USER CENTERED DESIGN IN THE CREATION OF A SURGICAL INTENSIVE CARE LOG

by

Eugene Wai Ching Leung

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biomedical Informatics

The University of Utah

May 2011

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Eugene Wai Ching Leung**

has been approved by the following supervisory committee members:

Nancy Staggers	, Chair	**12/9/2010** Date Approved
Bruce Bray	, Member	**12/9/2010** Date Approved
Charlene Weir	, Member	**12/10/2010** Date Approved

and by **Joyce Mitchell** , Chair of

the Department of **Medical Informatics**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

The Residency Review Committee (RRC) requires that general surgery residents document their Surgical Intensive Care Unit (SICU) experiences. To satisfy these requirements we created a web based intranet log to make it easier for residents to track their patients and determine when these requirements were complete. A premium was put on usability to promote acceptance by surgical residents. A prototype web site was designed with input from an attending general surgeon. Three general surgery residents were selected to participate in the iterative design phase. They went through three iterations using a "think-aloud" method while performing tasks on the prototype web site. Each iteration led to improvements to the web site. In a comparison test, a group of seven medical students performed 14 typical web site tasks using both the prototype and the final versions. They were asked to complete a Questionnaire for User Interaction Satisfaction (QUIS) for each version. The time for completion of these tasks was also recorded. The user interaction satisfaction did not show any improvement ($F(1,6)=0.13$, $p=0.912$). Similarly, there was no improvement in times for delete and add tasks (Delete $F(1,5) = 0.949$, $p=0.375$, Add $F(1,5)=0.267$, $p=0.628$); however, the time to complete edit tasks was faster for the final version of the web site ($F(1,5)= 14.3$, $p=0.013$). The primary reason for not detecting other differences

between the two web sites is likely that the comparison study did not have sufficient power. This was suggested by the participants whose comments favored the final version over the prototype as well as a trend of consistently higher mean subset scores in the final version. The results indicate that differences may be seen when more complex tasks are completed (editing information) versus the two simpler tasks (adding or deleting a patient record in a web site). Future studies should focus on the impact of navigation strategies on speed and data warehouse approaches to creating the application. This study shows the benefits of using an iterative design approach to create a usable web site and demonstrates the importance of further research in the field of usability.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

General surgeons are responsible for acute surgical care in various environments [1]; accordingly, surgeons must be able to handle a wide variety of patient conditions in intensive care units. Toward this goal, the residency review committee (RRC) now requires that general surgery residents prove that they have achieved this breadth of experience. The RRC for General Surgery, one of 27 committees responsible for accrediting graduate medical education residency programs, requires that general surgery resident programs document all surgical ICU experiences [2]. The goal is to ensure that surgeons completing an accredited program will have had experience managing a wide variety of problems; including managing patients in a nontertiary care Intensive Care Unit (ICU) setting until transport to a tertiary care facility can be arranged.

The requirements laid out by the RRC are complex (see Appendix A). There are seven essential categories that need to be addressed in a critical care residency program. Within each category other required information must be recorded. By the end of the residency program, the resident is required to have seen at least one patient in each of the seven categories, with each patient having had problems in at least two categories.

The conventional approach to tracking the information accurately would entail the use of a paper log; however, there are many problems with this approach. Paper logs are easily lost or inaccessible; if misplaced in a public place, the loss may represent a breach in patient confidentiality. Legibility of physicians' handwriting is often suspect [3] and can be difficult to interpret even by its author. Collecting logs from residents and manually confirming their compliance with the RRC criteria can be extremely time consuming and prone to error. Administrators stated that it often takes office staff one to two weeks of full-time effort to determine whether all residents conform to the requirements of the RRC. [4]

Our solution to these issues was to create a web-based log. A web-based log can never be lost since it resides on a hospital-based server. The omnipresence of the web offers residents ample opportunities to enter their data. The web-based log also enjoys the same security and other protections as the hospital's medical record system. Use can be restricted by having it accessible only via the hospital's intranet, a process that is relatively secure. In addition, because the data is discrete, checks can be run to ensure that the residents' logs are in compliance with RRC criteria.

The overarching goal of this project was to develop a web-based log and evaluate its usability. Usability was a key concern in the development of the web-based log. If residents felt that it was too complicated to use, they would resort to a paper-based system with all of its pitfalls. With that in mind, we

sought a process that would provide leeway to improve the usability of the program.

We set several objectives for this research study. First, we planned to create paper mock-ups of the web site. After these were approved, the paper mock-ups were used as a prototype for the web site connected to a dummy database. To improve the prototype we would use an iterative design process for the prototype until minimal changes were necessary. Finally, we planned to run a usability comparison test to confirm the results.

BACKGROUND

In this section I will discuss the meaning of "usability" as it applies in this paper. I will demonstrate the importance of an iterative approach in usability studies and show the benefits of this approach. There are other approaches that may be appropriate in usability evaluation, and these will be reviewed as well as rationales for their use.

Definition of usability

User centered design is both a philosophy and a process [5]. "It is a philosophy that places the person (as opposed to the 'thing') at the center; it is a process that focuses on cognitive factors (such as perception, memory, learning, problem-solving, etc.) as they come into play during peoples' interactions with things." To paraphrase, the design process should focus on the users and the context the users find themselves in. Users are the most knowledgeable about their own workflow, and so it is important to include them in the design process [6]. This is also a way to build user acceptance of the product.

Gould and Lewis in their classic paper on usability, note three principles in user centered design: 1) An early focus on users and tasks; 2) Empirical measurement of product usage; and 3) Iterative design whereby a product is designed, modified and tested repeatedly [7]. An early focus on

users and tasks will inform the vision needed to guide the creation of the product. In the past, the design of many products considered only the ability of a product to function but left it to the user to figure out how to get it to work. If products are designed with an eye to the user's workflow and perception of ease of use, it is more likely to be successful [8].

Rationale for usability

The National Academy of Sciences was asked to study how best to use the computer science research community to determine how to use technology to improve health care [9]. In their research they note that many current IT systems do not take advantage of existing human-computer interaction principles and, as a result, "increase the chance of error, add to rather than reduce work, and compound the frustrations of executing required tasks." [9] Consequently, one of the recommendations was for government to encourage research in three critical areas, one of which was usability.

The experience of the Children's Hospital of Pittsburgh (CHP) best illustrates this point. In October 2002, CHP purchased a Computerized Physician Order Entry (CPOE), becoming one of the first children's hospitals to become 100% CPOE [10]. Its goal was to decrease incidence of medication error rates and improve hospital resource utilization. Unexpectedly, analysis of the mortality rate pre- and postimplementation showed an increase in mortality rate from a baseline of 2.8% to 6.57%. The nonsurvivors were more likely to be admitted to the ICU, younger, premature or referred for surgery.

Further analysis indicates that before CPOE the hospital met national guidelines for timeliness of administration of critical medications; however, after implementation this occurred only half of the time. The authors noted that difficulty in entering the medication might have been a contributing factor, especially for time-sensitive medications. The authors went on to describe the onerous way orders have to be entered into the system.

The authors note several methodological flaws in this study. The authors note that system errors may also have contributed to the increased mortality. In the example that was given, the authors note that the nursing staff occasionally would dose antibiotics according to the times noted by the computer and not by dosing intervals, that is time between medications. The medication example cited by the authors was antibiotics, but it is conceivable that close dosing of other medications (such as vasoactive medications) may have adversely impacted mortality. The authors also note that the patient population (interfacility pediatric transport patients) comes with a different level of acuity and may not be representative of the general hospital population. Furthermore, they note the short post implementation period and that the changes noted occurred during an "adjustment period that commonly follows any sweeping change."[10] Indeed, other researchers at the same hospital noted a decrease in the rates of adverse drug reactions over the same period of time.

