Usability Testing Finds Problems for Novice Users of Pediatric Portals

Maria T. Britto, MD, MPH, Holly B. Jimison, PhD, Jennifer Knopf Munafo, MA, Jennifer Wissman, MS, Michelle L. Rogers, PhD, William Hersh, MD

Abstract

Objective: Patient portals may improve pediatric chronic disease outcomes, but few have been rigorously evaluated for usability by parents. Using scenario-based testing with think-aloud protocols, we evaluated the usability of portals for parents of children with cystic fibrosis, diabetes or arthritis.

Design: Sixteen parents used a prototype and test data to complete 14 tasks followed by a validated satisfaction questionnaire. Three iterations of the prototype were used.

Measurements: During the usability testing, we measured the time it took participants to complete or give up on each task. Sessions were videotaped and content-analyzed for common themes. Following testing, participants completed the Computer Usability Satisfaction Questionnaire which measured their opinions on the efficiency of the system, its ease of use, and the likability of the system interface. A 7-point Likert scale was used, with seven indicating the highest possible satisfaction.

Results: Mean task completion times ranged from 73 (± 61) seconds to locate a document to 431 (± 286) seconds to graph laboratory results. Tasks such as graphing, location of data, requesting access, and data interpretation were challenging. Satisfaction was greatest for interface pleasantness (5.9 ± 0.7) and likeability (5.8 ± 0.6) and lowest for error messages (2.3 ± 1.2) and clarity of information (4.2 ± 1.4). Overall mean satisfaction scores improved between iteration one and three.

Conclusions: Despite parental involvement and prior heuristic testing, scenario-based testing demonstrated difficulties in navigation, medical language complexity, error recovery, and provider-based organizational schema. While such usability testing can be expensive, the current study demonstrates that it can assist in making healthcare system interfaces for laypersons more user-friendly and potentially more functional for patients and their families.


Introduction

Electronic access to health information, medical records, and health care providers can support new partnerships between patients and providers by promoting self-care, enabling informed decision-making, promoting information exchange and enhancing social support.1 One promising method for providing such access is the personal health record (PHR), defined by the Markle Foundation as an “electronic application through which individuals can access, manage, and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment.”2 3 While PHRs take many forms, the American College of Medical Informatics concluded that PHRs integrated with electronic health records (EHRs) were likely to provide greater benefits than stand alone records.3 These integrated systems have typically been called gateways or patient portals.2

Patient portals have been developed for primary care records4–8 and for specific patient populations, especially those with chronic conditions9 such as diabetes mellitus10,11 and heart failue.12,13 These systems are feasible, secure, and well accepted by patients.4,5,7,9,10,12,14–16 Although usability17,18 is likely to have a major impact on uptake and effectiveness of PHRs, usability has been studied less extensively and mostly by questionnaire or interview rather than by performance testing. In those studies, systems were perceived to be generally easy to use,4,5,10,12,14,15,19 although there was some variation in preferences for information presentation.

We located only four performance or scenario-based studies of PHR usability. In a 2004 UK study, patients systematically
viewed each aspect of their National Health Service Record for the first time while engaging in a semi-structured interview. The majority found it easy to use and useful. Kim et al. investigated the performance of 11 patients entering data related to care of their thyroid disorder, such as free text entry of diagnosis and prescriptions. Free text entry was more accurate for diagnosis than for therapy goals or prescriptions. Marchionini et al. compared presentation to patients of medical test results using bar charts and tabular formatting. Bar charts demonstrated consistently faster task completion times than tables, and inconsistently better accuracy. Finally, Tran et al. used scenarios to investigate a prototype of a locally developed PHR at the University of Washington and found difficulties with jargon which were improved in a subsequent iteration.

Other performance-based studies related to electronic health information use by patients may inform PHR design. A 2005 study investigated participants’ preferences for Web sites providing information about cancer diagnosis and treatments. After completing four scripted tasks, participants favored the prototype they felt was easiest to navigate and most clearly organized compared to a more “graphically appealing” Web site or one that was difficult to navigate and offered too many choices.

