Mobile Interface Design for Low-Literacy Populations

Beenish M. Chaudry
Kay H. Connelly
Department of Computer Science
Indiana University
Bloomington, IN
1-413-461-6440
bchaudry, connelly
@cs.indiana.edu

Katie A. Siek
Department of Computer Science
University of Colorado
Boulder, CO
1-303-492-5066
ksiek@cs.colorado.edu

Janet L. Welch
Department of Adult Health
Indiana University
& Purdue University
Indianapolis, IN
1-317-274-8091
jwelch@iupui.edu

ABSTRACT
Despite the proliferation of mobile health applications, few target low-literacy users. This is a matter of concern because 43% of the United States population is functionally illiterate. To empower everyone to be a full participant in the evolving health system and prevent further disparities, we must understand the design needs of low-literacy populations. In this paper, we present two complementary studies of four graphical user interface (GUI) widgets and three different cross-page navigation styles in mobile applications with a low-literacy, chronically-ill population. Participant's navigation and interaction styles were documented while they performed search tasks using high fidelity prototypes running on a mobile device. Results indicate that participants could use any non-text based GUI widgets but preferred and performed efficiently with large radio buttons. For navigation structures, users performed best when navigating a linear structure, but preferred the features of cross-linked navigation. Based on these findings, we provide some recommendations for designing accessible mobile applications for low-literacy populations.

Categories and Subject Descriptors
H.5.2 [User Interfaces]: Graphical User Interfaces (GUI), User-centered Design

General Terms
Design, Human Factors.

Keywords
Low-literacy, interface design, mobile, widget, navigation style

1. INTRODUCTION
Researchers have found that individuals with low-literacy skills are at the risk of poor health outcomes. This is because they misunderstand written medical information and instructions [1]. This is a matter of great concern because about 43% of the US population has been found to have inadequate literacy skills needed to search and comprehend written information [2]. Furthermore, low- education and literacy are linked to other disadvantages such as low-socioeconomic status and unemployment [3]. These factors also affect the quality of health care that can be afforded and received by low-literacy people.

Given the large segment of the population with low-literacy skills and the recent vision to have patients fully involved in the collection and reflection of their health data to improve health outcomes, there is an increased awareness on developing health applications for low-literacy populations [4,5,6]. Mobile devices are poised for this purpose because research has shown that people with low-literacy skills find it easier to learn mobile devices rather than desktop computers [7]. Moreover, two recent Pew studies [29, 30] have shown that smartphone ownership and internet access through mobile phones is rapidly increasing in low-income American adults. These individuals are not only comfortable with mobile devices but are also mostly relying on them rather than on desktops for internet access.

To date, however, a review of the five most popular mobile nutrition-monitoring applications available through Apple's App Store[1] shows they rely on high- literacy and numeracy skills. Indeed, analyzing 25 screens from these applications using SMOG [8] shows they require an average reading level of over the 10th grade, with only three screens below the 9th grade reading level. This effectively cuts out an entire segment of the population. Guidelines informing how to design interfaces for low-literacy users remain incomplete. What graphical user interface components can easily be interpreted and used by low-literacy populations? What kind of cross-page navigation style or structure (i.e., connections between different pages or screens of a mobile application) results in the least amount of errors? What complexity of the navigation tree (depth and breadth) is appropriate for this population? Answering these questions is crucial to ensure everyone, independent of literacy levels, can use technology for his/her benefit.

In this paper, we describe two user studies that were conducted on a chronically-ill, varying-literacy population with an overall low-literacy and low-technical familiarity. The design recommendations reflect the needs of low-literacy users because being inclusive of the entire population means designing for the lowest literacy levels. In the first study, we tested four different GUI widgets of varying sizes. The purpose was to investigate which widget type and size is optimal for and usable by our target population. In the second study, we tested three different cross-page navigation styles each with four tasks of different complexities. The second study aimed at finding the navigation style that was least prone to and easiest to recover from errors, and the navigation tree that had appropriate complexity for our target users.

IHI'12, January 28–30, 2012, Miami, Florida, USA.

[1] http://www.iphoneappsdir.com/top/health-and-fitness/ accessed on June 27, 2011.

Our results show that a low-literacy population was able to use all graphical widgets but large radio buttons were most efficient and preferred. They were also able to successfully navigate with all three navigation styles. However, with linear navigation and with navigation depth and breadth of five they made the least errors and recovered faster after an error.

2. RELATED WORK

We provide a brief overview of existing design guidelines for low-literacy users, identifying the gaps in the literature. In particular, few studies perform direct comparisons between different GUI components and navigation styles, but instead report on the successes or failures of particular interface features embedded in a larger application. In addition, the majority of the literature addresses low-literacy populations in developing countries, in which culture, infrastructure and education all differ greatly from first world countries and can impact results.

2.1 User Interfaces for Low-Literacy Users

There have been a number of approaches to the design of user interfaces including graphics, icons, speech and video for low-literacy populations.

M. Huenerfauth proffered design recommendations mostly involving speech for non-literate users of a proposed PDA-like device based on a literature review [9]. Sherwani et al designed, developed, and evaluated speech-based health information access interfaces for low-literacy community health workers in Pakistan [10]. However, they and Siek et al. [11] found that a low-literacy population needed detailed instruction from a human facilitator on what to say while using the speech-based interfaces. Furthermore, Siek et al. found that a United States based low-literacy population was not able to use a phone-based integrated voice recognition menu system to navigate a voice-based application [11].

Grisedale et al. worked on a PDA-based interface designed for rural community health workers in India [12]. This work was concerned with designing usable interfaces for recording health-related information by a low-literacy population. Their interface was based on a graphical language that relied on some text, specifically designed for the target users.

Medhi et al. developed a text-free interface that performed well in a usability test with a low-literacy population in India [13]. They emphasized the need for semi-abstracted graphics with voice feedback and a help feature on every screen.

There have been no studies comparing the usability of different non-text based widgets (such as radio buttons, check boxes and icons) with low-literacy populations.

2.2 Navigation Structures for Low-Literacy Users

Walton et al. examined the extent to which novice and experienced web users in South Africa were able to reproduce classificational taxonomies or tree structures of various websites [14]. They found that the cultural dimensions influence the interpretation of these structures. This further encouraged us to explore design guidelines for low-literacy users in developed countries.