Although the true cause of the mortality is unclear, the authors raise

some valid issues. The steps required to order medications are difficult, and the implication – that the rise in mortality is related to poor usability – is certainly plausible. There is evidence to support this assertion outside of this study. [11]

Usability is particularly important for physicians in training. An article by Ash [12] notes that at one site residents threatened to strike because of their dislike of the physician order entry systems. In this article it is very clear that one way to avoid such confrontations is to engender a feeling of trust. Ash notes that physicians hate to be forced to think like a computer or forced to use workarounds to complete documentation. Nonetheless, in the hospital where residents almost went on strike, they kept working with the IT staff to continue to modify the system until the outstanding issues were resolved and the system became a source of pride.

Although many authors identify usability as important, finding articles that comply with Gould's principles is surprisingly more the exception than the rule. Of the articles reviewed only a few are complete from this perspective. [13-16] Gould [7] notes in his paper that, in general, a single iteration is not enough to discover all the usability weaknesses in a program. Authors of several studies [17-18] claimed that their design was iterative but actually had only one iteration. Gould does not explicitly state the number of iterations that are required. Rubin [8] offers a reasonable approach stating that the number of iterations should depend on whether

significant issues are discovered by the subjects. When there are no further major issues identified by the subjects, the iterative rounds are complete.

Gould's three principles in practice

Johnson's usability study [13] was an effort to redesign a family history/ pedigree drawing product that had already been released. The software product needed to be redesigned because it lacked several user required features and had a host of usability problems. In this redesign, there was an extensive focus on users and tasks using multiple methods to define user requirements for the new prototype. These methods included questionnaires to over 1200 people and comparative analysis of three competing products. The prototype was then subjected to an iterative heuristics evaluation and later a "think aloud" evaluation by eight subjects performing 12 common tasks. When the versions were compared, typical tasks required on average 48.7 seconds in the original but only 34 seconds ($p<0.001$) in the completed version. Furthermore, there were significant improvements in ability to complete tasks and in user satisfaction with the redesigned product. The authors noted that, while they were successful in their redesign, it would have been less costly if the process had occurred in the initial creation of the software. This is a common observation made by others as well [8, 19]. Also, after the heuristics evaluation of the prototype, the "think aloud" evaluation was still able to detect major usability problems.

Boyington's study [14] was performed in an attempt to create de novo an educational web site to facilitate continence promotion in patients. Interestingly, this is one of the few studies reviewed that attempted to model expert thought processes using knowledge engineering, which is at its heart an iterative design process. The model was created after intensive (six hours) questioning of a continence expert. The model created was validated by another expert in continence. The iterative portion was composed of four rounds and involved 12 subjects using the web site. Of the subjects, two had never before used a computer. Then, the web site underwent a user satisfaction evaluation, which demonstrated positive user satisfaction ratings. They made interesting use of an "Interface Metaphor" in an attempt to create a familiar setting for the subject. The metaphor in this case was of a health clinic, since the subjects were selected from this population, and thus facilitated the subjects' use of the web site.

Taylor's study [15] represents an excellent example of classic employment of usability testing as described by Rubin [8]. Taylor's research involved creation of a web site to screen for amblyopia, a relatively common eye disorder in pediatrics. He performed rounds of iterations for the web site design using six sets of parents with their children. After the final iterative portion, a validation test was performed by asking parents to fill out a user satisfaction questionnaire after using the final site. In the verification test his web site met 21 of 22 criteria for improved usability. During the study he

observed that after the fifth subject there were no new usability issues observed and that a more efficient use of time would have had more iterative rounds with fewer subjects in each round. This observation is probably true in his example but would likely vary depending on the complexity of the task.

Wachter's research [16] involved the design of graphical displays for use of anesthesia. These displays were used to represent anesthesia scenarios such as airway obstruction or hypoxia. These displays began with a prototype created by a multidisciplinary team that included clinical psychologists, clinicians, bioengineers and human factors engineers. The iterative design process led to four other designs. After further testing on whether subjects were able to match a display with a clinical scenario, one display was chosen as the best. Subjects' abilities to interpret the displays improved from 70% to 98% after redesign.

Redesign of the Federal Emergency Management Agency (FEMA) web site also adopted these practices; the results were available only on a government web page rather than in a peer reviewed journal. In April 2005, FEMA started a year-long process to improve its web site [20]. In this year-long redesign, they did a comprehensive analysis of numerous sources of data: online user surveys, call data, and development of personas. Subsequently, they recruited representative samples of users of the web site, including disaster victims, emergency personnel, insurance agents, architects and others to iteratively improve the web site. From this comprehensive

approach, they noted a 93 % improvement on user performance based on ability to find specific pieces of data, nearly a 50 % reduction in time required to find data, and an improvement in user satisfaction by 22 %.

The studies mentioned above are interesting because they demonstrate the benefits of adherence to Gould's three principles. Each of the articles was able to reveal improvements in a variety of usability scenarios. These studies showed improved time savings, a better ability to perform tasks, and improved user satisfaction. In particular, Gould mentions that most designers have only minimal contact with their users and often misestimate the value of that relationship [7]. An iterative approach involving users can be a way of getting insight not available from the developer's viewpoint. Another benefit of iterative approaches is that they tend to create a relationship between the developers and the users, which will in turn encourage collaboration. This is demonstrated in the improvement in satisfaction with the application as demonstrated in the previous articles.

Heuristics in usability

A heuristics approach offers some advantages over an iterative one. In one study, a web site was designed to address the nutritional needs of those in a disadvantaged rural population [21]. This design was done in three rounds – a requirements gathering round, an iterative design round, and a confirmatory testing round. Interestingly, the iterative development portion lasted from February to "the end of summer" or approximately 7 months. In

contrast one article used heuristics to evaluate a web site, and this evaluation took approximately 80 hours to complete [22]. The study noted difficulty in their recruitment phase, which may have impacted the time it took to complete the study.

A heuristic evaluation involves the assistance of a usability expert or a domain expert who is given a series of usability guidelines to follow and asked to make recommendations for improvement of the software. In one study [23], several well known heuristics, such as Nielson's 10 heuristics [24] and Shneiderman's 8 golden rules [25], were combined into a list of 14 heuristics that were deemed to be the most pertinent. These heuristics were applied to a paper mock-up of the hospital's Computerized Information System (CIS), and recommendations for change were made. Using these heuristics, they were able to identify several usability issues and they were able to do this within a month.

"Think aloud"

Other usability studies have attempted to capture the users' experience in other less time-consuming ways. In one study, the authors redesigned the Medication Administration Module [26] commonly used by nurses in a hospital setting. In this study, the subjects were asked to "think aloud" while performing their usual tasks in a normal workday. The data collected were deconstructed into a workflow diagram and later, with what

kind of outcome, were included into the program. The use of "think aloud" in this study served as a useful tool to capture user requirements and workflow.

Another study evaluated a computerized patient record using nine physicians as subjects [27]. The subjects were given tasks and asked to use the think-aloud process to communicate their thoughts. Content analysis was performed on transcripts of the sessions as well as their videotaped interactions with the system. Based on this information, recommendations for changes to the application were handed to the programming team and, after implementing the changes, a 10-fold decrease was noted in the average number of user problems.

The think-aloud process is not without its critics. In one study, the authors created a web site to assist in the management of depressive symptoms in HIV patients [28]. Web site development started with a heuristic evaluation by three usability experts who uncovered 14 usability problems. A single iteration was performed based on the usability experts' evaluation. This iteration was validated by a confirmatory test by six potential users. The authors found that the experts were more focused on information design and the users tended to focus on navigation and access. However, they noted that their subjects had some difficulty in adapting to the think-aloud process and noted that the users' comments were subject to interpretation.

Usability in intranets

Very few studies conduct usability testing of intranet sites in a healthcare setting. Regarding intranet sites in general, Nielson [29] notes that two crucial points differentiate intranets from web sites on the internet. First, the intranet belongs to your company, and it is unlikely the employee will need to search for the right site. Second, the employee's familiarity with the intranet site will grow as the employee uses the site more. Because there is a difference, one would expect that there will be usability differences as well, but this is poorly documented in the literature. In a search conducted using the keywords "Intranet" and "usability," over 300 articles appeared although only one was applicable to the topic at hand.

