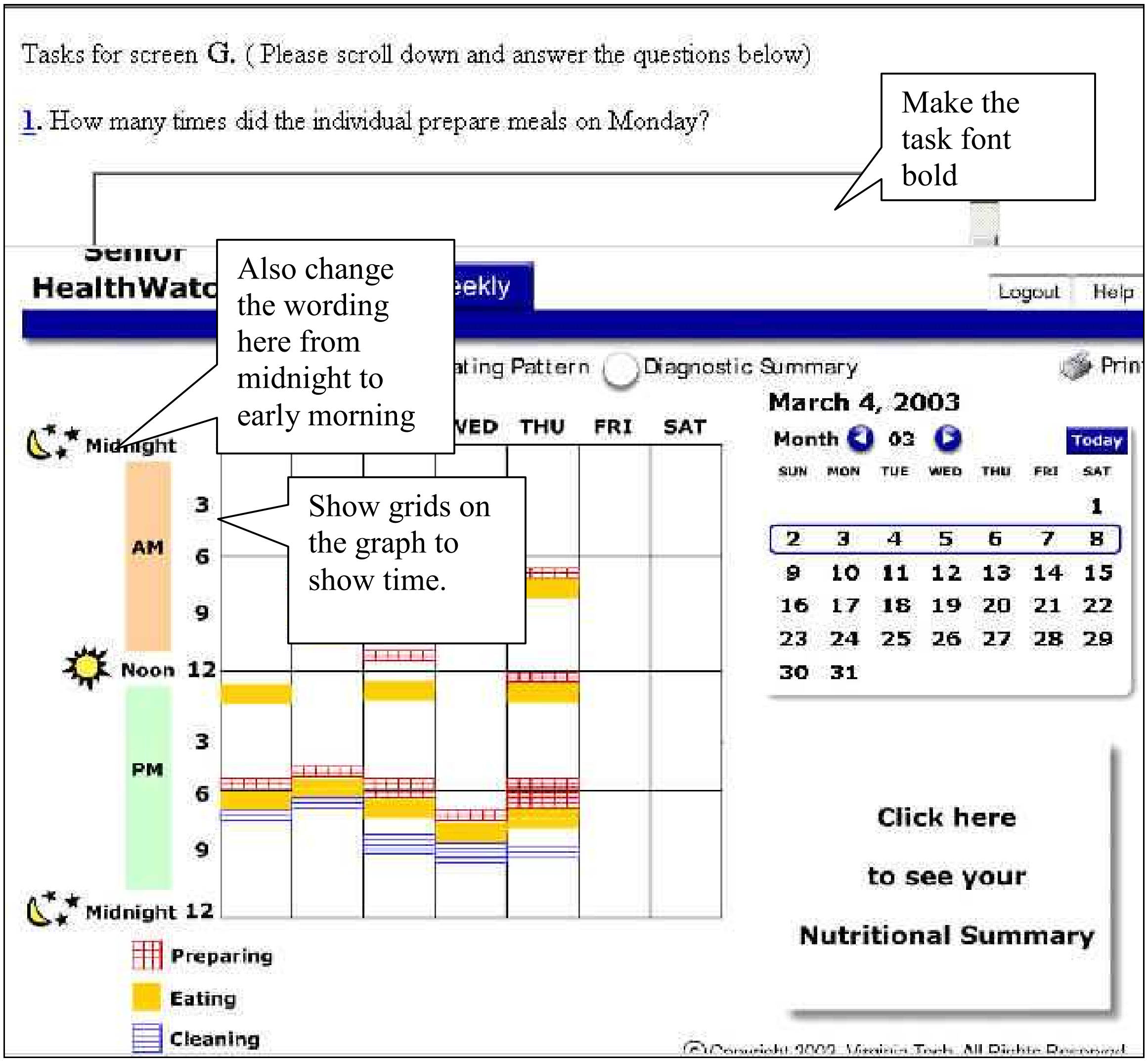


Figure 28: Redesigns for Graphical display

6.3. Conclusions

6.3.1. Relationship to Existing Literature

Supporting the statement of Tullis (1984) that the truly strong predictive role for performance outcomes is density of the screen, this study found that the predictors with the strongest relationship was density of screen. Supporting Hutchin et al.'s (1985) hypothesis that concrete icons have direct mapping between form and function and are easier to infer, the results of the study did not show any significant difference in performance between concrete and abstract displays but, it was noted that participants were more accurate while performing tasks in concrete icons. Also the subjective ratings indicated that users significantly felt the concrete icons were easier to learn than abstract icons. The trends in the current study indicated that a relationship may exist. Wiedenbeck (1999) stated that there is significant advantage for text only interface and icon text interfaces when compared to icon only interfaces for correctness and time. However in this study participants were more accurate while performing tasks in concrete displays and tabular displays compared to graphical and abstract displays, no significant results were obtained in terms of time and accuracy. Rogers (1989), state that icons depicted that concrete icons tend to be more effective but people can learn the arbitrary relationships of abstract icons when few are being used, but not when embedded in a whole set of abstract icons. The results of this study do support this argument, this argument could possibly explain the significant high preference for abstract displays, though participants performed more accurately and learned concrete displays faster. There was not enough significance to merit drawing compelling conclusions about performance.

Davis et al. (1989) stated that the perceptions of the ease of use of the system are affected by surface features of the system such as the use of icons in the interface, the color and information presentation formats. The subjective ratings of the study indicated that older adult's preferences of the interfaces were influenced by use of icons, color and information presentation formats.

Martin (2000) stated that a change in behavior can be observed due to the experimenter's paying attention to the participant, rather than the effects of the independent variable, which he referred to as the Hawthorne effect. This might explain the increase in number of user comments collected in inlab testing as compared to remote testing. This design would hope to get information on the display design for seniors while additionally capturing higher subjective ratings by older adults.

6.3.2. Limitations and Future Research

None of the results referring to age, and experience were accounted for in a pre-planned way. Therefore the generalization of the results regarding these factors is limited. Pre-planned comparisons that examine order effect and computer experience should be conducted using a greater sample size and more equivalent numbers in each group to obtain more reliable results. Future research should also limit or account for age and gender, perhaps by designing levels of these into their experimental design. Due to lack of control group we cannot be sure if the results are specific only to older adults, therefore another interesting way for future studies to build on this research would be to compare different displays, on older adults and young adults and compare the differences on an approach that was altered based on the recommendations given in this paper. Ideally, this would include the above factors (gender, experience, order effect etc.) to get a better idea about interactions with this new approach, and it would also look at both performance and subjective data.