Kodogoda et. al. combined open-card sorting technique with cognitive task analysis to find reasons to explain the premature abandonment of online searches by low-literacy users [15]. Their comparative study between low and high literacy users residing in UK found that low-literacy users lose focus while navigating a multi-leveled taxonomy. Similarly, Doe et al., through evaluation of a digital library by low-literacy immigrants from developing countries residing in New Zealand, found that hierarchies in information architecture should be avoided. Browsing multiple depths of information (e.g., as in a Web page, or when using navigation metaphors of “up a level” is difficult for low-literacy users [16].

Parikh et al. described the evolution of a successful design of a user interface for managing community-based financial institution in rural India. The target users were low-literacy village women. The researchers listed several important features that contributed to the success of their interface including recommending a linear navigation structure which guides the user from task to task through a sequence of consistent screens [17].

We extend these findings in two ways: by comparing different navigation information hierarchies and targeting a low-literacy population in a developed country.

3 CONTEXT OF STUDIES

The two studies reported in this paper are part of a larger project looking at how to assist a low-literacy, inner-city, chronically-ill patient population monitor their health. In particular, we are working with patients who have Chronic Kidney Disease (CKD) – stage 5, which means their bodies cannot remove excess fluids and toxins on their own and they must undergo hemodialysis to survive. Since they do not receive “kidney” function except for three times a week, they have an extremely restrictive prescribed diet that most are unable to adhere to, even when they try [18]. A mobile fluid and nutrition monitoring application is an ideal mechanism to help these patients better understand their consumption patterns and make dietary changes. However, currently there is no application suitable for this population, which has the added constraint of a high portion of users with low-literacy skills.

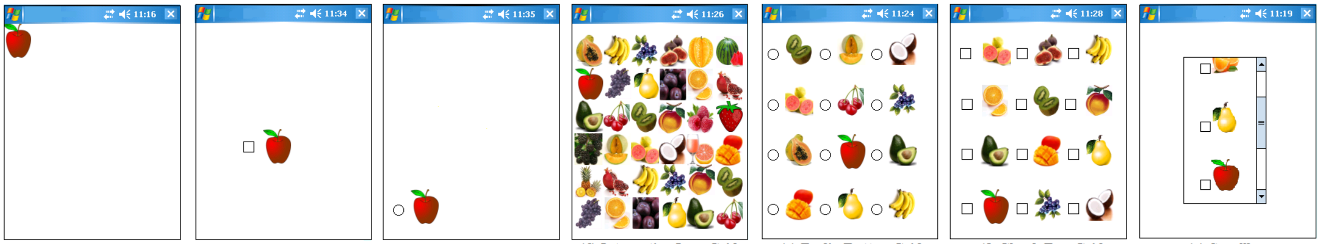
To design such an application, we must first better understand what non-text based graphical interface components and what type of navigation style this population is able to use.

We used our target CKD patient population in the current studies; thus, the studies were conducted in a dialysis unit during the first two hours of dialysis to accommodate participants' schedules. We were not allowed to use video or audio recording as per the dialysis unit rules. In both studies, we recruited participants who had a reading level below the 9th grade, as measured by the Rapid Estimate of Adult Literacy in Medicine (REALM) test [19]. They were also required to have no vision problems and also not experiencing any problems viewing the mobile device screen used in the study. Moreover, only alert and oriented patients were selected, scoring greater than or equal to 25 (out of 30) in the mini-mental state exam [20].

4 STUDY 1: GUI COMPONENTS

The aim of the first study was to evaluate the usability of several different non-text based graphical widgets by a low literacy population. Specifically, our research questions were:

- With which widget type and size do low-literacy users perform most efficiently?
- Which widget type and size is preferred by low literacy users and why?



(a) Interactive Icon (b) Check Box (c) Radio Button (d) Interactive Icon Grid (e) Radio Button Grid (f) Check Box Grid (g) Scrollbar

Figure 1: Participants played a game by selecting the apple on each screen for the Widget Study (all widgets shown in small sizes)

- Are low-literacy users able to find and select what they are looking for when presented with several choices?

4.1 Methods

We obtained approval for the widget study from the Institutional Review Board (IRB). The study was conducted in a dialysis unit during the first two hours of dialysis to accommodate participants' schedules and comfort. The participants voluntarily participated and were compensated with $5. We only recruited participants who had low-literacy skills according to REALM. The first part of the study consisted of task-based interactions with the prototype followed by a multiple choice survey with the participants. The entire study lasted for 15-20 minutes.

4.2 Prototype

A high-fidelity mobile prototype game was developed for this study. To keep in line with the nutrition monitoring application we were developing, the food-themed game required participants to select an apple on each of its 21 screens. Each screen had only one type of widget on it: an interactive icon (Figure 1.a and 1.d), a check box (Figure 1.b, 1.f and 1.g), a radio button (Figure 1.c and 1.e), and a scrollbar (Figure 1.g). These widgets were chosen because they are frequently used in interfaces that allow people to make selections when presented with multiple options [21] and could be adapted for non-text applications. Selecting the target item with these widgets involved: clicking directly on its interactive icon, or clicking inside the check box or the radio button associated with it.

Each widget was tested in three different sizes (shown in Table 1). The size selection was based on the empirical observation of how much area each one covers with respect to the rest of the screen. These sizes also lie within the range of icon size preferred and/or viewable by this population [22]. All widgets were tested for their usability both as singles and multiples. The purpose of testing widgets as singles was to see if participants could use them, and as multiples to observe if an item could be picked from many choices. Three different lengths of scrollbars were incorporated in the list view to assess which length the population performed best with. The apple was placed at a different location in each list view. Scrollbars were not used on any other screens.

The prototype was programmed to randomize the screen (hence the widget) order every time a new instance of the game started. This was done to make sure that we received an unbiased assessment of participant's performance for each widget.

The prototype ran on a mobile device having the form factor of a smartphone, and a touch screen. The screen resolution was 240 X 320 pixels and measured 3.5 inches diagonally. The screen's contrast and brightness were set at default values during the study.

Table 1 Sizes of Widgets

Widget	Large (L)	Medium (M)	Small (S)
Interactive Icon	28mmx35mm	21mmx25mm	13mmx15mm
Check Box	18mmx18mm	12mmx12mm	5mmx5mm
Radio Button	18mm	12mm	5mm
Scrollbar	190mm	114mm	38mm
Number of Widgets: width-wise x length-wise			
Interactive Icon Grid	3 x 3	4 x 4	6 x 6
Check Box Grid	1 x 2	2 x 3	3 x 4
Radio Button Grid	1 x 2	2 x 3	3 x 4

4.3 Participants

We recruited a total of seventeen participants consisting of eight women and nine men. Sixteen were Black/African American and the remaining one was Latino/Hispanic. The average age of the participants was 54.4 years (S.D. = 12.7 years). While they had between 9 to 12 years of education, they all read below 9^{th} grade as measured by REALM. Specifically, six participants read at 7th to 8th grade, three at 4^{th} to 6^{th} grade, and eight at or below 3rd grade. The mean MMS score was 27.8 with the highest being 30 and the lowest 26.5.