In the applicable article [22], a company's internal studies revealed under utilization of an organization's intranet resources. The author had two experts review a web site using 10 of Nielson's heuristics to guide the experts (Appendix B). A single evaluation round was done because of time limitations. The evaluation round was then followed by a confirmatory study involving 18 physicians whose task was to find certain pieces of data. In spite of this noniterative approach, the experts, guided by a heuristics approach, were able to suggest a significant number of changes that improved speed, decreased errors and increased user satisfaction.

Summary

The studies above highlight the benefits and problems with some of the approaches to evaluating usability. Iterative approaches, involving actual users, give direct input from the users, building both user acceptance and allowing users to warn designers of possible pitfalls to workflow. The downside is the time required to develop such products. Heuristics is a faster approach; however, it often requires expert opinion and leaves users out of the loop.

The think-aloud method was used in a number of studies [26, 30] and has been shown to be helpful in gathering information. One study raised valid concerns about the need for vigilance with the interpretation of the subjects' comments. We will address those concerns in this study by following a study protocol and use of open ended questioning.

Given how important it is that general surgery residents accept use of the software and the impact that the studies can have on future users, an iterative approach to design the web site makes sense.

The paper will first discuss the iterative design portion of the study, which includes the creation of a case report website and the iterative design process. The paper concludes with discussion of the comparison test which measures the prototype against the final version.

ITERATIVE DESIGN: METHODS

In this section, I will discuss the methodology used to create a case report web site developed using an iterative design process and then discuss the results surrounding this part of the research. I will describe the sample used, the setting and the instruments and procedures used to carry out the study. I will also discuss how the scenarios were created, and the think-aloud methodology. I will describe how the data were collected and the content analysis process.

Institutional Review Board approval from the University of Utah was obtained and the study was determined to be exempt under 45 CFR 46.101(b).

Sample

We recruited three subjects to perform the iterative portion of the study. Recruitment was originally done by email; however, due to poor response, faculty aided in finding three surgical residents to volunteer. These residents were compensated by additions to their CME fund. Sessions were held in the late afternoon after the residents had signed out for the day.

Setting

The University of Utah's University Hospital is a teaching hospital located in Salt Lake City, Utah, and home to the University of Utah's General Surgery program. The study was performed in a conference room at the hospital. This room was distinctive in that it came equipped with multiple monitors attached to a single desktop computer. This was convenient as it allowed researchers to videotape one of the unused monitors, minimizing the subjects' sensitivity that they were being videotaped. The subjects were told that they were being videotaped before each session.

Instrumentation

The RRC criteria (see Appendix A) were reviewed by the author and formed the basis of minimum criteria for the paper prototype. The other source of content for the paper prototype was a sketch (see Appendix C) by Dr. Holman, Assistant Professor of General Surgery, of how he envisioned the web site. The paper prototype was essentially a copy of the sketch with navigation added around it.

The web site was developed by Ming Tu, programmer, in a Cold Fusion (® Adobe Systems, San Jose, CA) environment with extensive technical support from University of Utah's Information Technology Services (ITS). Data were stored on a dummy Oracle database that was kept on the hospital's information system to replicate a live environment during evaluation sessions.

The pick lists were populated with options for areas such as modes for the ventilator. This task was performed by two surgical ICU attendings. The separate lists were compiled into a common list and incorporated into the appropriate parts of the program.

Procedure

There were basically three tasks the subject could perform. The subject could enter data for a "New" patient, one that is not in the log already and needs to have all relevant sections completed. The subject could also "Delete" a patient that is already in the log. Finally, the subject could "Edit" certain details about a patient who is already in the log. In the "New" and "Edit" tasks, we deliberately stayed away from text entry as subjects may have different levels of typing skills. The last two tasks were deliberately designed to enable completion by using drop down lists.

At the beginning of each round, the subjects were given a form that outlined the goal for that session, and consent was obtained for video and audio taping the session. The tasks typically started out with the addition of new patients. The addition of new patients to the database would create the data for other actions such as edits and deletes. The tasks also included activities such as describing what items were required to complete the log. The tasks were simple enough that no training was required. This assumption was later corroborated by the test subjects' comments.

Scenario creation/Think aloud

The tasks for each round were created by examining what functions would be performed by the residents when using the application. The tests involved entering data on new patients, editing data on patients already entered and deleting patients. The tasks also led the test subjects to areas that had recently been changed to get immediate feedback. The author acted as facilitator and assisted in the use of the application. The risk of undue bias was mitigated by using an open-ended questioning format. The subjects used the think-aloud method to communicate what they thought about the application. Their actions on screen while using the web site were captured by video camera, and their comments were captured on microphone.

The think-aloud method [31] can be an effective way of documenting what subjects are thinking while they are performing the tasks. This method is simply having the subject speak aloud while performing set tasks. At the beginning of each session they were given an instruction sheet which they read aloud and performed. A copy of this instruction sheet can be found in Appendix C. This method can be a rich source of information allowing the subject the opportunity to express thoughts and feelings as the subject performs the tasks. [8]

Data collection

Data from such sessions can be collected in a number of ways. [8] Handwritten notes are perhaps the most flexible, least costly method. Screen

capture devices are also available, although one must be careful that the recorders do not have an impact on the application. Video and audio tape sessions are also options. The key point is that in collecting the data the test monitor must be careful not to influence the subject in an undue fashion.

The data for this study were collected using all of these methods except screen capture devices. Screen capture was accomplished by the camcorder which recorded the subjects screen actions as well as audio. While subjects completed tasks using one monitor, a Sony DCR-TRV 530 digital video camcorder recorded the subject's voice and activities on another monitor. This allowed researchers to record the session discretely.

Content analysis

Video and audio content need to be translated into a usable format. Content analysis helps create a formalized, systematic way of achieving this goal. [32] For our study, the preceptor notes were reviewed before viewing the videotape. As the subject made observations, we noted how the subject interacted with the application. These comments were written on a summary sheet for that subject and round and compared to comments from the other subjects. These comments, in combination with observations on mouse movements, were interpreted and compiled into a list of changes for the development team to make. We measured time to complete tasks by using a stopwatch.

If trends in the comments were noted or if the resident was observed to have difficulty in a particular area, this was discussed with the subject and a technical translation occurred where data from content analysis and these observations were turned into software requirements. These were incorporated into the next version of the web site.

ITERATIVE DESIGN: RESULTS

In this section we will discuss the subjects, the improvements made in each iteration, and the results of the time trials that were performed as pilot work for the comparison test.

Subjects

The three surgery residents came from varying technology backgrounds ranging from someone who just used technology for shopping to another who had done some programming in the past. One subject was a 31-year-old female who reported using Microsoft Excel (® Microsoft, Redmond, WA) once, had never used Microsoft Access (® Microsoft, Redmond, WA), and had no programming experience. The second subject was 29-year-old male who had used Excel and Access and had done some web programming. The third subject was a 30-year-old male who had used Excel and Access but had no programming experience.

Improvements

The recommendations for improvement collated from notes, video and audio are summarized in Table 1. One item that is not obvious from the figure is the number of pick list changes that the residents requested for nutrition. The choices that were listed are markedly different than

Table 1. Summary of improvements from the "thinking aloud" method.

Round 1	1. Added a previous and a next button for improved navigation. 2. Pop up appears if patient already in the log. 3. Graphical representation of which categories are completed. 4. Improved location of the "Add new patient" feature. 5. Changed pick list to checkboxes to account for multiple weaning modes. 6. Added, deleted or altered pick list choices.
Round 2	1. Check to see if category is completed if subject moves to another category before required sections are filled out. 2. Add "Hourglass" for delays in page loads. 3. A concise summary of the RRC requirements (i.e. 20 patients with 2 complete categories, at least 1 patient in each of the 7 categories) was placed on the summary page to clarify what was required. 4. Change the graphical representation from light bulb to checkbox, add question mark for patients that are incomplete. 5. Added, deleted or altered pick list choices.
Round 3	1. Too easy to delete a patient. A pop up would appear when deleting a patient checking to make sure that you really wanted to delete a patient. 2. Access to the original RRC document describing the requirements (Appendix A). 3. Option to print record. 4. Added, deleted or altered pick list choices.

the choices the SICU attending physicians had recommended. In many cases the residents did not agree with the available pick list choices. Screenshots displaying the prototype and final versions are available [33].