The possible lack of sufficient motivation may also have had an unwanted effect, and future studies should be designed so that motivation is ensured and payoffs of the goals are evident. In addition, the deep and abstract nature of the displays may have caused some unanticipated effects. Future research would do well even to simply replicate this study using a less abstract and complex topic, or environment.

Future research can also be conducted on how information processing is effected by the display design. Research on how the issues such as redundant coding, visibility of high density screens, search time influence of density and mental workload would affect subjective and objective measures of display performance will be useful for future purposes.

No control group was used in the study. Therefore, it was not determined if the results found were more specific to older adults or it could have pertained to younger users as well. Future research would do well even to simply replicate this study using younger adults as well as older adults.

One last point for future research is that it would be useful to examine both remote and inlab testing of seniors using online software for reporting user comments, which will save all the data into a centralized database. This way it makes the procedure easier on the user and the experimenter during the test. Although this research might be more difficult and time consuming to conduct, it is believed that results of a study that examine this idea would be a great contribution, since users are generally interested in what they will be able to do on their own at their homes.

REFERENCES

Arend, U., Muthig, K.P., and Wandmacher, J. (1987). Evidence for global feature superiority in menu selection by icons. *Behaviour and Information Technology*, *6*, 411-426.

Ausubel, D. P. (1968). *Educational psychology: A cognitive view.* New York: Holt, Rinehart and Winston.

Benbasat, I., and Todd, P. (1993). An experimental investigation of interface design alternatives: icon vs. text and direct manipulation vs. menus. *International Journal of Man Machine Studies, 35*, 363-377

Berge, R.L., and Cassells, J.S. (Ed.). (1990). *The second fifty years: Promoting health and preventing disability*. Wasington, DC: National Academic Press.

Berkowitz, J.P., and Casali, S.P. (1990). Influence of age on the ability to hear telephone ringers of different spectral content. In W.A. Rogers (Ed.), *Designing for an aging population* (pp.184-188). Santa Monica, CA: Human Factors and Ergonomics Society.

Bewley, W.L., Roberts, J.L., Schroit, D., and Verplank, W.L. (1983). Human factors testing in the design of Xerox's 8010 "Star" office workstation. In *Proceedings of the CHI'83 Conference* (pp. 72-77). New York: ACM.

Birren, J.E. (1965). Age Changes in speed of response: Its central nature and physiological correlates. In A.T. Welford and J.E.Birren (Ed.), Behaviour, aging and the nervous system. Springfield, IL: Thomas.

Birren, J.E. (1974). Translations in gerontology-from lab to life: Psychophysiology and the speed of response. *American Physiologist, 29*, 808-815.

Birren, J.E., Woods, A.M., and Williams, M.V. (1980). Behavioral slowing with age: Causes, organization and consequences. In L.W. Poon (Ed.), *Aging in the 1980's Psychological issues*. Washington D.C.: American Psychological Association, 293-308.

Blankenberger, S., and Hahn, K. (1991). Effects Of icon design on human-computer interaction*. International Journal of Man –Machine Studies, 35,* 363-377.

Blattner, M.A., Sumikawa, D.A., and Greenberg, R.A. (1989). Earcons and icons; their structure and common design principles. *Human –Computer Interaction, 4*(1), 11-44.

Brown, C. (1988). *Human Computer interface design guidelines.* Norwood, NJ: Ablex.

Byrne, M.D. (1993). Using icons to find documents: simplicity is crucial. In *Proceedings of the INTERCHI'83 Conference* (pp.446-463). New York: ACM.

Carter, J. H., & Honeywell, R. (1991). Training older adults to use computers. *Performance and Instruction, 30,* 9-15.

Carrol, J.M., and Mack, R.L. (1986). Metaphor, computing systems, and active learning. *International Journal of Man Machine Studies, 22*, 39-57.

Castillo, J.C., Hartson, H.R., Hix, D. (1997). The user reported critical incident method at a glance.

Chan, M., Hariton, C., Rigeard, P., Campo, E. (1995). Smart House Automation System for the Elderly and Disabled. *IEEE, 1586*-1589.

Charness, N., Kelley, C. L., Bosman, E. A., & Mottram, M. (2001). Word-processing training and retraining: effects of adult age, experience, and interface. *Psychology and Aging, 16*(1), 110-127.

Czaja S. J. (1997). Computer technology and the older adult. In M. Helander, T. K. Landauer, & P. Prabhu (Ed.), *Handbook of human-computer interaction* (pp. 797-812). New York: Elsevier.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F.D., Bagozzi, R.P., and Warshaw, P.R. (1989). User acceptance of computer technology: a comparison of two theoretical methods. *Management Science, 35* (8), 982-1003.

delGaldo, E. M., Williges, R. C., Williges, B. H., & Wixon, D. R. (1987). A critical incident evaluation tool for software documentation. In L. S. Mark, J. S. Warm, & R. L. Huston (Eds.), *Ergonomics and Human Factors* (pp. 253-258). New York: Springer-Berlag.

Egidio, C., and Patterson, J. (1988). Pictures and category labels as navigational aids for catalog browsing. In *Proceedings of the CHI'83 Conference* (pp. 127-132). New York: ACM.

Familant, M.E., and Detweiler, M.C. (1993). Iconic reference evolving perspectives and an organizing framework. *International Journal of Man-Machine Studies, 39*, 705-728.

Fisher, D., and Tan, K. (1989). Visual Displays: The highlighting paradox. Human Factors, 31, 17-30.

Fisk, A., Scerbo, M., and Kobylak, R. (1986). Relative value of pictures and text in conveying information: Performance and memory evaluations. In *Proceedings of the Human Fcators Society 30th Annual Meeting* (pp.1269-1272). Santa Monica, CA: Human Factors Society.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*, 4, 327-359.

Galitz, W. (Ed.). (1985). *Handbook of Screen format design* (2nd ed.). Wellessley Hills, MA: QED Information Sciences.

Gillian, D.J., Wickens, C.D., Hollands, J.G., and Carswell, C.M. (1998). Guidelines For presenting quantitative data in HFES publications. *Human Factors, 40,* 28-41.

Gilly, M.C., and Zeithaml, V.A. (1985). The elderly consumer and adoption of technologies. *Journal of Consumer Research, 12*, 353-357.