In addition, scenario-based usability studies that had children and adolescents as the end-users revealed specific needs and preferences for an electronic system. Children completing an online physical activity questionnaire did not use provided directions, but they did not have any problems entering information from drop down menus or selecting activities from a rolling list of options. Finally, two studies assessed electronic diaries for tracking pain of children with juvenile idiopathic arthritis (JIA). In one study, children were randomized to a paper or electronic diary. After 1 week of use, both paper and electronic diaries were acceptable and easy to use, but electronic diaries were significantly more complete and accurate than paper diaries. A subsequent study by Stinson et al. demonstrated the utility of usability testing in correcting ease of use issues. After entering pain data on a hand-held electronic device, adolescents reported several design issues that made the program difficult to use. Once the issues were addressed, no more ease of use issues were reported in a second round of testing.

Since usability is so important to effective electronic health applications and has been infrequently studied with regard to PHRs, we undertook scenario-based usability testing of three condition-specific patient portals designed specifically for parents of children with chronic illnesses. Our primary goal was to improve the usability of these applications, but we also expected to provide more generalized information to inform interface design for other PHRs.

Methods
This study took place at Cincinnati Children’s Hospital Medical Center (CCHMC). The study was approved by CCHMC’s Committee for the Protection of Human Subjects. Participants were parents of children with chronic illnesses and were paid $50 for their participation.

Description of Portals
The targeted system, MyCare Connection, is a secure web-based application developed at CCHMC. It allows clinicians and families to view key elements of the medical record as well as exchange secure electronic messages. Currently, three MyCare Connection portals are in use: juvenile idiopathic arthritis (JIA), diabetes mellitus (DM), and cystic fibrosis (CF) portals.

Each portal has features customized for the chronic condition it serves in a disease-specific tab. For example, the JIA tab contains quality of life scores which are not captured for other chronic illnesses, and the CF tab contains pulmonary function tests. Functions found in all portals include: demographic and contact information; laboratory, radiology, and pathology reports; inpatient and outpatient encounters; medications; secure, electronic messaging. Each function appears on a separate page, and navigation is guided by tabs for each function arranged across the top of the page (see Figure 1).

Participants complete confidentiality and system use agreements and are granted a user name and password. Multiple improvements to the portal interface were previously made based on formal and informal user feedback and heuristic usability testing.

Recruitment
Parents of children seen in the JIA, DM, and CF clinics were considered potential subjects if they had never enrolled in any MyCare Connection portal and were not computer or healthcare professionals. Participants were recruited by phone following an introductory letter from their clinicians.

Procedure
Participants were audio and video recorded while using the portal on a laptop computer with Internet Explorer to complete scripted tasks (detailed in Table 2, available as an online data supplement at http://www.jamia.org) on a prototype portal specific to their child’s disease. Usability specialists moderated tasks and acted as primary observers. Prior to portal testing, participants self-reported demographics and verbally responded to questions about their computer abilities, amount of time spent on a computer at home and work, and familiarity with their child’s condition. Participants’ computer skills were not tested. Then, a researcher demonstrated the “think-aloud” technique using a popular shopping Web site and asked participants to think-aloud as they attempted each task. Thinking aloud slows the thought process and increases mindfulness, which might prevent errors that might have normally occurred. However, when users are asked to perform simple tasks, the method has been shown to have no effect on user performance. Since the tasks used in this test are not considered complex, we chose to use the think-aloud method.

Scenarios were designed to test various areas of the portal reflecting expected use cases (uses of the site). Context was given to the use cases to create scenarios with which users expected to easily identify and relate. The tasks were chosen because they were representative of what portal using parents had previously told us through surveys and interviews were most important and common. Test, rather
than actual, data were used for all scenarios. Participants were encouraged to think-aloud and the moderator only directed a participant when he or she gave up or incorrectly completed a task. The wording of tasks was modified to reflect each disease-specific portal; however, the purpose of the task remained the same. Task time was measured from the time each participant finished reading the task directions aloud until the time the participant completed or gave up on a task, based on an embedded clock.