In terms of level of familiarity with the computers, ten had never used a computer before. The rest had some exposure but were not regular users: they reported using computers once a month or less to play games or browse the Internet. These participants either owned a computer or had used one at the public library or old job.

4.4 Study Procedure & Tasks

After obtaining informed consent from the participants, we briefed them on the purpose and procedure of the experiment. We then demonstrated the prototype so participants could become familiar with it. The study task was to play the game three times in a row. The widget order was randomized every time a new instance of the game started. During the game play, the mobile device automatically recorded the total number of clicks made on each screen and the total time spent with each widget. At the end of the study, participants told us their preferred widget type and size.

4.5 Results

Our main findings were that most participants not only performed the best with radio buttons, but also preferred them over the other widgets. They could find and select target items with any widget when presented with several options. Large sizes of all widgets were not only more efficient than small sizes but also preferred by most participants. In terms of inner page navigation, scrollbars were successfully used. Here we provide a detailed analysis of the data leading us to these results.

Table 2 Performance Metrics Analysis

Widget	Unsuccessful Clicks (UC)			Interaction Time (IT) in seconds		
	Mean	S.D	Median	Mean	S.D	Median
Check Box (S)	0.89	0.88	1.00	5.06	4.83	2.67
Check Box (M)	0.51	0.55	0.33	2.61	3.06	**1.67**
Check Box (L)	0.42	0.57	0.33	2.40	2.02	**1.67**
Radio Button (S)	0.65	0.65	0.67	2.48	1.35	1.83
Radio Button (M)	0.50	0.46	0.67	2.43	1.50	2.00
Radio Button (L)	**0.29**	0.32	**0.22**	**1.96**	1.44	**1.67**
Icon (S)	0.38	0.40	0.33	2.42	1.07	2.17
Icon (M)	0.31	0.42	0.22	2.60	1.43	2.17
Icon (L)	0.40	0.52	**0.00**	2.75	1.46	2.33
Scrollbar (S)	0.98	0.92	1.00	17.87	10.9	17.33
Scrollbar (M)	0.92	0.92	1.00	16.43	7.72	15.33
Scrollbar (L)	0.96	0.67	0.67	13.36	9.05	10.33
Icon Grid (S)	0.45	0.47	0.33	4.31	3.20	3.33
Icon Grid (M)	0.35	0.43	0.33	3.24	1.88	3.00
Icon Grid (L)	0.31	0.35	0.33	3.08	1.73	2.67
Check Box Grid (S)	0.30	0.35	0.17	4.83	3.07	4.33
Check Box Grid (M)	0.56	0.57	0.50	6.30	6.49	3.33
Check Box Grid (L)	0.31	0.22	0.00	2.30	1.16	2.17
Radio Button Grid (S)	0.57	0.45	0.33	3.46	1.21	3.17
Radio Button Grid(M)	0.37	0.35	0.33	2.53	1.33	2.67
Radio Button Grid (L)	**0.30**	0.21	**0.00**	**2.07**	1.29	**2.00**

4.5.1 Widget Performance

Participants' performance with each widget was gauged by: interaction time and number of unsuccessful clicks. Table 2 shows the mean, standard deviation and median of these metrics (averaged for all trials) across all participants. The columns with the white background indicate results for single widget screens and the gray columns are for multiple widget screens. We analyze them separately to account for any differences that could have affected performance in these different situations.

We ignored outlier values of these metrics in calculating these statistics. A metric value was considered to be an outlier if it was more than 10 units away from the majority of values. Figure 2 shows interaction time scatterplots of three different widgets with three outlier points. Such outliers were expected because the participants could be distracted during the study, for example due to a medical attendant checking the dialysis machine to which the participant was connected.

Among single widget screens, large radio button had the smallest median and mean interaction time showing that most participants were fastest with it. This was followed by medium and large check boxes. However, large icon, followed by large radio buttons, had the highest number of participants with least unsuccessful clicks.

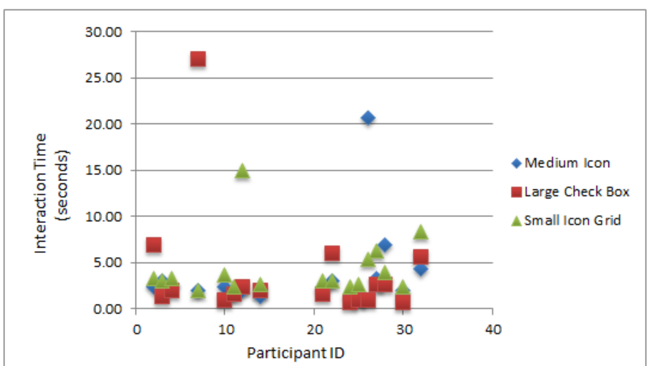


Figure 2 Interaction Time Scatterplots

Table 3 Number of Preferences for the Widgets

Widget	Large (L)	Medium (M)	Small (S)
Interactive Icon	-	2	-
Check Box	3	-	-
Radio Button	-	8	-
All of them	4		

For multiple selection screens, the large radio button grid was fastest and had the fewest errors. This was followed by the large checkbox grid.

A linear mixed model repeated measures analysis on interaction time was performed with widget type and size as independent variables and participant as a repeated effect across widgets. The diagnostic statistics suggested that the residuals were approximately normally distributed. The scrollbar was found to have taken significantly longer than the other widgets but no significant difference was determined between the interaction times of the other widgets ($p<0.001$). There was a marginal difference between the three sizes ($p=0.007$). The analysis was re-done without the scrollbar. This time, a significant difference was found between the interaction times of different widgets ($p=0.001$). The radio button grid ($p=0.001$) and icon grid ($p=0.013$) were significantly faster than the check box grid. There was a significant difference between the sizes – small sizes took significantly longer than large sizes ($p=0.002$).

In terms of number of clicks, there was a significant difference between widget types ($p=0.021$) with scrollbars excluded. The check box grid had significantly more unsuccessful clicks as compared to the icon grid ($p=0.013$), and the radio button grid had less unsuccessful clicks than the check box grid (Bonferroni $p=0.028$). There was no significant difference between the three sizes ($p=0.159$) in terms of the number of unsuccessful clicks.