Gittens, D.1986.Icon based human computer interaction. *International Journal of Man Machine Studies, 24*, 519-543.

Gould, J. D., and Lewis, C.H. (1985). Designing For usability- key principles and what designers think. *Communications of the ACM, 28,* 300-311.

Gould, J. D., and Lewis, C.H. (1983). In *Proceedings of the 1983 Computer Human Interaction Conference,* (pp.50-53).

Green, A.J.K., and Barnard, P.J. (1990). Iconic interfacing: the role of icon distinctiveness and fixed or variable screen location. In *Proceedings of the INTERACT'90* (pp.457-462). Amsterdam: North-Holland.

Hammontree, M., Weiler, P., and Nayak, N. (1994). Remote Usability Testing. *Interactions (July),* 21-25.

Hartson, H.R., Castillo, J.C., Kelso, J., and Neale, W.C. (1996). Remote evaluation: The network as an extension of the usability laboratory. In *Proceedings of CHI'96, ACM, Human Factors in Computing Systems*. (pp.228-235). New York, NY: ACM.

Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting with Computers, 12,* 507-528.

Hendrix, C. C. (2000). Computer use among elderly people. *Computers in Nursing, 18*(2), 62-68.

Hix, D., and Hartson, H.R. (1993). *Developing user interfaces: Ensuring usability through product and process*. New York, NY: John Wiley and Sons Inc.

Horton, W. (1994). *The Icon Book: Visual Symbols for Computing Systems and Documentation*. New York: Wiley.

Hutchins, E.L., Hollan, J.D., and Norman, D.A. (1985). Direct manipulation interfaces. In D.A. Norman and S. Draper (Ed.),*User Centered System Design* (pp.87-124). Hillsdale, NJ: Lawrence Erlbaum.

Johansen, R. (1988). *Groupware: Computer Support for Business Systems*. New York: The Free Press.

Lansdale, M. W. (1988). On the memorability of icons in an information retrieval task. *Behaviour and Information Technology, 7*, 131-151.

Larkin, J.H., and Simon, H.A. (1987). Why a diagram is sometimes worth ten thousand words. *Cognitive Science, 11*, 65-99.

Lin, J.J., Williges, R.C., and Beaudet, D.B. (1995). Accessible remote controls for older adults with mildly impaired vision. In *Proceedings o f the Human Factors and Ergonomics Society 39th Annual Meeting* (pp.148-152). Santa Monica: Human Factors and Ergonomics Society.

Martin, D. (2000). *Design of Psychology Experiments*. Belmont, CA:Watsworth.

Mead, S. E., Batsakes, P., Fisk, A. D., & Mykityshyn, A. (1999). Application of cognitive theory to training and design solutions for age-related computer use. *International Journal of Behavioral Development, 23*(3), 553-573.

Morrell, R. W., & Echt, K. V. (1997). Designing written instructions for older adults: Learning to use computers. In W. A. Rogers, and A. D. Fisk (Eds.), *Handbook of human factors and the older adult* (pp. 335-361). San Diego, CA: Academic Press, Inc.

Muter, P. and Mayson, C. (1986). The role of graphics in item selection from menus. *Behaviour and Information Technology, 5*, 89-95.

Navon, D. (1984). Resources: a theoretical soupstore? *Psychological Review, 91,* 216-23.

Kelley, C. L., Morrell, R. W., Park, D. C., & Mayhorn, C. B. (1999). Predictors of electronic bulletin board system use in older adults. *Educational Gerontology, 25,* 19-35.

Kwahk, J., Smith-Jackson, T.L., and Williges, R.C. (2002). An application of neural network modeling to diagnose eating behavior of seniors. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting,* (pp. 180-184), Santa Monica: Human Factors and Ergonomics Society.

Rogers, Y. (1989). Icons at the interface: their usefulness. *Interacting With Computers, 1,* 105-117.

Rogers, W., Gilbert, D.K., and Cabrera, E.F., (1994). An in-depth analysis of automatic teller machine usage by older adults. In W.A. Rogers (Ed.), *Designing for an aging population* (pp.184-188). Santa Monica, CA: Human Factors and Ergonomics Society.

Rohr, G. and Keppel, E., 1984, Iconic interfaces: where to use and how to construct? In H.W. Hendrick, and O. Brown (Ed.), *Human Factors in Organizational Design and Management,* (pp. 269-275). New York: Elsevier.

Sanders, M. S., and McCormick, E.J., (1993). *Human factors in engineering and design*. New York : McGrawhill, Inc

Scheiber, F., and Kline, D.W. (1994). Age differences in the legibility of symbol highway signs as a function of luminance and glare level: A preliminary report. In W.A. Rogers (Ed.), *Designing for an aging population* (pp.184-188). Santa Monica, CA: Human Factors and Ergonomics Society.

Shackel, B. (1984). Information technology-A challenge to ergonomics and design. *Behavior and Information Technology, 3*, 263-275.

Shepard, R.N., 1967. Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156-163.

Shneiderman, B. (Ed.), 1997, *Designing the User Interface: Strategies for Effective Human Computer Interaction* (3rd ed.), Reading, MA: Addison-Wesley.

Smith-Jackson, T.L. and Williges, R.C. (2001). Remote measurement methods for user-centered design of telesupport systems. *Assistive Technology*, **13**(2), 106-115.

Standing, L., (1993). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207-222.

Stephanidis, C., Salvendy, G., Akoumianakis, D., Bevan, N., Brewer, J., Emiliani, P.L., Galetsas, A., Haataja, S., Iakovidis, I., Jacko, J., Jenkins, P., Karshmer, A., Korn, P., Marcus, A., Murphy, H., Stary, C., Vanderheiden, G., Weber, G., and Ziegler, J. (1998). Toward an information society for all: An international R& D agenda. *International Journal of Human Computer Interaction, 10*, 107-134.

Stephanidis, C., and Salvendy, G. (1999). Toward an information society for all: HCI challenges and R&D recommendations. *International Journal of Human Computer Interaction, 11*(1), 1-28.

Story, M.F. (1998). Maximizing usability: The principles of universal design. *The Assistive Technology Journal, 10*(1), 4-12.

Tang, P., and Venables, T. (2000). 'Smart' homes and telecare for independent living. *Journal of Telemedicine and Telecare, 6* (1), 8-14.

Tullis, T. (1983). The formatting of alphanumeric displays: A review and analysis. *Human Factors, 25*, 657-682.