Time trials

In the second and third rounds, time trials were performed as pilot work for the comparison study. The subjects were timed using tasks similar to what they used during the think-aloud sessions and similar to what was planned for the comparison test. The point where user times plateau

determined how many trials it took for the learning effect [34] to extinguish. The learning effect accounts for variability in any subject's tendency to improve their performance after repetition of a particular task. The graph suggests an inflection point at seven trials for data entry of a new patient in Figure 1. There are several times when subjects' times unexpectedly increased, such as with Subject 3 at the 4th trial and Subject 2 at the 6th trial. Both subjects were noted to be post call during these episodes. Attempts to retime led to similar results, because during the retiming trials they were also post call. Each trial was carefully crafted only to vary in the item picked in the pick list. No free text was required in the time trials.

The data for deletion (Figure 2) and editing (Figure 3) of a patient are also presented. The deletion of a patient was relatively simple, although one of the subjects missed a trial because the subject flipped two pages instead of one. As a result, one of the subjects is missing a data point. The inflection point for deletion is at about two trials. In editing a patient, the inflection point appears to be about two to three trials. Subject 1's time ended up increasing for unknown reasons, although it is suspected that it was related to this subject being post-call.

Summary

The iterative section led to several improvements in the application. Improving the navigation, selecting pick list choices that did not lead to cognitive dissonance, and introducing a graphical user interface for the

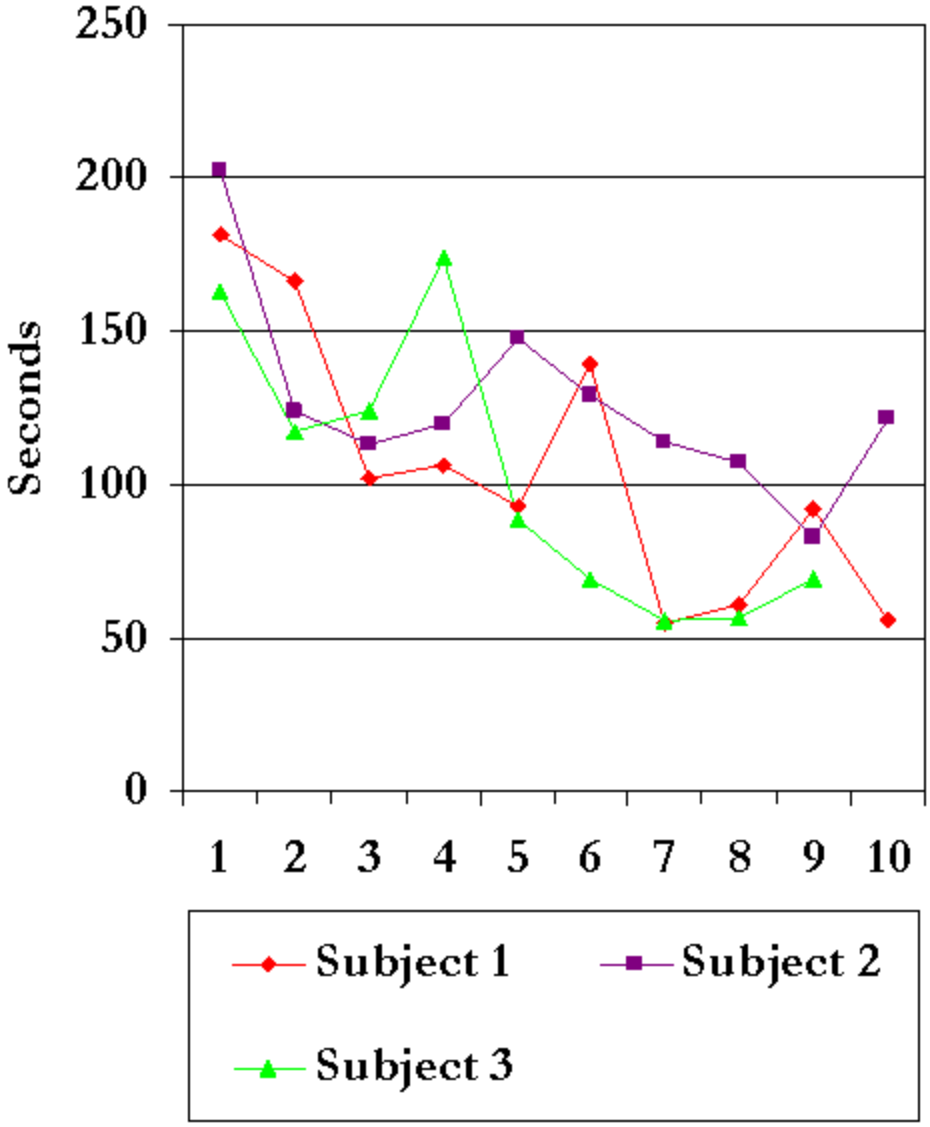

Figure 1. Time spent for data entry.

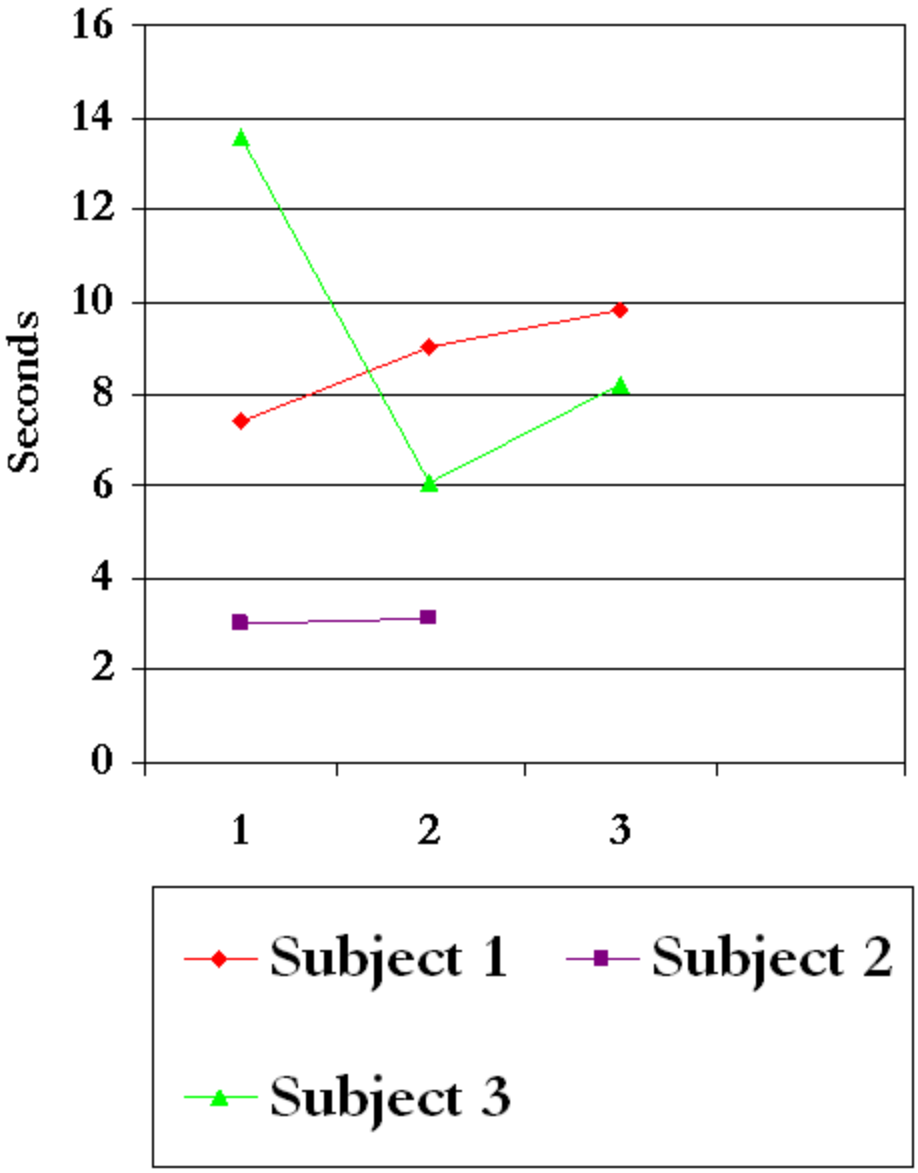

Figure 2. Time spent for deleting data.