4.5.2 Widget Preference

At the end of the experiment, we asked the participants to vote for their preferred widget type and size. We counted the number of votes for each option. The results are presented in Table 3. Participants preferred the large and medium sizes of the widgets in general. Eight participants thought that the medium radio button was the ideal widget. We are somewhat surprised that most participants preferred radio buttons, because before learning to use radio buttons and check boxes, they were directly clicking the apple icons associated with them. Even though participants performed moderately well with the smaller widgets, no participants preferred them. This is consistent with previous results on icon sizes for chronically-ill patients [22].

4.5.4 Scrolling

We evaluated participants' ability to use scrolling input by making them scroll through a list of items to select an apple (Figure 1.g). The participants took longer to locate the target item with all three lengths of scrollbars because it involved searching through multiple screens of icons. We thought that interaction time would be directly proportional to the length of the scrollbar but searching for the target was fastest with the longest scrollbar. We believe this happened because the target was located at the bottom of the list that could be reached easily by keeping the arrow pressed for some time. Participants took longer to locate the target item with small and medium scrollbars. In these cases, a rather skillful manipulation of scrollbar was required to locate the target item because it was in the middle of the list. Nevertheless, the participants were successfully able to manipulate all three scrollbar lengths to locate the target item.

4.5.5 Summary of Findings

Overall, participants could use any of the tested widgets, but performed fastest and made least errors with radio buttons. The large sizes of all widgets were significantly faster than the small sizes. Participants also preferred large and medium sized widgets, and most preferred to use radio buttons. They were also successful in locating the target item in a multitude of options. Finally, scrollbars increased task completion times but low-literacy users manipulated them successfully.

5. STUDY 2: IDEAL NAVIGATION STRUCTURE

After identifying the optimal widget for our low-literacy population, we explored which navigation structure was suitable for them. Specifically, we investigated:

- What kind of cross-page navigation style or structure results in the least number of errors and fastest error recoveries?
- What complexity of the navigation tree (depth and breadth) is appropriate for this population?
- What usability issues surround these navigation styles?

5.1 Methods

We obtained approval for the navigation study from the IRB. The participants voluntarily participated and were compensated with a $10 gift card. We administered REALM test to recruit only low-literacy participants. The study required participants to perform task-based interactions with three high-fidelity prototypes, each based on a different navigation structure, followed by answering a multiple choice survey. The study, including the time to train participants, lasted for 1 – 1.5 hours on average. Participants were encouraged to proceed slowly and take breaks during the study to minimize exhaustion and tiredness towards the end. The prototypes automatically logged all the interactions.

5.2 Prototypes

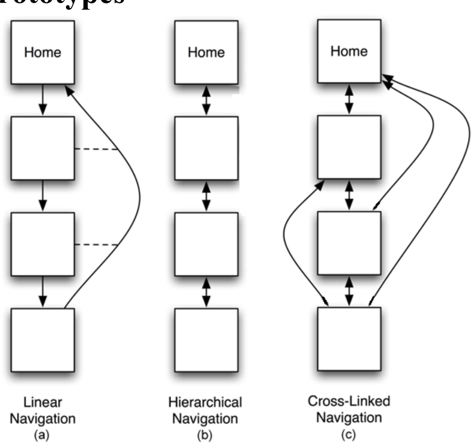


Figure 3 Box and arrow diagrams for three navigation styles

Three different prototypes were used in this study. The underlying application of each prototype was the same. It enabled the user to select food items from specific food categories. (These food categories were based on the way low-literacy people think about organizing food [22]). The prototypes differed in the way various screens of the application were interconnected. The interconnections were based on linear, hierarchical or cross-linked modes of navigation (Figure 3).

Linear navigation (Figure 3.a) resembles the navigation used in a multimedia slide show or a photo gallery. The user follows a particular sequence of steps determined by the designer of the application. It is possible to return to the beginning (or main page of the application) from any step in the sequence. At the end of the sequence, the user is automatically taken back to the beginning.

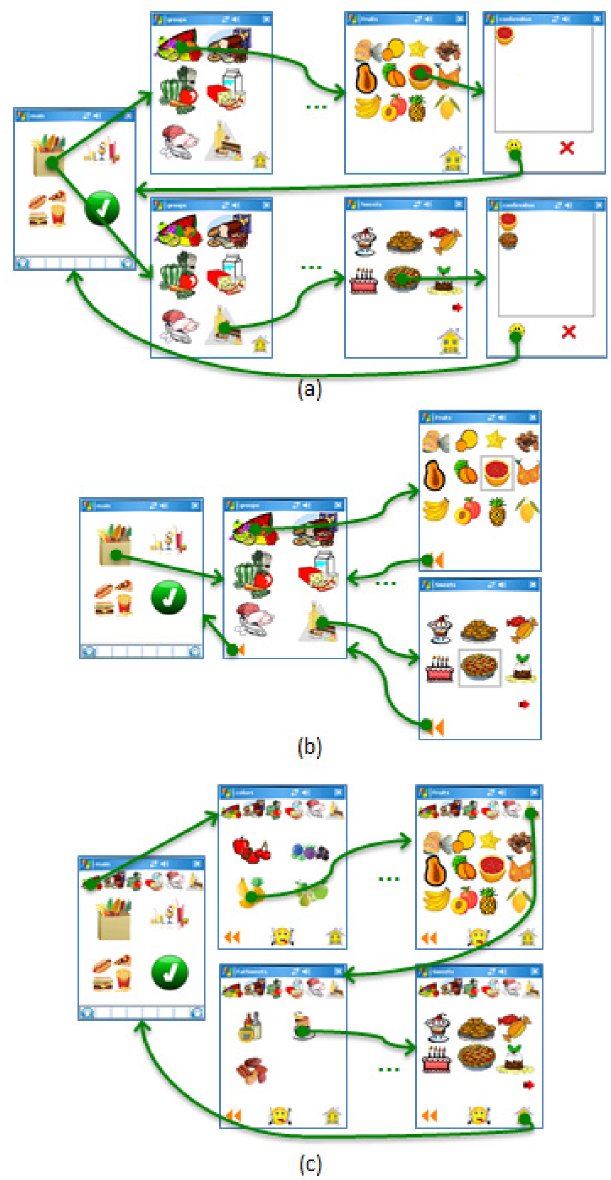


Figure 4 Screen shots from three prototypes to select a grapefruit then a pie: (a) linear (b) hierarchical – back arrows (c) cross-linked – navigation tab. "…" represents missing screens in the sequence

Hierarchical navigation (Figure 3.b) resembles decision support systems that are based on prior answers. It allows the user to go down different paths to finish subtasks from the main page, follow the same steps backwards to get back to a previous screen and then down another path for a different subtask. It typically incorporates back arrows to return to a previous screen. It does not bring the user back to the start on (sub)task completion.