Tullis, T. (1984) A computer-based tool for evaluating alphanumeric displays. In *Proceedings of INTERACT'84 Conference on Human-Computer Interaction*, London, Sept. 1984.

Tullis, T. (1986a). Optimizing the usability of computer generated displays. In *Proceedings of HCI'86 Conference on People and Computers: Designing for Usability* (pp.604-613). London: British Computer Society.

Tullis, T. (1988). Screen design. In M. Helander (ed.), Handbook of human computer interaction (pp.377-411). Amsterdam: Elsevier Science.

Walker, N., Philbin, D. A., & Fisk, A. D. (1997). Age-related differences in movement control: Adjusting submovement structure to optimize performance. *Journal of Gerontology: Psychological Sciences, 52B,* P40-P52.

Waterson, S., Landay, J.A., Matthews, T.,(2002) "In the lab and out in the wild: remote web usability testing for mobile devices." In *Human Factors in Computer Systems: CHI 2002 Conference Extended Abstracts*, Minneapolis.

White, J. & Weatherall, A. (2000). A grounded theory analysis of older adults and information technology. *Educational Gerontology, 26,* 371-386.

Williges, R.C., Williges, B.H., and Elkerton, J. (1987).Software Interface Design. In G. Salvendy (Ed.), *Handbook of human factors* (pp.495-504). NewYork, NY: John Wiley and Sons, Inc.

Wiedenbeck, S. (1999).The use of icons and labels in an end user application program: an empirical study of learning and retention. *Behaviour and Information Technology, 18* (2), 68-82.

APPENDIX A: TASK INSTRUCTIONS

Thank you so much for your participation in this evaluation. We appreciate your willing ness to help us evaluate the Information Displays, which we hope will eventually help to develop information displays for seniors.

This session involves going through step-by-step instructions and then evaluation of 4 information displays (4 different computer screens). All the screens that would be shown to you consist of some aspects of the eating pattern or nutritional aspects of a senior. Two of the screens have displayed the number of servings in each food group an individual has had in that particular day. One of the screens has displayed the number of servings of each food group an individual has had in that particular week. And one has the eating pattern of the individual for the whole week.

During this session you will be asked to view each screen and answer questions related to that screen on the task questionnaire. After each screen you would be asked to fill out an online questionnaire related to that screen. You will also be given some time just to play around with it, to try whatever you would like to do. While performing some of the specific tasks, we may be timing how well the environment helps you with these tasks. Therefore we would like for you to work through each task without taking a break; you can take time to relax while filling out the questionnaire if you wish.

Because we are interested in why this environment is easy or difficult to use, we would like you to report comments using the remote evaluation tool. You should report using the remote evaluation tool both positive and negative aspects of how you perform the tasks and about different aspects of the screens shown to you (for eg: anytime you are confused or could not understand the meaning of a label or button or anytime you liked something about the screen).

Remember that you are helping us evaluate the information displays: we are not evaluating you. You should feel free to say and report whatever you think about any aspect of the environment or the tasks you are asked to perform.

Now please go through the step-by-step instructions to perform the test.

This session should last little over one hour.

Please report comments using the remote evaluation tool during your session.

STEP-BY-STEP INSTRUCTIONS

1. Open the remote evaluation toll which is located on your desktop
2. Enter your participant ID number given to you.
3. Click Enter.
4. Read the instructions on screen.
5. Click “Begin Session” button once you have completed reading instructions.
6. Click “OK” to continue.
7. Enter your participant id number on the field given as “Please enter your participant ID number”
8. Click “Submit Survey” button.
9. Click “Submit Survey” button once again.
10. Note down the screen name for the screen which is given at the very top of the screen. For eg, if the heading says “Tasks for screen B”, here screen name is B.
11. Now scroll down and view the screen, and answer the task questions.
12. Once done answering the questions, click on the remote evaluation tool below on the task bar, to report comments
13. Click on “Report Comments” button that pops up.
14. For the first question, please type in some name you can think of for the comment you are reporting.
15. For the second question, type in the screen name which you had noted down in step 10.
16. For the third question, please enter the comment you wanted to enter for the screen you just viewed. The comment can be a positive or negative feedback about the screen.
17. Click on “Next” button.

18. Click on "Next" button again once you have entered "Yes" or "No".
19. Click on "Next" button again once you have entered "Yes" or "No".
20. Click on "Finish" button.
21. Click on "OK" when the pop-up appears which says "The record has been written to the file successfully. Thank you".
22. Click on "OK" when the pop-up appears which says "Please mail the files in c:\Remote Evaluation directory to snarayan@vt.edu".
23. Click "Submit Survey" button at the bottom of the task questionnaire if you have completed answering all the task questions.
24. Now answer the questions in the questionnaire and click on "Submit Survey " button once you are done.
25. From this screen on follow the same procedure ad you did from step 10 through 24.

APPENDIX B : SUBJECTIVE QUESTIONNAIRE

Concrete Icons (Icons are pictures used to represent a particular idea, for eg: picture of a face, animals, shapes etc).

Instructions: Please respond to the following statements by filling in or marking the circle that most accurately depicts your opinion

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Overall, I found the interface (the computer screen previously shown) appealing	○	○	○	○	○
2. I liked the colors used in the screen	○	○	○	○	○
3. Overall, the interface (the computer screen previously shown) was not frustrating	○	○	○	○	○
4. I had no difficulty in answering the questions.	○	○	○	○	○
5. The fonts were clear to read.	○	○	○	○	○
6. I liked the face icons.	○	○	○	○	○
7. The icon representation was clear.	○	○	○	○	○
8. I am confident that I can use this interface (the computer screen previously shown) easily.	○	○	○	○	○
9. I will not feel uncomfortable trying to use this interface (the computer screen previously shown) again.	○	○	○	○	○
10. The interface (the computer screen previously shown) meets my expectations.	○	○	○	○	○
11. The icons were easy to learn	○	○	○	○	○

Please write responses to the following questions.

Did you like the interface (the computer screen previously shown) Why or why not?

What, if any, changes will you make to the interface (the computer screen previously shown)?

What did you like best about the interface (the computer screen previously shown)

Would you have preferred another font size or style?

What did you not like at all about the interface (the computer screen previously shown)

Abstract Icons (Icons are pictures used to represent a particular idea, for eg: picture of a face, animals, shapes etc).

Instructions: Please respond to the following statements by filling in or marking the circle that most accurately depicts your opinion.