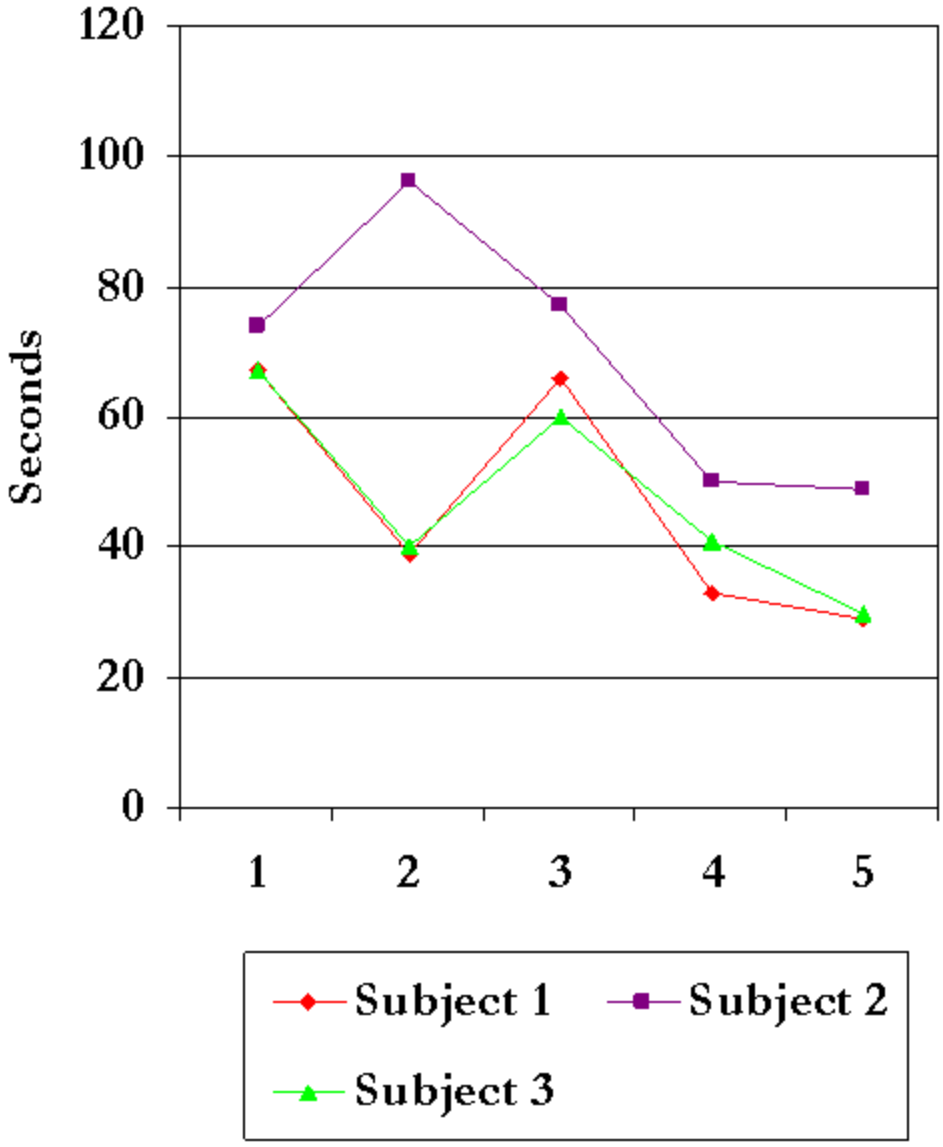

Figure 3. Time spent for editing a patient.

summary page appeared to make the subjects more enthusiastic about using the application. Time trials revealed that the learning effect for Add new patient, Edit, and Delete tasks seemed to plateau at the 7th, 2nd and 3rd trial respectively.

ITERATIVE DESIGN: DISCUSSION

The iterative rounds led to successive improvements to the web site, and all of the subjects noted that they would be happy to use it to document their log. Major improvements included improved navigation, addition of GUI elements in the summary screen, and modification to pick lists.

Number of subjects

In an optimal study, one would have four to five subjects which, according to Virzi's classic article [6], should predict 80% of the major usability problems. Based on his studies on usability, there is an inflection point at four to five subjects where, for each additional subject added, you would yield minimal additional usability problems discovered.

Some evidence supports the use of lower numbers in the iterative phase, especially in settings where the task is not complex and the users are relatively uniform. "Discount usability" has been used with some success, and Nielson comments in one study that in some cases only one subject was needed using a modified think-aloud method. The results of these iterations was later validated in a larger study (n=38). [19] The homogeneity of the population (all surgical residents) also improves the likelihood that this will be valid. [35] The factors noted above, plus the relatively small number of

potential users at any one time (n=15 to 20), suggest that three is a reasonable number for the iterative portion of this study.

Navigation

In the iterative rounds, we were surprised that the subjects requested a previous and next button. The subjects had the opportunity to navigate using the left side of the screen, which appeared to be a more efficient way to navigate. While reviewing the video, it was noted that two of the three subjects tended to intuitively bring their mouse to the lower left hand corner of the screen when it was time to go to the next organ system. As a result, it was decided to put the previous and next buttons in this location. As the subjects became more familiar with the web site by the end of each round, all were using the left sided navigation.

It is interesting how the addition of these buttons mirrors the constructivism theories by Piaget. [36] Piaget's contemporaries note that to know an object, one must be able to recreate it. Piaget suggests that in doing so one will only be able to create the original object. Piaget argues that true learning comes from being able to take the object and build from it in some meaningful way. More recently this ability to start with something known and build from it has been termed "scaffolding" and plays an important part in learning theory. [37]

To put this in context, the residents, likely from previous experiences, found that they learned best going in a step by step fashion and, as a result,

asked for previous/next buttons. The previous/next buttons would, very inefficiently, walk one through each of the seven organ systems, whether documentation was needed in these areas or not. However, as they got used to the buttons, they learned that the faster way to navigate was to use the left sided navigation. Allowing for multiple ways to navigate through a web page in this case appeared to function as training wheels serve a future bicyclist. With two ways to navigate, the user can take the training wheels off when the user is comfortable and then go to left sided navigation for faster performance.

Graphical user interface

Users also asked for a graphical interface for the summary page, which is understandable since at least 20 patients need to be recorded. With over 20 patients on the page, the page will start to look cluttered and difficult to interpret. This request is consistent with research done by Staggers and Kobus. [38] They compared a GUI versus a text-based interface and demonstrated that the GUI interface led to twice the response time (speed) of the text-based interface. Also, users of the text-based interface experienced six times the error rate.

Pick list changes

One of the other striking features from the iterations was the number of changes the residents made to the pick lists. When designing the web site,

I had presumed that the SICU attendings would be using common terminology since the residents and the attendings work together in the SICU; however, this apparently was not the case. For example, in hemodynamic instability the residents were able to come up with seven more etiologies than the attendings. The items included hypovolemia, cardiac, neurogenic, anaphylactic, sepsis and adrenal insufficiency.