Following testing, participants completed the Computer Usability Satisfaction Questionnaire (CUSQ), which includes 19 items pertaining to efficiency, ease of use, and likeability of the system interface. Participants responded on a Likert scale ranging from 1 (“strongly disagree”) to 7 (“strongly agree”) and could also provide free text.

**Sequencing of Tasks and Testing Rounds**

As this was a formative evaluation whose primary purpose was improving the portals, developers made adjustments to the portal after each round of testing. The portal was then reassessed in the next round of testing, usability tasks were added, changed, or deleted to reflect condition specific content and learning from prior rounds (see Table 3, available as an online data supplement at http://www.jamia.org). In total, 6 JIA and 5 CF participants completed 14 tasks and 5 DM participants completed 13 tasks.

Based on feedback from Round 1 (JIA), interface modifications were made to general instructions, instructions for signing-up and logging-in, and abbreviations. The laboratory results page was redesigned, removing the horizontal scroll bar and providing more explanations of values and abbreviations (Figure 1). Side effects were added to medications, the amount of information was reduced, and column headings were clarified. To improve navigation, tabs were renamed; landing pages (the pages on the Web site where traffic is sent specifically to prompt a certain action or result) were created for medications and visits (Figure 2); the Personal Information tab was relocated to the final tab position, and a support menu was moved to the lower left hand corner. In addition, the size of page titles was increased and choices on the left hand navigation bar were highlighted.

Due to the changes in the logon process and instructions, Tasks 1 and 2 (requiring the user to request access and logon to site with a password) were added. For the next two rounds, these tasks were always administered first and in a logical sequence, followed by the remaining tasks in a random order.

Following the second testing round (DM), additional instructions regarding scrolling and normal ranges were included on the laboratory page and a larger button for...
logging-in was created. A confirmation was created after a message was sent. For clarity, the term “Ask a Question” was changed to “Send a Question”, and “X-Ray” was added after “Radiology.” Pull down submenus were added to each tab. A new tab allowed participants to assign rights to others, and a new task related to this function was added for Round 3 (CF).

**Analysis**

Data were analyzed using SPSS version 12.0 and Stata version 8.0. Means and proportions were used to describe the study population, time to complete each task, percent successful completion, and satisfaction scores. T-tests and one-way analysis of variance were performed to determine if each round of testing produced significant differences in time to complete tasks. The one-way analysis of variance was followed by Tukey HSD posthoc comparisons. Fisher’s exact test was performed to determine if each round of testing produced significant differences in satisfaction scores. Qualitative analysis included review of testing tapes and open-ended questions by two investigators. Emerging themes were abstracted and compiled from each round of testing, and main themes across all rounds of testing were identified.

**Results**

**Participant Characteristics**

Participant characteristics are shown in Table 1. The average age of participants was 39 years and 81% were female (two males participated in Round 1, and one male participated in Round 2). All participants defined themselves as White, except for one participant in Round 2 who identified herself as African American. Participants were very familiar with their child’s condition (64%), but their computer knowledge was low (44%) to medium (50%). Participants reported from 0 to 15 hours of computer use per week at home, and employed participants \( n = 12 \) reported 0 – 40 hours of use per week at work. There were no significant differences among groups regarding familiarity with the condition, computer knowledge, computer use, participant age, or years since diagnosis.

**Quantitative Analysis**

**Time to complete tasks:** Mean task completion times ranged from 73 (± 61) seconds to locate a document to 431 (± 286) seconds to graph laboratory results. Tasks requiring the least amount of time to complete required viewing only. The lengthiest tasks required more than one step or the interpretation of medical information (e.g., trends in laboratory...
Mean task completion times for tasks completed in all rounds are ranked from longest to shortest in Figure 3. Four tasks produced significantly different completion times between rounds. In order of decreasing significance they were: locate and view a letter from the hospital (F2,13 = 9.1, p < 0.01); locate and complete a previsit questionnaire (F2,13 = 8.5, p < 0.01); locate and view pathology reports (F2,13 = 6.5, p = 0.01); and locate and view radiology reports (F2,13 = 4.8, p = 0.03).

Percent successful completion: The percentage of participants who successfully completed each task is presented in order from highest to lowest in Figure 4. When analyzed in the order in which they were administered, there was no trend for ordering effects.