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Overall, I found the interface (the computer screen previously shown) appealing	○	○	○	○	○
2. I liked the colors used in the screen	○	○	○	○	○
3. Overall, the interface (the computer screen previously shown) was not frustrating	○	○	○	○	○
4. I had no difficulty in answering the questions.	○	○	○	○	○
5. The fonts were clear to read.	○	○	○	○	○
6. I liked the icon representation	○	○	○	○	○
7. The icons were clear and easy to understand.	○	○	○	○	○
8. I am confident that I can use this display (the computer screen previously shown) easily.	○	○	○	○	○
9. I will feel comfortable trying to use this display (the computer screen previously shown) again.	○	○	○	○	○
10. The interface (the computer screen previously shown) meets my expectations.	○	○	○	○	○
11. The icons were easy to learn	○	○	○	○	○

Please write responses to the following questions.

Did you like the interface (the computer screen previously shown)? Why or why not?

What, if any, changes will you make to the interface (the computer screen previously shown)?

What did you like best about the interface (the computer screen previously shown)

Would you have preferred another font size or style?

What did you not like at all about the interface (the computer screen previously shown) and why not?

Tabular representation

Instructions: Please respond to the following statements by filling in or marking the circle that most accurately depicts your opinion

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Overall, I found the interface (the computer screen previously shown) appealing	○	○	○	○	○
2. I liked the colors used in the screen	○	○	○	○	○
3. Overall, the interface (the computer screen previously shown) was frustrating	○	○	○	○	○
4. I had no difficulty in answering the questions.	○	○	○	○	○
5. The fonts were clear to read.	○	○	○	○	○
6. I liked the tabular representation	○	○	○	○	○
7. The table was clear and easy to understand.	○	○	○	○	○
8. I am confident that I can use this display (the computer screen previously shown) easily.	○	○	○	○	○
9. I will feel comfortable trying to use this display (the computer screen previously shown) again.	○	○	○	○	○
10. The interface (the computer screen previously shown) meets my expectations.	○	○	○	○	○
11.The icons were easy to learn	○	○	○	○	○

Please write responses to the following questions.

Did you like the interface (the computer screen previously shown) Why or why not?

What, if any, changes will you make to the interface (the computer screen previously shown)?

What did you like best about the interface (the computer screen previously shown)?

Would you have liked more graphs than tables? Why?

Would you have preferred another font size or style? Why?

Graphical representation

Instructions: Please respond to the following statements by filling in or marking the circle that most accurately depicts your opinion

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Overall, I found the interface (the computer screen previously shown) appealing	○	○	○	○	○
2. I liked the colors used in the screen	○	○	○	○	○
3. Overall, the interface (the computer screen previously shown) was not frustrating	○	○	○	○	○
4. I had no difficulty in answering the questions.	○	○	○	○	○
5. The fonts were clear to read.	○	○	○	○	○
6. I liked the graphical representation	○	○	○	○	○
7. The graph was clear and easy to understand.	○	○	○	○	○
8. I am confident that I can use this display (the computer screen previously shown) easily.	○	○	○	○	○
9. I will feel comfortable trying to use this display (the computer screen previously shown) again.	○	○	○	○	○
10. The interface (the computer screen previously shown) meets my expectations.	○	○	○	○	○
11.The icons were easy to learn	○	○	○	○	○

Please write responses to the following questions.

Did you like the interface (the computer screen previously shown) why or why not?

What, if any, changes will you make to the interface (the computer screen previously shown)?

What did you like best about the interface (the computer screen previously shown)

Would you have liked more tables than graphs? Why?

Would you have preferred another font size or style? Why?

Final questionnaire

Please write responses to the following questions.

Which interface (computer screen) did you like the most and why?

Which interface (computer screen) was the easiest to comprehend and why?

Which interface (computer screen) are you most confident to use and why?

APPENDIX C: INFORMED CONSENT FORM

Participant Assent Form

Virginia Polytechnic Institute and State University

Informed Consent of Investigative Projects for Participants

Title of Project: Comparison of Various Display Representation Formats Using a Smart House Interface for Older Adults Using Inlab and Remote Usability Testing.

Principal Investigator: Sajitha Narayan, Grado Department of Industrial and Systems Engineering, 540-818-2388, snarayan@vt.edu

I. The Purpose of this Research

You are invited to participate in a research project that compares the information display formats such as abstract, concrete, tabular and graphical representations to identify the best possible information display for seniors and their likes and dislikes. The experiment will also be conducted both in the lab and at your home computer (remotely) to test the effectiveness of remote testing. The project will help in developing a display format, which is suitable to the senior population.

II. Procedures

Participants involved in the research will be recruited by several means such as emails and personal invitation from the Blacksburg and the surrounding area. If you agree to participate, you could be asked to take part in the following research activities:

- Some of the participants will have to do the experiment in the lab. These individuals will have to evaluate 4 sets of information display formats and perform a set of tasks. They will also have to fill out few questionnaires. We will like all participants to complete questionnaires throughout the study. Some of these questionnaires will ask for general background information (e.g., computer experience, profession) and some will be asking questions about the display with rating scales.
- Some of the participants will have to do the experiment in the house. These individuals will also have to evaluate 4 sets of information display formats and perform a set of tasks. They will also have to fill out few questionnaires. We will like all participants to complete questionnaires throughout the study. Some of these questionnaires will ask for general background information (e.g., computer experience, profession) and some will be asking questions about the display with rating scales.

III. Risks

There are less than minimal risks to you in this research. A small number of participants may experience eyestrain from using a computer screen.

IV. Benefits of this Research

While there are no direct benefits to you from this research (other than payment for lab-based activities), you may find the research interesting. Your participation and that of other volunteers should make it possible to better understand the information display preferences of seniors and the display formats suitable for seniors.

V. Extent of Anonymity and Confidentiality

No one other than investigator will have access to your questionnaire results without additional written consent from you. Likewise, no printed or electronic rendition of information that could be attributed directly to you will be available to anyone other than the investigators without additional written consent from you. Any visual data included in professional presentations or publications will be used anonymously. No information identifying you will accompany visual material.

You will be identified by a 3- digit study code. No reference will be made in oral or written reports that could link you to the data nor will you ever be identified as a participant in the project.

VI. Compensation

You will be paid $ 7.50 per hour.