Several other features that were added could have been predicted with more foresight or training in usability. The main piece of data to be loaded was supposed to be demographic data; in spite of this, server performance was an issue. At times some frustration was felt because the cursor did not turn into an hourglass as expected and the user kept trying to enter data. The ability to delete a patient was too easy and should have included a popup warning of the deletion. The subjects also requested the ability to view the formal recommendations of the RRC when in the summary view.

Summary

The iterative design process identified many major usability issues with the original web site. The findings support the assertion that iterative design is a lightweight method that, for very little cost and effort, can provide someone like the author who started with minimal usability experience a way to uncover flaws in the web site.

Conceptually, the iterative approach presented here offers several advantages over a heuristics based approach. Involvement of actual users,

especially when the overall end user population is relatively small, as is the case here, potentially can generate super-users and, as a result, others who can help create a good environment for the software to succeed. In addition, going through several iterations increases the probability that many of the problems will be found. [7] Finally, involving users at the end of development is one of the things Rubin complains about with respect to people's conception of user-centered design. In this late-involvement scenario, users end up being rubber stamps to the process and are able to contribute only when the development process is near the end and when changes are more costly.

COMPARISON TEST: INTRODUCTION

Comparison tests are performed at the end of the iterative rounds to assure that the design process produced a usable product. [8] Often these tests compare the product to a benchmark or predetermined standard. The measures used can include such things as speed or a user's subjective evaluations of the product.

In this comparison test, we sought to show that the iterative process led to a web site with improved usability.

COMPARISON TEST: METHODS

In this section, I describe the methods used to evaluate the improvement in the prototype after the iterative design.

Study design

The goal of this section is to determine whether the changes from the iterative design portion had any impact on user satisfaction or human performance. This will be measured by using time to complete tasks and questionnaire. Error rate was not available because the database where the results were stored is no longer available.

The study was set up as a 1-by-14 within-subjects design. The within variables were which version of the web site was used. The subjects were also measured on time for each of the tasks. At the end of each version of the web site, they completed a questionnaire to evaluate their perceptions of the web site's usefulness.

The null hypothesis for the analysis of the questionnaire is that there is no difference in user satisfaction between the different versions. The null hypothesis for the time analysis is that there is no difference in the time it takes to complete tasks between the different versions.

Sample

In the initial trial for the comparison test, the general surgery residency class was recruited during one of their didactic sessions. Unfortunately, a fatal database error as well as a number of server performance issues occurred during the test, preventing subjects from completing their trials. The comments section was still mostly positive; however, the collected data was discarded. It was decided that the current population would be biased and would not be used for future studies with this web site.

The next best option was to recruit medical students. The subjects were third and fourth year medical students, chosen because at this stage they had already had some clinical experience. They were recruited by email. To compensate them for their time, they were offered either a $10 bookstore gift card or a gift certificate to a local pizzeria.

After several rounds of recruitment efforts, a total of seven volunteers were recruited. Five of the subjects were women and two were men. All had had experience using a web browser and only one had past programming experience.

Setting

The study was performed in a computer lab where the subjects were free from distractions.

Instrumentation

Time was measured by timestamp. The timer started when the subject opened a page to enter data (e.g., the ventilator section). The time ended when the subject selected the complete button. User experience was captured by questionnaire. The questionnaire used was the Questionnaire for User Interaction Satisfaction (QUIS) version 7.0. The QUIS is a questionnaire that has been validated for internal consistency, reliability and validity. [39] Modifications were made to the questionnaire as allowed by the instructions; a facsimile is included in Appendix E. Two items were removed from the "Overall reaction to the software"; one item was left off of the "Screen section"; four items were left off "Terminology and system information"; four items were left off of "Learning"; and "System Capabilities" was not included. These items were left out because of lack of relevance to the project or because they were redundant.

At the end of the study, the participants were asked to fill out a form to determine which of the two versions they preferred and were provided an opportunity to contribute ideas for improving the site. A paper form was also handed to each participant asking which version they preferred and allowing significant space for free text/drawings (Appendix D).

Procedure

The subjects consented and then were randomized to start with one of three tasks (Add new, Edit or Delete). All subjects completed eight add new,

four edit and two delete tasks using a version of the intranet site. After completing the tasks, they crossed over into the other version. For example, a person who started with eight add new, four edit and two delete tasks in the prototype version would follow with eight add new, four edit and two delete tasks in the final version (Table 2). The pilot study in the iterative phase determined the number of times each task was performed.

A separate web site was created as a framework for the prototype and the final version of the website. This web site allowed us to insert questionnaires as well as time stamp the beginning and end of each task. At the end of each section, a computer-based questionnaire asked the subjects for their impressions about the application immediately after they have had a chance to test it.

After the questionnaire, subjects completed the next set of tasks in the same order as the first round, except they used the alternate application (prototype or final version) (Table 2). Like the first round, the second round also ended with the same computer-based questionnaire. At the end of the trial, they were given a summary sheet that asked them for overall comments and an opinion of which version they preferred to use.

Data analysis

The data were analyzed for two separate dependent variables: user interaction satisfaction and time to complete types of tasks for the two interfaces. A repeated measures ANOVA was performed for the QUIS data

Table 2. Sample of tasks completed by the subject.

	Case #	Task
Prototype version	1	Edit
	2	Edit
	3	Edit
	4	Edit
	5	New
	6	New
	7	New
	8	New
	9	New
	10	New
	11	New
	12	New
	13	Delete
	14	Delete
Final version	15	Edit
	16	Edit
	17	Edit
	18	Edit
	19	New
	20	New
	21	New
	22	New
	23	New
	24	New
	25	New
	26	New
	27	Delete
	28	Delete

and the time study. A p value of <0.05 was used for both the questionnaire and the time analysis.

Subscale analysis for each of the sections of the QUIS data (i.e. Overall user reactions, Screen, Terminology and system information, and Learning) was also performed to account for the increased probability of finding an effect because of repeated testing [40]. This type of error occurs when one performs repeated measurements on a subject. Although an alpha of 0.05 is typically adequate for finding significance, if the measurements are repeated

there is an increased likelihood of creating a type I error (Rejecting the null hypothesis when it is true).

A repeated measures ANOVA was also run to compare the time taken to complete each task type. This was performed for the mean time per task (Add, Edit, Delete). Content analysis was performed on the comments derived from the comments section from the subject's final form (Appendix D).

COMPARISON TEST: RESULTS

Sample

Data for tasks #14 and #28 are missing for all subjects. The original database with the stored data was no longer available, so the analysis had to be done without these data points. Tasks 14 and 28 were matching pairs for the last task for the prototype and the final version.

On examination of the sample it was noted that Subject 2 for Case 3 was noted to have taken 206 seconds to complete an "edit" task. Because this data point was an outlier, we imputed the average for the other subjects and used that data point instead. The average of all subjects completing the 3rd task was 51.43. We substituted the 206 seconds with 51.43.

QUIS data

Analysis of the QUIS scores reveals a mean score of 6.39 (SD = 0.85) for the prototype and a mean score of 6.45 (SD = 1.47) for the final version. Comparisons of the means using repeated measures ANOVA results yielded an $F(1,6) = 0.13$, $p=0.912$. Thus, we accept the null hypothesis (i.e. there is no difference in user satisfaction between the two versions). The results are displayed in Table 3. Of note, all of the mean scores in the final version are higher than in the prototype. Thus, there was no statistically significant

Table 3. Results of subscale analysis for the QUIS questionnaire

	Prototype			Final		
	N	Mean	SD	N	Mean	SD
Overall Subscale	7	5.39	0.99	7	5.57	1.31
Layout Subscale	7	5.90	1.49	7	6.14	1.71
Terminology	4	7.38	2.14	6	7.75	1.94
Learning Subscale	7	7.50	0.50	7	6.86	1.03
Average Score	7	6.39	0.85	7	6.45	1.47

difference of improved user satisfaction between the prototype and the final version.