Only two tasks were successfully completed by 100% of participants: find password and logon to the site and locate and view a letter from the hospital. Three more tasks were completed successfully by all but one participant: locate historical medications, locate current medications, and locate radiology reports. One task, find the date of last pulmonary function test, was presented in Round 3 only and no one successfully completed it. The location of the pulmonary function test was not where participants expected it to be.

Two tasks had significantly different completion rates among the three groups. When asked to find and interpret laboratory results, 0% of participants in Round 1, 40% of participants in Round 2, and 80% of participants in Round 3 were able to do so successfully (p = 0.02, Fisher’s exact test). When asked to send an e-mail via the “Ask a Question”

![Figure 3](image-url)
function, 17% of participants in Round 1, 60% of participants in Round 2, and 100% of participants in Round 3 were able to do so successfully ($p < 0.03$, Fisher’s exact test).

Satisfaction scores: The means for each item can be found in Figure 5. Scores were normally distributed and therefore the mean was used as the measure of central tendency. Overall, there was no significant difference in satisfaction between rounds. Round 1 participants had a mean satisfaction score of 4.1 ($\pm 0.9$). Round 2 and Round 3 participants had nearly equal overall mean satisfaction scores (Round 2: $M = 5.3 \pm 0.7$; Round 3: $M = 5.3 \pm 0.9$). Satisfaction was greatest for interface pleasantness ($5.9 \pm 0.7$) and likeability ($5.8 \pm 0.6$) and lowest for error messages (text messages that appear on the screen when the user encounters an error in the system) ($2.3 \pm 1.2$) and clarity of information ($4.2 \pm 1.4$). Only one satisfaction item, “The system gives error messages that clearly tell me how to fix problems”, had significantly different means among the three groups ($p = 0.01$, Fisher’s exact test). The means were $M = 1.6 \pm 0.9$, $M = 4.0 \pm 0.0$, and $M = 2.5 \pm 0.7$ for Rounds 1, 2, and 3, respectively.

**Figure 4.** Percentage of participants ($N = 16$) successfully completing tasks from highest to lowest.

**Figure 5.** Mean and standard deviation of satisfaction scores from all rounds of testing ranked from highest to lowest. $N = 16$. 
Qualitative Findings
These findings fall into four categories: components that worked well, components that did not work well, desired enhancements, and unnecessary functions.

Components that worked well: There were numerous positive comments especially concerning convenience, assistance in dealing with insurance companies, and preparing for doctor’s visits. Several participants commented the portal would be particularly useful following doctor visits “because after a visit it is difficult to remember some of the things that the doctor told you.” In addition to being helpful, seven participants felt that navigation was easy, including spontaneous comments during testing such as, “This isn’t as intimidating as I thought”, and “After a while of messing with the system, you kind of figure out where everything is at.” Another participant found navigation to be the easiest part of the site, stating “you don’t have to use Back button a bunch of times to know where to go as on some sites.”

Components that did not work well: Medical jargon and terminology were problematic. Participants did not know what abbreviations such as “Fe” meant or what a pathology report was. There was also confusion caused by generalized phrasing. For example, one participant commented while reading about her child’s last eye examination, “What are normal levels, what are not normal levels?” Specificity was needed to let users know exactly what they were reading, especially regarding doctor visits.

Page organization was also an issue. Four people identified the disease specific tab as the most confusing component. The disease specific tab contains key information for a given condition, aggregated on one page. For example, in CF, it contains growth parameters, pulmonary function tests and key laboratory results. Although the page contained key information, participants rarely used it first. This persisted even after pull down submenus were added for the second and third rounds of the testing. This tab may have been considered a last resort as one participant reported, “Everything else seems to be under Diabetes, so I guess I’ll go there.”

Information overload was a worry for some participants who thought the portal could be simplified. One participant’s extreme viewpoint was “I can’t find anything on your website. There is too much detail for me to find anything!” Participants also requested help with the complexity and amount of some information.