VII. Freedom to Withdraw
You are free to withdraw from this study at any time without question.

VIII. Approval of Research
This research has been approved, as required, by the Institutional Review Board for projects involving Human subjects at Virginia Polytechnic Institute and State University and by the Department of Industrial and Systems Engineering.

IX. Participant's Responsibilities
As outlined above, if you agree to participate, your responsibilities may include:

- Possible participation in one or more 1.5-hour lab sessions at Virginia Tech, scheduled at your convenience.
- Possible participation in one or more 1.5-hour computer session at home.
- Completion of brief surveys in association with the study. Each survey is likely to require 5-10 minutes to complete.
- It is very important that you keep the activities and information discussed confidential, since others may be participating in this research.

X. Participant's Permission
I have read and understand the Informed Consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for my participation in the project.

If I participate, I may withdraw at any time without penalty.

______________________________ ______________________________ ________
Your Name (Please print) **Your Signature** **Date**

Should I have any questions about this research or its conduct I may contact:
Sajitha Narayan (Principal Investigator) ……........……..................... 540-818-2388
David Moore (Chair, IRB Research Division) ....…….. 540-231-4991

APPENDIX D: DEMOGRAPHIC QUESTIONNAIRE

Instructions: Please write a response to the following questions.

1. What is your age? _____

2. What is your gender? ___________

3. What job positions have you held?

____
__

____
__

4. Do you have any vision problem? _______________

Instructions: Please answer the following questions by circling the response that most accurately depicts your background and experience.

1. How many years have you used a computer?	0-1	2	3	4 or more
2. On an average, how many times do you use a computer per week?	0-1	2-5	5-7	8 or more
3. On an average, how many times do you e-mail and/or online chat per week?	0-1	2-3	4-5	6 or more
4. On average, how many times do you use an internet browser, such as, Internet Explorer, AOL, or Netscape per week?	0-1	2-3	4-5	6 or more
5. Have you ever used spreadsheets, such as in Microsoft Excel?	Yes		No	
6. Have you ever build WebPages?	Yes		No	
7. Have you ever made greeting cards or flyers online?	Yes		No	

APPENDIX E: INLAB TESTING QUESTIONNAIRE

Instructions: Please respond to the following statements by filling in or marking the circle that most accurately depicts your opinion

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Overall, I found the lab testing experience good.	○	○	○	○	○
2. I felt comfortable participating in the study.	○	○	○	○	○
3. I was able to answer the questions easily.	○	○	○	○	○
4. I had no difficulty in answering the questions.	○	○	○	○	○
5. The comments were easy to find.	○	○	○	○	○
6. I was able to report a lot of comments.	○	○	○	○	○
7. I will have preferred the experimenter taking down the comments	○	○	○	○	○
8. I had difficulty in reporting comments.	○	○	○	○	○

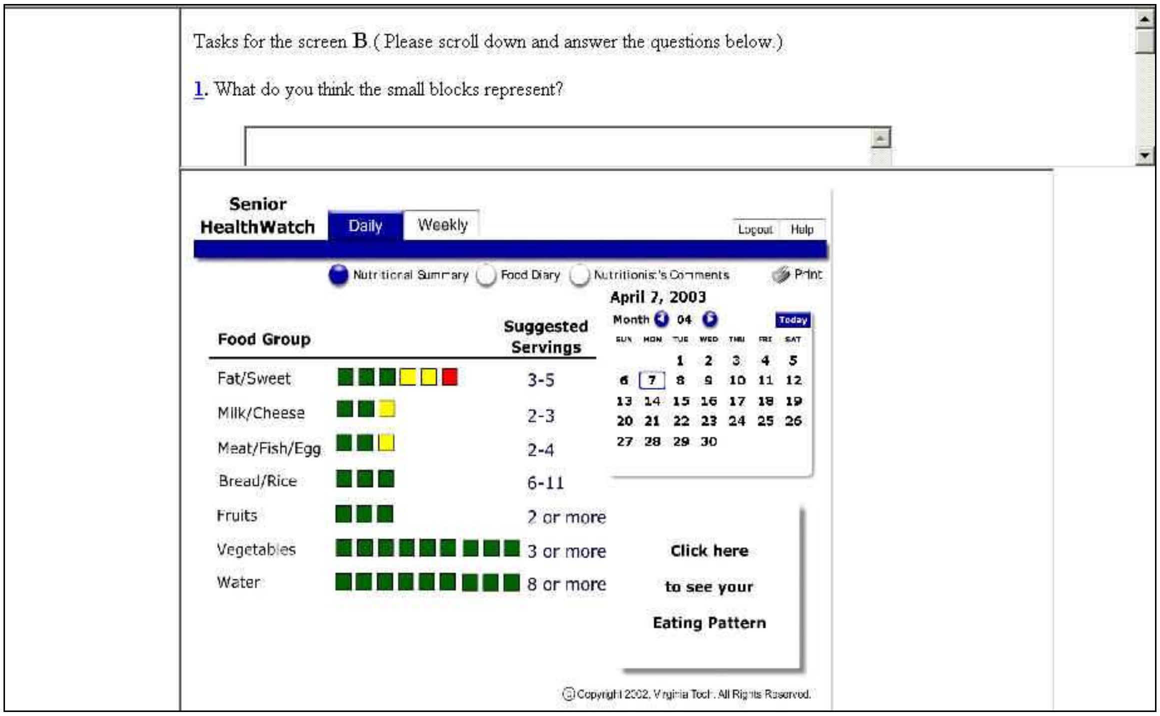
APPENDIX F: REMOTE TESTING QUESTIONNAIRE

Instructions: Please respond to the following statements by filling in or marking the circle that most accurately depicts your opinion