Cross-linked navigation style is often found in websites where information has an organizational structure. At every step of a task, this structure (Figure 3.c) provides access to the starting steps of all the tasks supported by the application while combining the features of linear and hierarchical navigations. Tabs or a navigation bar with links typically gives access to the starting step of all the tasks. Users are not returned to the beginning of the application on a (sub)task completion.

In all three prototypes, two features assist with the task of meal creation: (1) a meal review panel that is present at the bottom of the main screen (first screen in Figure 4.a-c) and (2) a meal review page that opens when the meal review panel is clicked (last screens in Figure 4.a). On meal review screen, participants can select items and then click on the cross to delete them or just click on the smiley to save them and return to the last page. In linear navigation, the meal review page automatically opens at the end of a task completion (selection of a target food item). In cross-linked, it can be opened by clicking the bottom-center icon of every screen (Figure 4.c).

In the linear prototype (Figure 4.a), the user chooses a food category on subsequent pages until they reach the target item page. Upon selecting the item, the user returns to the main page. The HOME icon returns the user to the main page from any page in the sequence if they make an error. The hierarchical prototype (Figure 4.b) was based on the same idea as linear, but had a BACK button instead of the HOME. The BACK icon helped a user step backwards through the sequence of visited pages. So if a user made a mistake, she could step backwards to ultimately reach the page on which the incorrect navigation decision was made. This navigation also did not need to return the user to the main page at the end of a task completion. The cross-linked prototype (Figure 4.c) had a link to each food category page (called a navigation bar) at the top of every page, in addition to the HOME and BACK icons at the bottom. In case of a mistake, the user could manipulate any of the top or bottom navigation icons to navigate to the correct page.

All prototypes were deployed on a mobile device with a touch screen and a smartphone form factor. The 3.5 inch screen had a resolution of 240 X 320 pixels. The contrast and brightness of the screen were set at default values for the study.

5.3 Definitions & Hypothesis

We define *traversed path length* as the number of interface interactions (or clicks) needed to move from one target page to another in the application. Thus, a traversed path length associated with the task of selecting a food item depends on the starting screen, the navigation style and any errors. An *error* is an interaction that takes the participant to a screen that is not on a direct path to the target. The number of interactions made after an error and before coming back to any point in the direct path is the *recovery length*. A path to the target item which is free from errors is a *correct path*. The *shortest path* is the path to the target item requiring the minimum number of interactions. In cross-linked navigation, not all correct paths are the shortest path because a user could correctly use the navigation tab or back arrows to find the next target, but one may be shorter than the other. Finally, the *breadth* of navigation is the number of choices offered on a screen at any point during the navigation.

In our linear and cross-linked navigation prototypes, the *maximum shortest path length* or *depth* was 5 and 6 respectively. While for hierarchical navigation, the maximum shortest path length was 8 because the participant was forced to go back to the main page step-by-step using the BACK button. Previous work has shown that as path length increases, the task completion time and the number of errors made in completing a task also increases [23]. This led us to hypothesize that *linear navigation is the most appropriate navigation structure for this population*.

5.4 Participants

We recruited nineteen participants, ten of whom were female. All except one participant were Black/ African Americans, the remaining one was White. The average age was 49.75 years (S.D. = 13.6 years) and years of education ranged from 10 – 14 (average = 11.75 years). We only recruited people with low-literacy skills (reading level at or below 8th grade) according to REALM test. Specifically, we had eight participants reading at 7th to 8th grade, three at 4th to 6th grade and the rest at or below 3rd grade. The mean MMS score was 28.9. The highest score was 30 and the lowest 26.5.

There were three different categories of computer familiarity within the participants. Seven had never used a computer. Eight identified themselves as beginners – they used a computer at most once a month to play games or surf the web. Five claimed to be daily users. They regularly used a computer to browse the internet, chat and play games. These participants either owned a computer at home or had used one at the public library, church or an old job.

5.5 Study Procedure & Tasks

After obtaining consent, a researcher explained how each navigation structure worked to the participant with the help of the diagrams shown in Figure 3. Participants were familiarized with specific features such as the BACK and HOME icons. Before performing study tasks, participants were encouraged to engage in playful experimentation with the prototypes so they could get familiar with them. They were assigned random tasks until their performances plateaued. We observed an improvement in performance during the training session but we do not report on the learning in this paper. We also noticed that increased computer familiarity did not offer any significant advantage to the participants – all participants needed training and some experimental time with the prototype to develop some proficiency in using it.

The study tasks consisted of building a meal by identifying and entering four food items present in a randomly assigned picture. Entering one food item was equivalent to one task. Each participant had to perform four tasks (i.e. enter all food items in one picture) with each navigation style. Participants received a different picture for each navigation style. We tried our best to present food items with different path lengths in each picture, so we could analyze performance with various path lengths. However because of the order in which participants chose to enter the food items, some participants had more than one task of same path length. Therefore, we will try to analyze how each type of task was completed across all participants. The order in which the navigation styles were used was randomized to offset any advantages resulting from using the prototypes in a particular order.

Participants had to complete all the tasks on their own. Help was only provided when they were stuck or confused. After finishing the experiment, participants commented on their experiences and level of satisfaction with each navigation style. They also answered a multiple choice survey about their preferred navigation style and explained their preferences.

Table 5 Averages of Navigation Metrics for Different Path Lengths

Shortest Path Length	Average Traversed Path Length			Average Number of Errors			Average Recovery Length		
	Linear	Hierarchical	Cross-linked	Linear	Hierarchical	Cross-Linked	Linear	Hierarchical	Cross-Linked
2	2.28(n=18)	2.90(n=10)	2.55(n=11)	0.17	0.50	0.27	0.06	0.6	0.27
3	3.62(n=13)	3.75(n=8)	4.15(n=23)	0.38	0.38	0.43	0.54	0.25	0.65
4	4.56(n=17)	4.8(n=10)	4.84(n=19)	0.12	0.3	0.42	0.18	0.6	0.79
5	5.74(n=19)	6(n=9)	5.5(n=5)	0.26	0.67	0.8	0.53	0.78	0.8
6		7.11(n=9)	10.8(n=6)		0.11	2.67		0.11	2.83
7		8.33(n=6)			0.67			0.83	
8		9(n=3)			0.33			0.33	

5.6 Results

Our main findings are that a low-literacy population performed best when navigating a linear structure. We attribute the better performance with participants' preference to always starting a new task at the same location and having the ability to start over whenever they wanted. Participants had difficulty navigating structures where they could go back and forth between subsequent screens, but they liked the ability to jump to any screen as provided by the cross-linked navigation. Following is the detailed analysis that led to these results.