Time trial: Prototype interface vs final interface

Comparing the time to complete tasks for the prototype interface vs the final yields an $F(1,5)=6.17$, $p = 0.056$, just outside our level of significance. In this instance, I have to accept the null hypothesis. There is no difference in the two sites.

Figure 4 shows the graph for a repeated measures ANOVA looking at time taken per task vs whether the tasks were completed first (Task 1 to 13) or second (Task 15 to 27). In spite of the fact that there was no statistical difference in whether the task occurred in the first or second interface, the lines intersect suggesting an interaction effect.

Time trial: Prototype vs final version

Comparing the time to complete tasks values for the prototype vs the final version yields $F(1,5) = 0.146$, $p = 0.718$ (Main) and the interaction effect

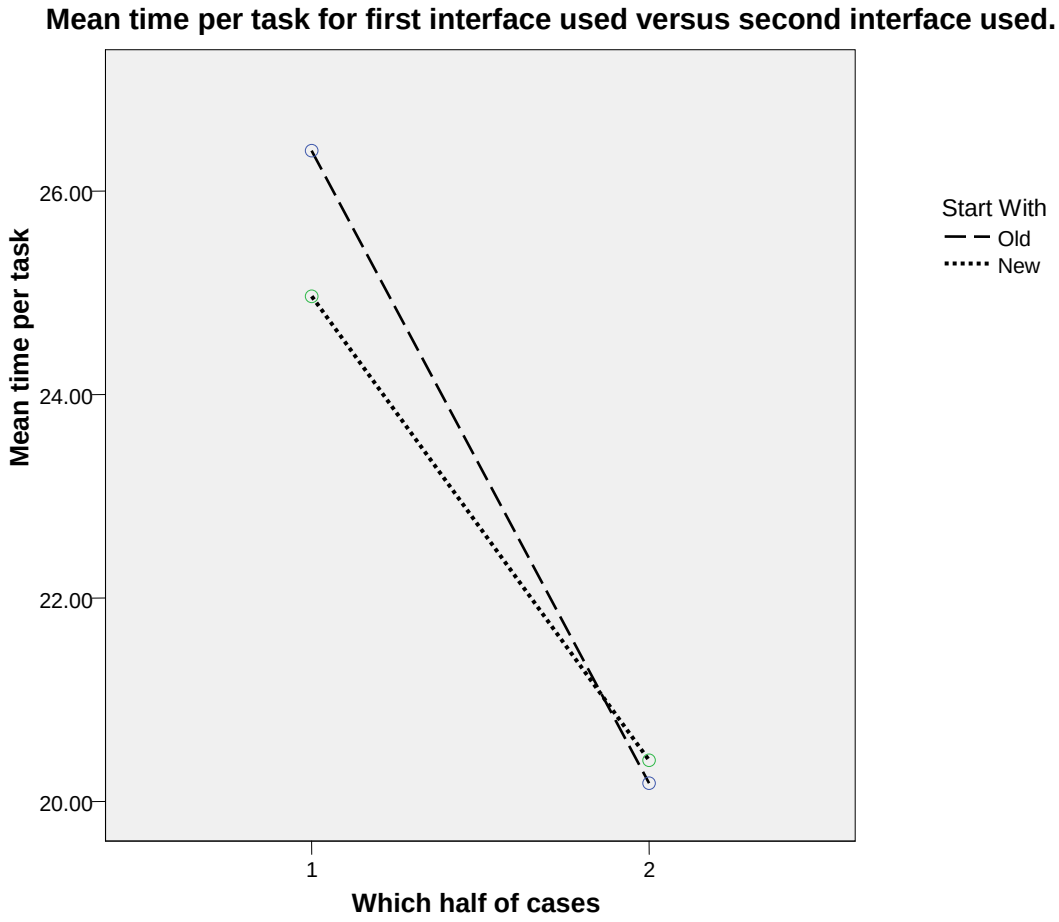

Figure 4. Mean time per task for first interface versus second interface (1= 1st interface task (task 1-13), 2= 2nd interface task (task 15-27)).

yields an $F(1,5) = 6.172$, $p = 0.056$. The interaction was just outside our level of significance. While statistically nonsignificant, the trends in the data suggest that the first interface used was slower than the second, regardless of which interface was used first.

Time trial: Tasks

The times were also compared within the individual task categories (Add new, Edit and Delete) regardless of whether the prototype or final

version was used. "Add new" tasks and "Delete" tasks showed no statistical significance when comparing the prototype vs the final version ("Add new" Main $F(1,5)= 0.267$, $p = 0.628$, Interaction $F(1,5)= 1.09$, $p = .345$; "Delete" Main $F(1,5)= 0.949$, $p = 0.375$, Interaction $F(1,5)= 0.895$, $p = 0.388$). However, the edit tasks did show significance (Main $F(1,5)= 14.3$, $p = 0.013$ Interaction $F(1,5)= 55.2$, $p = 0.001$) (Table 4).

Content analysis

Content analysis was performed on the comments by the subjects. The subjects' comments relating to the usability of the web site are summarized in Figure 5. Four of the seven subjects preferred the revised site because of improvements made to the summary page [33]. Four of the seven subjects commented on how both sites were easy to use. Three of the seven noted that they liked the previous and next buttons. These are juxtaposed with the usability problems found in the iterative phase (Figure 6).

Table 4. Time in seconds for "Edit" task completion.

	Old GUI	New GUI
Start with old first	49.0	32.8
Start with new first	41.2	46.5

1. Delete should include patient name to confirm that the correct patient is being deleted
2. More prominent patient names
3. Font should be larger
4. Easy to pick the wrong pick list item because of the small type

Figure 5. Usability issues discovered in comparative testing

1. Add a previous and next button
2. Pop up to indicate patient is already in the log
3. Graphical representation of summary page
4. Improve location of "Add new patient"
5. Confirm a category (e.g., ventilator) is complete before allowing someone to leave the screen
6. Hourglass icon when loading data
7. Summary page should include the original RRC recommendations
8. Summary page should tell user how many more patients they have left to complete the log
9. Icons for summary page (checkbox for complete, question mark for incomplete)
10. Pop up warning when deleting a patient
11. Improvements in pick list choices
12. Opportunity to print list of subjects

Figure 6. Usability issues discovered in iterative design

COMPARISON TEST: DISCUSSION

This section discusses sample data, the results of the QUIS data, the results from the variety of time analysis and content analysis of the comments from the subjects.

Impact of missing data

The data were missing tasks #14 and #28, and the original database was also not available. The lost data occurred at the end of each version, so the matching data was lost. Furthermore, the data loss occurred at the end of the series of a specific task but not consistent tasks. The lost data would have had the most impact on Edit tasks as four of the seven subjects ended with Edit type tasks. The other three ended with Delete tasks where, because of the nature of the task, it is unlikely to have much of an impact. As the last trial tends to be where the learning effect had been fully accounted for, it is likely that, had the missing tasks been included, the significance would have been improved.

Data discrepancies

It was noted that one of the time values for Subject 2 Task 3 was significantly slower than others in the same category (206 seconds compared to an average of 51.43 seconds to complete the third task). The exact reason

for this is unknown, but speculation suggests either subject inattention or the need to adjust to the new task.

User interaction satisfaction

The results did not show a statistical difference in user satisfaction between the two versions. Several possibilities exist that may explain the results.

One possibility is that the results are valid. Of the seven subjects, four noted that the two versions were fairly similar. Based on discussions with the surgical residents and the responses on many of the comment forms, I suspect that there was improvement in the web site; but if the statistical analysis is so close, one would have to consider whether the value obtained was worth the time and effort of a full usability test.

Another possibility is that the comparison test had inadequate power to detect differences in user satisfaction. This is suggested by the consistently higher mean in the subset scores for the final version as well as by noting the subjective comments. A within-subjects design was used for the comparison test for several reasons. A within-subjects design offers better power than a traditional between-subjects design. [41] Because each subject acts as his or her own control group, there are in essence twice as many subjects as in a between-subject study with the same number of subjects. The advantages of this design include the adequacy of smaller sample sizes and control for individual differences. Another is a reduction in error

variance. In studies involving smaller numbers of subjects random variations will have large effects. In a within-groups design, the same subject will be participating in all of the trials, so those variations are accounted for. However, for relatively similar user interfaces, larger numbers are needed.