Desired enhancements: Clarifications for numerical values and medical jargon were often requested. For example, participants preferred heights and weights in English rather than metric units and wanted an interpretation of the height and weight chart. One participant stated “These graphs are a pain. Do you have to figure it out? At the doctor’s office they explain it to me.” In addition, participants frequently asked for medical interpretations and explanations. This was particularly prevalent for laboratory abbreviations, values, and results. Although there was a hover feature on the laboratory results page that explained what most tests were used for, these were not adequate for the participants. Participants also requested more help options and bolder and more eye-catching sidebars and instructions.

They also wanted changes to the interface that would make finding things more obvious or available with fewer clicks, such as allowing other care providers access to the portal with one step and functionality to enroll online rather than on paper. Further, participants preferred a summary total of all inpatient, outpatient, and future visits rather than having to count each visit themselves. Finally, although most participants reported that the medication table was clear and understandable, some requested additional medication information such as side effects for past and current medications, allergic reactions, and why and when medications were discontinued.

Participants also wanted more direct feedback that their actions had the desired effect, such as confirmation of electronic messages or forms completed online. While completing a task requiring that a message be sent to a care provider, one participant commented, “I assume that this will be directed to the appropriate person” and another participant stated that she would like a time stamp or confirmation that the e-mail had been received. Participants also requested additional reminder functionality, links to research, and information regarding who would receive their e-mails.

Unnecessary functions: Participants found demographics and personal information unnecessary except to see if the information was correct. One commented that she thought more information would be under the demographics section, as she said that “I kept checking to see if anything I was looking for was there, which it never was.” Additionally, participants found text reports to be cumbersome and unnecessary, as one participant said that “all the text stuff” such as the documents found in “Procedures”, “didn’t mean much” and thought “the only way those may be useful is if [she] could print [the documents] out and give to her regular physician.”

Conclusions
Using scenario-based testing with participants unfamiliar with the system, we found numerous problems with terminology, portal navigation, task completion, satisfaction, and ease of use. These had not been uncovered in our previous heuristic usability testing or in focus groups and questionnaire feedback from parents who were portal users. Despite the difficulties they encountered, participants had many positive perceptions about the system.

Terminology: Medical terminology was a significant obstacle for participants, which could be because participants are not used to having to decipher medical terms by themselves. Our findings are consistent with past research. In studies by Pyper et al.10 and De Clercq et al.,11 participants also had problems with jargon even when they had experience with medical terminology. Similar to our experienced parents, the research by De Clercq et al. revealed that even patients who are regularly exposed to medical terms in office visits with their doctor did not comprehend terminology when they encountered it in their health records. Tran et al. also reported difficulties with jargon in their initial evaluation of a prototype portal. They redeveloped the PHR to include a “user-centered vocabulary.”23 We attempted to reduce jargon and improve explanations during development, although our final testing demonstrated some ongoing problems. In contrast, participants in a study by Cimino et al.4 did not have difficulty understanding the medical
information. However, the participants all had a college level education or higher and used a computer daily. Therefore, they might have more knowledge of what the medical terms mean or be able to find the definitions of these terms more quickly on the Internet.

**Portal navigation:** Adding landing pages and drop down tabs helped address the issues raised. Other studies have demonstrated that patients who use electronic health records have specific preferences for layout and navigation, favoring structures of sites that were most clearly organized even if they were less “visually appealing.”

Similar to our methodology, Tran et al. asked participants unfamiliar with the test system to complete tasks and provide feedback, although not in the “think-aloud” method we employed. His participants preferred a navigation scheme that resembled the organization of physical file folders, which is similar to the drop down tabs we added to make navigation easier.

Participants also encountered problems viewing test results and graphing. Similar to problems with medical jargon, participants were not prepared to interpret this information alone. Based on feedback in Round 1, we converted to bar charts for laboratory results, with some improvements in subsequent testing rounds. These findings mirror those of Marchionini et al., whose participants also performed better with bar charts than tables. Difficulty with graphs was also reported by Cimino et al. and Earnest et al. Unlike our participants, Cimino’s participants had, on average, 19 months experience with the site, giving them numerous opportunities to explore it. Despite this added experience, even these users had trouble with the graphing function.