5.6.1 Task Percentages

Overall, the task completion rate was high for all navigation styles. With linear navigation, participants not only had highest percentage of completed tasks but also had the highest percentage of tasks completed without any errors. This was followed by cross-linked and hierarchical navigations. Table 4 shows the percentage breakdown of the tasks completed and incomplete tasks. We attempt to understand these differences in more detail in the following sections.

Table 4 Percentages of Completed and Incomplete Tasks

Navigation Style	Completed Tasks		Incomplete Tasks
	Without Errors	With Errors	
Linear	69.7%	30.3%	0%
Hierarchical	46.1%	43.4%	10.5%
Cross-linked	59.2%	38.2%	2.6%

5.6.2 Traversed Paths, Errors and Recoveries

The metrics used for gauging performance with a navigation structure were: traversed path length, number of deviations and recovery length. In Table 5, we summarize the averages of our metrics for each shortest path length in all navigation structures. We report *n* which is the number of instances of the traversed path that was used to calculate the averages. (To report average metrics that are representative of our participants, we did not consider outlier path lengths in our calculations. A path length was considered to be an outlier if it was 5 or more clicks off). In the later sections, we will analyze the problems experienced by the outlier group.

From Table 5, we can deduce that independent of the shortest path length, participants were more likely to achieve the shortest path with the linear navigation. For example, average traversed path length for the shortest path length of 2 was 2.28 in linear but 2.90 and 2.55 in hierarchical and cross-linked respectively. This shows that almost one extra click was involved in navigating the same path length in hierarchical navigation.

The average number of errors per task was the smallest with the linear navigation for all shortest path lengths except 3 where hierarchical had the same average number of errors. The average number of errors were highest with cross-linked navigation for all shortest path lengths except for 2.

The average number of recovery clicks was also highest for cross-linked for all path lengths except 2 where hierarchical had highest average recovery clicks. Linear had the smallest recovery clicks for all path lengths except path length=3 where hierarchical had the smallest average.

From our analysis, it appears that for all navigation structures up until shortest path length 5, participants completed tasks with a traversed path length closer to the shortest path length, with fewer errors and also required fewer recovery clicks. Therefore, five appears to be an appropriate navigation depth for this population.

This analysis also supports our hypothesis that linear navigation is the most appropriate navigation structure for this population, but with slight variations in the observations. We discuss these variations next along with a brief examination of the features of each navigation structure to identify the potential impediments for the successful navigation.

5.6.3 HOME Icon

With the HOME icon, the user could return to the main screen from anywhere in the application. This button was present in cross-linked and linear navigations only. In linear navigation, it was used a total of 36 times by 15 participants – each time to recover from an error. In cross-linked, it was used a total of 37 times by 15 participants – 37.8% times for error recovery and 62.2% times for navigation purposes.

When asked about how they used the HOME icon, participants noted that it helped them recover from mistakes easier by providing them with a recognizable point to restart the task gone wrong. Indeed, one participant said, the "HOME button was good because I knew where I was going." Another said, "I liked the HOME button because it lets me correct mistakes easier."

5.6.4 BACK Icon

On analyzing the interaction logs, we found that BACK icon had both advantages and disadvantages. In general, it was responsible for the increased traversed path lengths in hierarchical and cross-linked navigations. Sometimes it was responsible for creating the loss of situational awareness leading to errors such as enter wrong categories or move back and forth between two screens several times. In hierarchical navigation about 12% of the total BACK clicks consisted of such errors and in cross-linked about 5.5%.

On a positive note, BACK icon was helpful in recovering from errors that were only 1-2 clicks away from the correct path. In hierarchical navigation, 14 participants used the BACK icon on average 3 times (across all four tasks) to recover from such errors (amounting to about 23.4% of the total BACK clicks in hierarchical). Similarly, in cross-linked 14 participants used it 2.14

times on average to recover from such errors (totaling to about 33% of the total BACK clicks in cross-linked). The usefulness of the BACK icon for this purpose was also confirmed by two participants who commented that BACK arrows made the application easy to use.

5.6.5 *Cross-linked Navigation Bar*

We counted the total number of times the cross-linked navigation bar was accessed by all the participants to see if they utilized this bar to their advantage. Three participants used it, on average two times, to complete their tasks without making any errors. Eight used it, two times on average, to make both errors (mostly of entering incorrect categories) and successful navigations. Four used it between 9 to 15 times each but made errors more than 60% of the times, and finally, after making successive errors, resorted to the BACK icon for the rest of their interaction with the cross-linked navigation. The remaining five participants did not use the navigation bar at all.

Despite the mixed performance, we had a good response from participants on the usefulness of the cross-linked navigation bar. One participant commented, "I would rather use the small pictures on the top to save me from going HOME every time." Another said, "The tabs were not big enough but they were helpful in giving options." One more said, "It challenged me to think more about what I was doing."

Table 6 Navigation Preferences

Navigation	**Total with Preference** (number of those who performed best with this style)
Linear	4 (3)
Hierarchical	4 (0)
Cross-linked	8 (1)
All of them	2 (0)
None	1 (0)
Total	**19 (4)**

5.6.6 *Review of Entries*

We did not explicitly incorporate the use of the meal review page into the tasks, but provided it to participants in case they added an incorrect item during a task. Participants manually opened the meal review page a total of 14 times with the linear prototype, 39 times with the hierarchical, and 33 times with the cross-linked prototype. The clicks that were used to open meal review page were not counted as errors. While they usually opened this page to remind themselves of what they had already entered before finding the next item, some participants opened it (in hierarchical and cross-linked navigations) when they had navigated to an incorrect screen and were trying to recover, or when they wanted to delete an incorrect entry.

In the interview, several participants told us that they preferred feedback right after entering an item. This most likely explains why they manually opened the meal review screen in the hierarchical and the cross-linked navigation structures before moving to the next task. "I liked getting feedback. Smiley face told me whether I had gotten everything right." The review page also prevented them from making unnecessary entries as another participant stated, "It helps you remember certain parts of the picture that you did not memorize so you don't add it again."