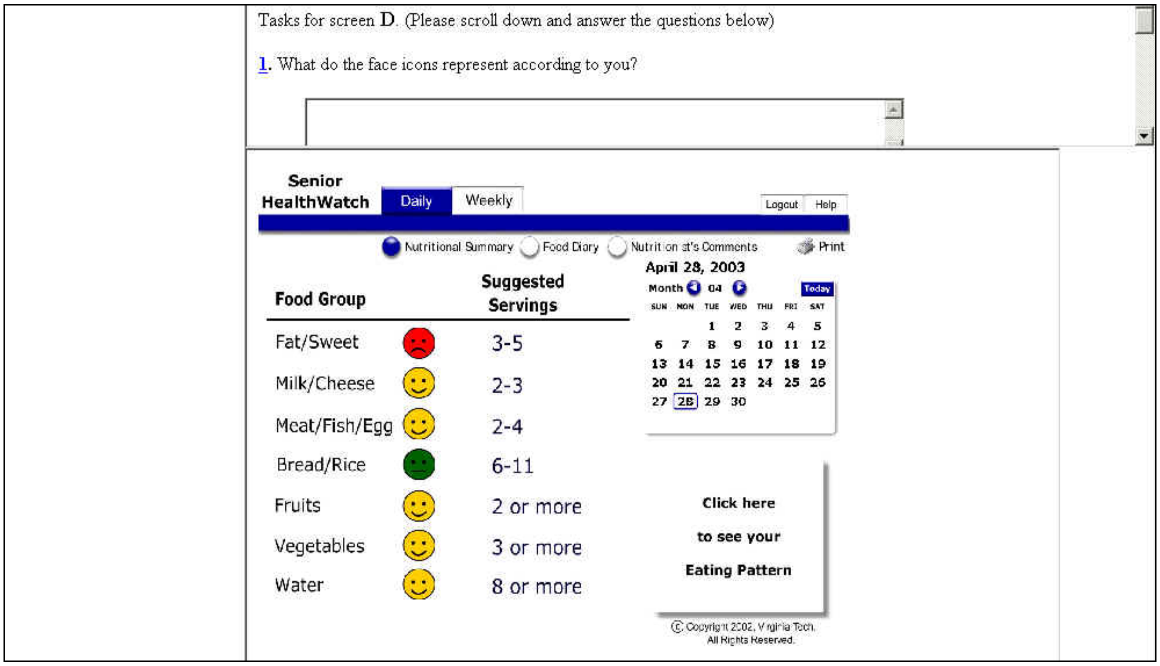
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. Overall, I found the testing experience good.	○	○	○	○	○
2. I felt comfortable participating in the study.	○	○	○	○	○
3. I was able to answer the questions easily.	○	○	○	○	○
4. I had no difficulty in answering the questions.	○	○	○	○	○
5. The comments were easy to find.	○	○	○	○	○
6. I was able to report a lot of critical incidents.	○	○	○	○	○
7. I will have preferred an experimenter to report comments.	○	○	○	○	○
8. I had no difficulty in reporting comments.	○	○	○	○	○

APPENDIX G: SCREENSHOTS

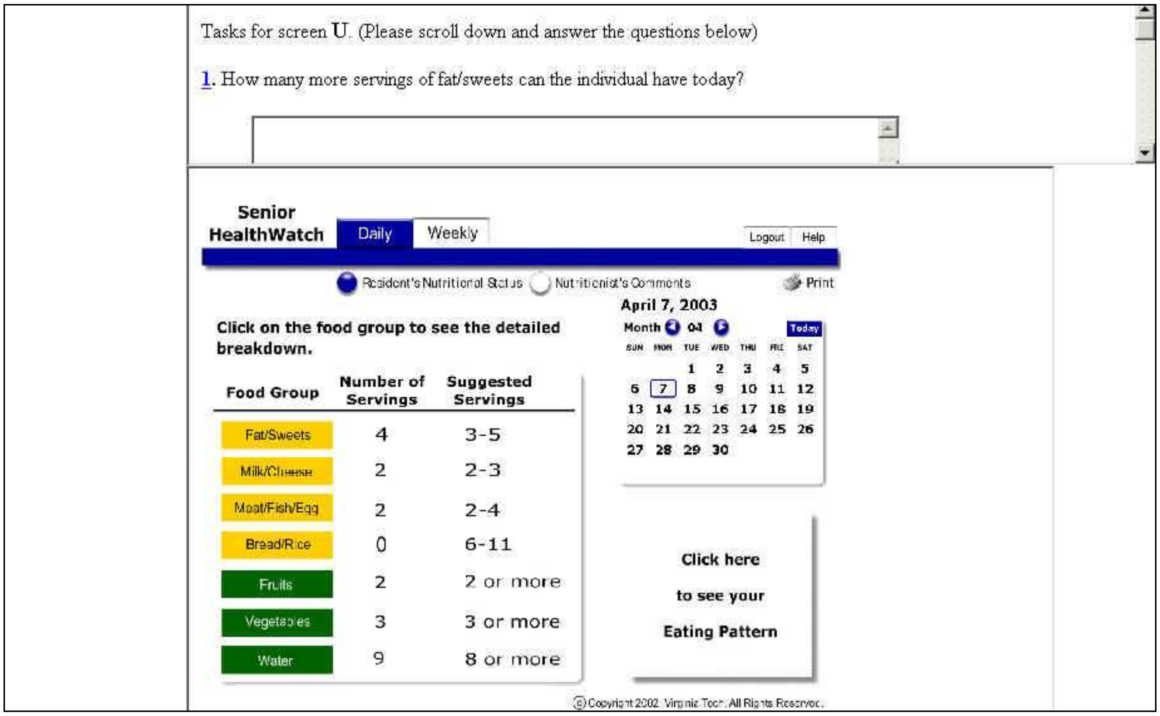
ABSTRACT DISPLAY FOR HIGH DENSITY SCREENS