**Task completion:** Participants in our study desired features that add “reassurance” that they had successfully completed a task. They also wanted to know where information was being sent and who received their messages. For users like ours that are concerned about the transfer of medical information via electronic communication, future usability testing should include tasks to assess if users understand who is concerned about the transfer of medical information. However, the participants all had a college level education or higher and used a computer daily. Therefore, they might have more knowledge of what the medical terms mean or be able to find the definitions of these terms more quickly on the Internet.

Despite the scores on satisfaction being highest for interface pleasantness and likeability, the low satisfaction with the error messaging system is of more importance in regards to system usability. In commercial systems, the focus tends to be more on having a likeable and pleasant interface while aspects of the program that would hinder use are often overlooked. Therefore, problematic error messaging should be addressed as an area of significant concern and necessary improvement because its effectiveness can ultimately determine if people will continue to use the system and use it as intended. We are addressing this in our portal.

In addition, less than half of participants said that they were satisfied with the system overall and almost 60% chose more neutral items when answering if they were satisfied with how easy it was to use the system. In previous non-scenario based studies, participants said the electronic health record they accessed was easy to use. The difference in the scores on ease of use between our participants and those in other studies may be due to several factors. For example, participants in the studies by Cimino et al. and De Clercq et al. all had experience with computers and may have been more computer savvy than our users. Further, previous scenario-based studies provided participants training before they used the electronic system, which we did not. We wanted to understand how users would respond with minimal training and experience. This design decision limited our ability to compare our study with others in which participants received prior training. Finally, we used a comprehensive prototype that replicated the complexity of the data parents would actually experience regarding their child with a severe chronic condition. Thus, it may have been more complex than the data presented in other studies, as most of those portals were designed for use in primary care.

Lastly, many problems encountered by participants could be traced back to assuming too much knowledge. Thus, content elements of the portal were not laid out clearly enough and spelled out sufficiently for a novice’s understanding. For example, without a clear outline of the process of how to sign up and log in, less computer savvy users were confused by the large amounts of text on introductory pages.

**Positive perceptions about the system:** Parents rated convenience and empowerment as key positive attributes of the system, a finding that echoes previous research. Participants in the study by Pyper et al. also cited benefits, including being better informed about their health, having medication lists, and having reminders of future appointments and screenings.

In addition to liking the convenience of the system, participants liked the medication pages and laboratory results, and the majority thought they would access both frequently. Knowing information about medications and laboratory results is very important to medical patients so it is expected that such functions that access them would be used often. Our findings are consistent with others who found that participants spent a majority of the time viewing laboratory results.

Participants perceived that understanding the layout of the portal was something they could learn reasonably quickly and they scored the CUSQ highly for learnability. Several
participants commented that once they started exploring the site, they could find their way around. Although we did not examine this area, future evaluators might investigate which tasks remain difficult over time for users and thus focus their redesign efforts.

**Limitations:** There were limitations to our research. First, we studied only parents who had never used the system. Additional study with experienced users would have broadened our knowledge of how people interact with the system over time. We also did not formally test participants’ computer skills. We assumed that parents who would participate in a computer-based study would have enough computer skills to navigate a web page effectively. However, because we made this choice, we could not distinguish difficulties due to limited computer skills from specific portal navigation problems. It is possible that many of the problems participants had resulted from a lack of computer skills than portal issues. This should be addressed in future work.

Participants might have experienced anxiety over not having someone available to help them interpret information. Portals, like other stand alone electronic systems, lack immediate interaction with clinicians, a fact discussed by some participants. Understanding parents’ experience with actual use of their children’s data would be an important area for understanding risks and benefits of portals.

Additionally, we modified the portals between rounds of testing, making improvements based on findings. We made an a priori decision to conduct formative testing since our main goal was to improve the portals and we could not afford to do sequential usability tests of each portal. This approach balanced available resources, desire for portal development, and scientific certainty.

Performance based usability testing is widely advocated but has rarely been carried out in developing patient portals. While such usability testing can be expensive, the current study demonstrates that it can assist in making healthcare system interfaces for laypersons more user-friendly and potentially more functional for patients and their families.

**References**