5.6.7 *Preferences*

At the end of the experiment, participants were asked for their preferred navigation style. The preferences have been summarized in Table 6. Surprisingly, eight participants preferred cross-linked while four participants preferred hierarchical and linear each. Two participants who preferred hierarchical actually used it in a linear style; that is, they always went back to the main screen using the BACK button to begin a new task. One participant did not like any of the navigation styles. This participant had never used a computer before. In terms of preference for navigation style, only 4 participants preferred the prototype they performed the best with. Several participants confirmed this by stating they liked the cross-linked style because it was challenging and made them think. This result is consistent with the findings of Siek et al. [24].

5.6.8 *Navigation Breadth*

Cross-linked navigation had highest number of average error per participant (3.26) as opposed to hierarchical (2.2) and linear (1.89). We attribute this to the breadth of cross-linked navigation. The multiple choices offered by this navigation seemed to overwhelm our participants' ability to think critically and hence make effective navigation decisions. We noticed the phenomenon of "distraction" among our participants which meant participants randomly clicked on various icons. This was detrimental to the task completion because several participants ended up stepping through a large number of screens (as high as 17) in order to locate their target items. Therefore, we do not recommend a navigation structure with the breadth of our cross-linked navigation prototype (10) for a low-literacy population. A number somewhere in between the breadth of our linear (5) and our cross-linked prototypes might be suitable because linear had the fewest number of errors.

5.6.9 *Summary of Findings*

Overall low-literacy users were capable of using all three navigation structures provided they received training. However, with linear navigation they completed tasks of various path lengths with the least number of errors. They were also able to recover faster from errors with this navigation. In terms of preferences, cross-linked navigation was the most favored option. The reasons were its navigation bar which provided more options and challenged the users to think. A depth of 5 and a breadth between 5 and 10 seemed to be appropriate for our population.

The meal review page helped participants orient themselves in the application for task completion. The features such as the HOME icon on every screen helped them exit out of wrong/ unfamiliar screens and to "start over". By returning to the HOME screen at the conclusion of a task, participants could smoothly transition to the next task and hence make fewer errors.

The BACK button had disadvantages such as increased traversed path and error recovery lengths. Stepping through previous screens led to the loss of location awareness causing participants to click on random icons and enter wrong screens. However, it was useful in recovering from errors when the recovery only required 1-2 clicks.

6. DISCUSSION

The motivation of these studies was to inform the design of mobile health applications that could empower low-literacy populations to better understand and manage their health. As a first step in designing effective and usable applications for people with low-literacy, we identified the appropriate widgets and navigation structures. We recruited low-literacy people with a varying level of computer familiarity. Here we present design recommendations for low-literacy people based on our findings.

6.1 Design Recommendations

Our studies show that the target low-literacy population can use any GUI widgets and navigation structures on the touch screen interfaces of mobile devices. Our widget study shows that people prefer widgets that are bigger or medium in size. They also preferred and performed well with radio buttons, possibly because of the feedback mechanism built into their designs, and also because the feedback is designed to match with the input (dot for a point click). Based on these findings, we encourage designers to ***design widgets that are bigger in size, and allow the user to visualize interactions with them***.

Although most participants had never used a scrollbar, they not only learnt to use it during the study but also skillfully manipulated it to find the target item. Therefore, we recommend incorporating ***scrollbars in the mobile applications for low-literacy users***.

The navigation structure study shows that participants performed most efficiently with the linear navigation structure, but they preferred the ability to quickly navigate to other pages provided by the cross-linked navigation bar. Thus, for this population, we suggest to ***use a hybrid navigation structure that combines linear navigation with a navigation bar*** - similar to a dock on a Mac operating system or bread crumb trail in webpage design. This combined design idea provides participants the ability to move between screens quickly, but the safety of being able to move sequentially when needed. More work is needed to identify the appropriate hybrid model for this population.

We learnt from the navigation study that it is very important to have error recovery mechanisms built into the interface. Our participants preferred to start over with the HOME icon when they made an error. When the recovery was only a page or two back, they successfully used the BACK icon. However, the difficulty of recovering from back errors on longer paths makes us believe that only BACK buttons are not enough for navigation or error recovery help. Instead, every screen should ***incorporate a BACK button for shorter recovery lengths, and a HOME button for longer recoveries***. This also extends the findings of Medhi et al who suggested incorporating a help feature on every screen for low-literacy users [13].

Our participants liked the ability to ***start every task from the same location*** as provided by the linear structure. We think that this feature can be used to empower and orient low-literacy users to perform specific tasks in more complex applications.

Based on their comments, we know that our participants liked having more choices per page but in reality it often detracted them from their original task causing them to make errors. Moreover, our participants completed and recovered much faster from tasks with smaller path lengths or depths. Based on these findings, we recommend structures with ***navigation tree of depth no more than five and a breadth between five and ten for our target users.***

Finally, designers should ***incorporate a review mechanism*** to help orient users and prevent premature abandonment of tasks. We found that some of our users appreciated having a review available to them whenever they wanted it, (e.g. the meal review page) – thus they knew that they were completing the tasks and were successfully using the application. This suggestion confirms Weiss' design guidelines for handheld devices [25].

These recommendations and considerations to design health interfaces for low-literacy populations come at a time when chronic illnesses, such as CKD, are the leading causes of death and disability in the world [26]. While those with higher socioeconomic status have the resources necessary to educate themselves and monitor their health conditions more rigorously, lower socioeconomic status populations have poorer health and less access to healthcare and resources [27]. Likewise, lower socioeconomic status has been linked to lower literacy levels [2]. Technology can help empower everyone to fully engage in his or her health and participate in the healthcare system. We, as a research and practitioner community, must design for everyone – including low-literacy populations – to ensure that health informatics can reach the ideal goal of decreasing healthcare costs and improving quality of care inside and out of healthcare institutions.

6.2 Future Work

The work presented here is just the beginning of what is required to identify the interface needs of low-literacy populations. Indeed, our results indicate that more research is needed to help low-literacy users establish a sense of location within the application. This challenge encompasses everything from the length of scrollbars (especially in devices with small screens) to how to provide contextual cues about the current location within a complex navigation structure. We intend to extend our current study to examine multiple hybrid structures with contextual cues.

In addition, since more features and assistance provided by an application results in a more complex navigation (in breadth, depth or both), we must identify a good balance between features and application complexity.

Unfortunately, while the sample size of our study was adequate for answering our research questions, it was too small to compare the results for different literacy levels. We would like to recruit more participants in each literacy level to see if how they compare with both widget and navigation performance.

Finally, user-interface research for low-literacy populations may inform and be informed by research being done for other populations with special needs. For example, older adults also have issues with navigation and location awareness in computer applications due to age-associated cognitive decline and other health problems. An article by G. Demiris et al shows that older adults prefer a flattened navigation structure [28]. It would benefit both communities if a population comparison could be performed to determine if results from one translate to the other.