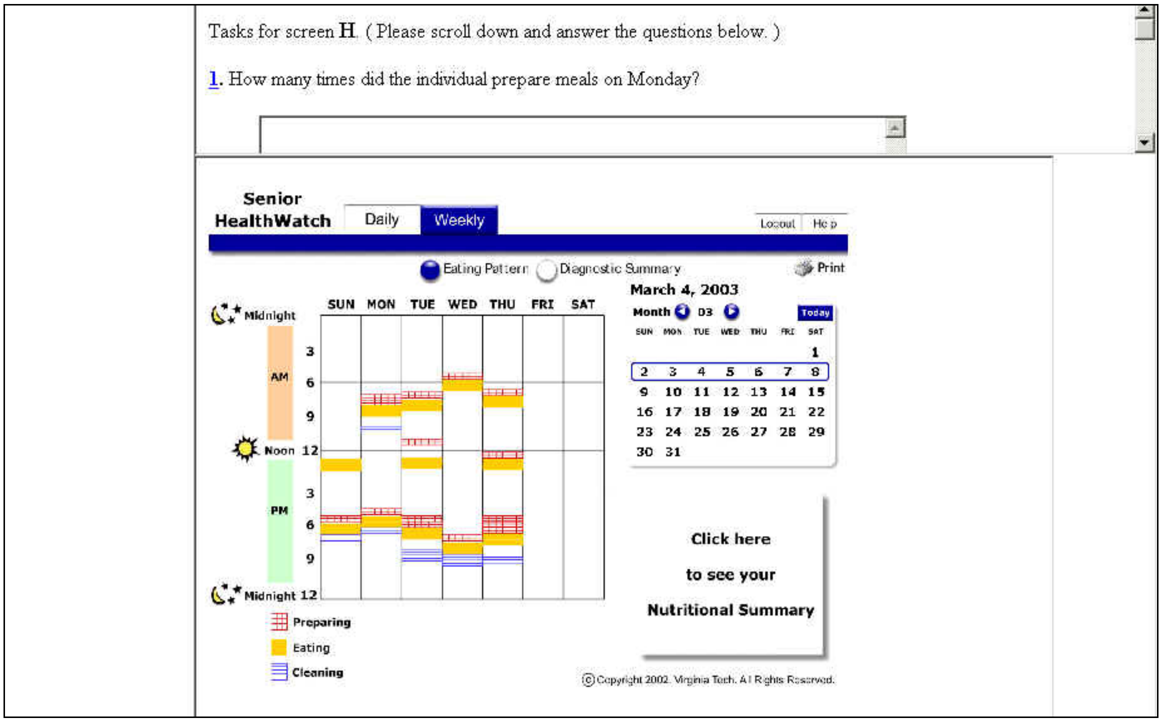
CONCRETE DISPLAY FOR HIGH DENSITY SCREENS



TABULAR DISPLAY FOR HIGH DENSITY SCREENS



GRAPHICAL DISPLAY FOR HIGH DENSITY SCREENS



APPENDIX H: TASK QUESTIONNAIRES

TASK QUESTIONNAIRE FOR ABSTRACT DISPLAY

Tasks for the screen B.(Please scroll down and answer the questions below.)

1. What do you think the small blocks represent?

2. Do you think this individual can consume more sweets?

Yes No

3. How is the individual's intake with respect to milk/cheese group?

4. How is the individual's intake with respect to fruits?

Please also report your comments using the Remote Evaluation tool which is below on your taskbar before you click the submit survey button.

Submit Survey

TASK QUESTIONNAIRE FOR CONCRETE DISPLAY

Tasks for screen D. (Please scroll down and answer the questions below)

1. What do the face icons represent according to you?

2. Do you think this individual can consume more sweets?

☐ Yes ☐ No

3. How is the individual's intake with respect to milk/cheese group?

4. How is the individual's intake with respect to fruits?

Please also report your comments using the Remote Evaluation tool which is below on your taskbar before you click the submit survey button.

Submit Survey

TASK QUESTIONNAIRE FOR TABULAR DISPLAY

Tasks for screen U. (Please scroll down and answer the questions below)

1. How many more servings of fat/sweets can the individual have today?

2. How many servings of bread/rice can the individual have today?

3. How many servings of Milk/Cheese can the individual have in a week?

4. Is the individual in a good condition in terms of Fruits diet?

Please report your comments using the Remote Evaluation tool which is below on your taskbar before you click the submit survey button.

Submit Survey

TASK QUESTIONNAIRE FOR GRAPHICAL DISPLAY

Tasks for screen H. (Please scroll down and answer the questions below.)

1. How many times did the individual prepare meals on Monday?

2. From when to when did the individual have their first meal on Tuesday?

3. How long did the individual spend on preparing a meal on Thursday?

4. How many times have the individual missed a meal during the week?

Please also report your comments using the Remote Evaluation tool which is below on your taskbar before you click the submit survey button.

Submit Survey

APPENDIX I: EXPERT COMMENTS AND DESIGN RECOMMENDATIONS

Abstract	Concrete	Tabular	Graphical
User Comment: Stoplight formatting is very effective.	**User Comment**: Faces are very cheerful.	**User Comment**: It is tough to get what the colors mean, since in this case there is no red, only yellow and green squares. **Recommendation:** Provide legends below the table explaining the colored squares	**User Comment**: Tough to understand the screen. **Recommendation:** Provide legends below the graph explaining the colored grids
User Comment: Legends will be more helpful. **Recommendation:** Provide legends below the table explaining the colored blocks	**User Comment**: Adding colors to faces is a good idea with redundant coding.	**User Comment**: Very easy to answer questions.	**User Comment**: Its tough to make out the actual time on the graph without any demarcation. **Recommendation:** Show grids on graph to show time
User Comment: More place to answer task questions needed. **Recommendation:** Provide a bigger area on the screen for tasks	**User Comment**: Explanation of color-coding required. **Recommendation:** Provide legends below the table explaining the colored faces	**User Comment**: Easy to understand.	**User Comment**: Very tough to view on high-density screen. **Recommendation:** Use low density screen, or reduce amount of information displayed when it is high density screen
User Comment: A column header for the boxes is needed.	**User Comment**: Also telling the user what the faces stand for is required.		**User Comment**: Wording is not clear, with the starting time being

Recommendation: Give clear column headers for the column with colored blocks	**Recommendation:** Provide legends below the table explaining the colored faces		midnight on the graph. **Recommendation:** Change the wording on the graph from "Midnight" to "Early Morning"
User Comment: Fonts need to be bigger on high-density screen. **Recommendation:** Make font 2 sizes bigger	**User Comment**: Font size is too small in high-density screen. **Recommendation:** Make font 2 sizes bigger		**User Comment**: Colors are nice and pleasing.
User Comment: Too much clutter on high-density screen. **Recommendation:** Use low density screen, or reduce amount of information displayed when it is high density screen	**User Comment**: Faces don't give any specific data.		**User Comment**: Extra information on the screen which is not being tested.

VITA

Sajitha Narayan was born in Kerala, India and raised in India and Kuwait. In 2000, she received her Bachelor of Technology from Indian Institute of Technology, Madras in Chemical Engineering. In 2005, she completed her Master of Science in Industrial and Systems Engineering at Virginia Polytechnic Institute and State University. Through the Human Factors Engineering option, she pursued research in the area of Human-Computer Interaction while working in conjunction with the Computer Science Department. While Obtaining her Master's degree, she also spent 1 year as a Graduate Research Assistant. Sajitha also pursued usability internships at Kyocera Wireless Corporation, VTTI, and Oracle Corp. Sajitha is a member of the Human Factors and Ergonomics Society and Bay Chi. At present Sajitha is working as a Usability Engineer with the Applications UI group at Oracle.