7. ACKNOWLEDGMENTS

This research was supported by a grant made available by the National Institute of Biomedical Imaging and Bioengineering (R21 EB007083), and Indiana University's Faculty Research Support Program. These studies were possible due to the permission granted by University Hospital Indianapolis. We also thank IHI reviewers whose comments helped improve the quality of this submission.

8. REFERENCES

[1] Barry, D. and Weiss, M. 2007. Health Literacy and Patient Safety: Help Patients Understand. American Medical Association Foundation and American Medical Association, 2nd Edition.

[2] http://nces.ed.gov/naal/

[3] http://www.dol.gov/oasam/programs/history/herman/reports/futurework/conference/nalsfina/nalsfina.htm

[4] Medline. http://www.nlm.nih.gov/medlineplus/easytoread/

[5] Gerber, B., et al. 2005. Implementation and Evaluation of a Low-Literacy Diabetes Education Computer Multimedia Application. Diabetes Care 28, 7 (2005), 1574-1580.

[6] Mechael, P. 2006. Exploring Health-related Uses of Mobile Phones: An Egyptian Case Study. Unpublished Doctoral Dissertation. London School of Hygiene, University of London.

[7] Kavanaugh, A., Reese, D., and Pérez-Quiñones, M. 2007. Mobile Phone as Scaffolding Technology: How Low Literacy Groups might Learn Computing. Presented at the 35th Annual Telecom Policy Research Conference 2007.

[8] McLaughlin, G. H. 1969. SMOG Grading — a New Readability Formula. Journal of Reading 12, 8 (May 1969), 639–646.

[9] Huenerfauth, M. 2002. Developing Design Recommendations for Computer Interfaces Accessible to Illiterate Users. Master of Science (MSc) Thesis, Department of Computer Science, National University of Ireland, University College Dublin.

[10] Sherwani, J., Palijo, S., Mirza, S., Ahmed, T., Ali, N. and Rosenfeld, R. 2009. Speech vs. touch-tone: Telephony interfaces for information access by low literate users. In Proceedings of ICTD 2009: Information and Communications Technologies and Development (Doha, Qatar), 447–457.

[11] Siek, K.A., Connelly, K.H., Chaudry, B., Lambert, D., and Welch, J.L. 2009. Evaluation of Two Mobile Nutrition Tracking Applications for Chronically Ill Populations with Low Literacy Skills. In J. Tan and P. Olla, editors, Mobile Health Solutions for Biomedical Applications, IGI Global (2009), 1–23.

[12] Grisedale, S., Graves, M. and Grunsteidl, A. 1997. Designing a Graphical User Interface for Healthcare Workers in Rural India. ACM CHI 1997, 471-478.

[13] Medhi, I., Sagar, A. and Toyama, K 2006. Text-Free User Interfaces for Illiterate and Semi-Literate Users. In Proceedings of Information Technologies and International Development 4, 1 (2006), 37-50.

[14] Walton, M., Vukovic', V. and Marsden, G. 2002. 'Visual Literacy' as Challenge to the Internationalisation of Interfaces: a Study of South African Student Web Users, CHI '02 Extended Abstracts on Human Factors in Computing Systems (2002).

[15] Kodagoda, N., Wong, B.L.W. and Khan, N. 2010. Open-Card Sort to Explain Why Do Low-Literate Users Abandon their Web Searches Early? BCS HCI. Dundee, United Kingdom: British Computer Society (2010).

[16] Deo, S., Nichols, D., Cunningham, S., & Witten, I. 2004. Digital Library Access for Illiterate Users. In Proceedings of 2004 International Research Conference on Innovations in Information Technology (Dubai, UAE), 506-516.

[17] Parikh, T., Ghosh, K. and Chavan, A. 2003. Design studies for a financial management system for micro-credit groups in rural India. ACM Conf. on Universal Usability, ACM Press, New York 2003. 15–22.

[18] Dowell, S., and Welch, J. 2006. Piloting the Use of Electronic Self-Monitoring for Food and Fluid Intake. Neph. Nurs. J. 33 (2006), 271–277.

[19] Davis T.C. et al. 1993. Rapid Estimate of Adult Literacy in Medicine: a Shortened Screening Instrument. Fam Med 25 (1993), 391–395.

[20] Folstein, M.F., Folstein, S.E., and McHugh, P.R. 1975. Mini-Mental State: A Practical Method for Grading the Cognitive State of Patients for the Clinician. Journal of Psychiatric Research 12 (1975), 189-198.

[21] Getting User Choices: Using Check Boxes and Radio Buttons. http://msdn.microsoft.com/en-us/library/.

[22] Siek, K.A. 2006. The Design and Evaluation of an Assistive Application for Dialysis Patients. Ph.D. Thesis, Department of Computer Science, Indiana University, Bloomington.

[23] Parush, A. and Yuviler-Gavish, N. Web Navigation Structures in Cellular Phones: the Depth/Breadth Trade-off Issue. International Journal of Human-Computer Studies 60, (2004), 753-770.

[24] Siek, K.A., Connelly, K.H. and Rogers, Y. 2006. Pride and Prejudice: Learning How Chronically Ill People Think about Food. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2006), 947–950

[25] Weiss, S. 2002. Handheld Usability. John Wiley & Sons Ltd, West Sussex, England.

[26] Strong, K., Mathers, C., Epping-Jordan, J., Beaglehole, R. 2006. Preventing Chronic Disease: a Priority for Global Health. International J. of Epidemiology 35, 2(Apr 2006), 492-494.

[27] N. E. Adler and K. Newman. 2002. Socioeconomic Disparities in Health: Pathways and Policies. Health Affairs 21, 2 (2002), 60–76.

[28] Demiris, G., Finkelstein, S., Speedie, S. 2001. Considerations for the Design of a Web-Based Clinical Monitoring and Educational System for Elderly Patients. J. of the American Medical Informatics Association 8, 5 (2001), 468-472.

[29] Smith, A., Mobile Access 2010, Pew Internet & American Life Project, July 7 2010, http://www.pewinternet.org/Reports/2010/Mobile-Access-2010/Summary-of-Findings.aspx, accessed on Sept. 30, 2011.

[30] Smith, A., Smartphone Adoption and Usage, Pew Internet & American Life Project, July 11 2011, http://www.pewinternet.org/Reports/2011/Smartphones/Summary.aspx, accessed on Sept. 30, 2011.