Another factor that may have contributed to the lower power is the missing data points. When asked which version the subjects preferred, they unanimously chose the final version. There were many common usability features that a usability expert would likely consider major flaws in the prototype version. For example, the lack of a warning message when a user attempts to delete a patient. The gratitude for such a feature can only truly be appreciated by someone who has nearly accidentally deleted the patient from their record. Similarly, if one was entering data in a category and forgot to fill in one of the required fields, one would be more appreciative of the ability of this application to check whether a category was complete. In addition, the tasks to be performed were relatively simple. With such a simple workflow, detecting small changes would only be possible with a larger study population.

A third factor concerns task complexity. In order to avoid confounding factors, such as typing ability, the tasks were set up to modify values on the pick lists. As a result, the tasks were relatively simple. As such, they may have underestimated the time savings the application could have achieved.

More complex tasks can potentially lead to improvements in both time improvements as well as user satisfaction scores.

Two tasks, Delete and Add, were unlikely to have shown an effect in time. The Delete task is a fairly straightforward task (e.g., click on record and delete), and time saving on this task is unlikely. This is also suggested by the short amount of time it currently takes to complete this task. The Add task is more difficult than Delete; however in contrast to an Edit task, the fields are blank to begin with. Cognitively, this may prove to be just enough of a hurdle to prevent benefit from being seen in adding new tasks. Further studies will have to be done to clarify this point.

Fewer confirmatory study comments

The fact that so many of the user interface errors were captured by the iterative study, compared to the confirmatory study, lends credence to Virzi's study [6] and reaffirms our assumptions about using three subjects in the iterative section. The fact that significant usability issues were found in the confirmatory study, however, also validates the importance of performing this final step. [8]

Heuristics

In Yao's study [22], Yao had access to usability experts. Although the best method of developing a web site would involve the users at the beginning of the development, a benefit of using a heuristics-based-expert approach is markedly decreased development time. In her study, Yao was able to

complete the evaluation in one month. In stark contrast, the FEMA web site design was a year-long process. Because of the difficulty in scheduling the residents, our study took approximately six months to perform. Many times the study was delayed because the subjects had rotations in outside hospitals and wouldn't have been able to travel to the study site. Notwithstanding the fact that we were using residents, delays for other reasons can occur and are valid threats to be considered when committing to an iterative process.

User involvement vs. usability expert

Many previously discussed articles [22-23] have used expert opinion to the exclusion of iterative user feedback. This seems to go against Gould and Lewis' original precept of early involvement of users. [7] Furthermore, Nielson argues that "[e]valuators are probably especially likely to overlook usability problems if the system is highly domain-dependent and they have little domain expertise." [42] This is a common problem in medicine.

Involving usability experts following heuristics has its place. Nielson [42] notes that usability experts are able to complement the users' observations by finding the issues that they may not notice or are unable to verbalize. Many of the studies that engaged usability experts were able to complete the design phase in a much shorter time than if the design were done in true iterative fashion. [22, 28] Consequently, Nielson recommends the use of an initial expert-based evaluation followed by an iterative user

evaluation. [42] This approach is likely to be very resource intensive, and I have been unable to find any research that validates this approach.

Performance time

The time analysis comparing the prototype and the final version was revealing in spite of the lack of statistical significance. The implication from Figure 4 is that there is an interaction going from the prototype to the final version of the web site. This implies that if you started with the final web site, you may require less time to complete tasks. This did not reach statistical significance, but it was very close to doing so ($p=0.056$). A similar finding occurred with the interaction effect comparing the prototype with the final version of the web site ($p=0.056$). All tasks in the 2nd interface were faster than the 1st interface as one would expect due to the learning effect.

Edit tasks were significantly different. (Main $p=0.013$ and interaction $p=0.001$). In this section subjects were able to "edit" more rapidly with the final version than with the prototype. One explanation is that the subjects may have benefited from the improvements in navigation. As the subjects learned how to use the system better using the previous and next buttons, either they were able to navigate to the category in question more rapidly or they learned how to use the system faster. Another explanation is that differences in usability become more evident with more complex tasks such as editing.

This study shows that the usability techniques employed in this study can be used to develop a usable intranet web site. The iterative design process generated important improvements to the web site as well as pointing out usability defects that would be helpful to usability novices. The difficulties encountered during this study illustrate the importance of good planning and thoughtful study design. Allying with multiple stakeholders should be encouraged to decrease risk of losing a single stakeholder and to help in securing resources to complete the study. This study illustrates the importance of usability and the need for further usability research. Biomedical Informatics should adopt these measures both to gain more experience in usability techniques for future research and to disseminate this information to trainees so it may be used in the field.

For residency program directors and administrative personnel, the above methods show ways to develop web sites that will track and compile data that are required by regulatory agencies. The methodologies can be reproduced with minimal training and cost.

Study limitations

The study was limited by not using surgical residents as test subjects, poor enrollment of medical students, lost data points, the use of a stopwatch and the use of email to recruit subjects. Ideally this study would have used surgical residents for the comparison test. Many of them had been exposed to the web site while using a defective server, so we considered them to be a

biased population. As these residents are replaced by other unbiased residents, the next residency class can be considered as future study subjects. The study was also limited by poor enrollment of medical students. It appears that larger more significant incentives may be necessary to attract the larger numbers necessary for the confirmatory study.

The pilot study achieved its original goal of determining the number of trials before the learning effect occurred. However, a trial run of the study might have averted many of the problems we ran into during the trial involving subjects.

The loss of data was also troubling and raises questions about data integrity. The systematic loss of data for trials 14 and 28 for all subjects suggests a systematic error perhaps as a result of programming defect. In one of the trials, because of data irregularities, we also had to use an average value for one of the values in the time to complete tasks data.

There were methodological issues as well. Our team used a stopwatch to measure time. Other devices allow for more accurate time capture but cost more. In comparing the two versions, there is also risk in having the tasks in the same order because of the possibility that the subjects may not be as attentive in the second round. Medical students were used as the subject population, and it is likely that they are not representative of the general surgery group.

The study could have been biased because the author completed the timing. Future studies should involve use of software designed to capture time spent on tasks or the use of time stamps as was used in the confirmatory study.

Finally, it should be noted that the selection of the confirmatory population was done by email. This may bias the population towards those who use email; however, given the prevalence of email and the fact that the school of medicine communicates to all of its students in this fashion, the bias created by this form of recruitment is negligible. Furthermore, there is data to suggest that at the 3rd and 4th year medical student level, this is of minimal significance [43-44]. I suspect, as technology becomes universal amongst this population, this will become less of an issue in future studies.

Unfortunately, this project suffered from inadequate medical student interest. Greater incentives are likely to increase medical student interest and discovery of other ways to fund this research would make reimbursement decisions easier. Furthermore, if higher level department staff were involved, they might also be able to facilitate recruitment of medical students via personal appeals for volunteers.

A key committee member was lost during the conduct of this study. This loss certainly made tasks more difficult. Without his advocacy, it was impossible to implement this project in a live environment and determine how the application performed in a real world setting. He also would have

had access to resources that would have facilitated the recruitment of medical students. Future efforts may benefit from broadening the number of people involved, especially including those in decision making positions, to decrease the risk of losing a sole stakeholder.

Future research

The results of this study leave many avenues for further research. Nielson's approach of using an initial expert-based evaluation followed by an iterative user evaluation has merit. A comparison study would be useful to determine how much additional benefit the combined expert review and iterative design delivers as opposed to each process individually. Repeating the study with more subjects would give the study the power to detect more subtle differences in the designs.

It would also be interesting to see whether iterative design has produced other statistical failures in other web sites. Some studies reviewed [21, 28] have not subjected their sites to such rigorous statistical analysis. In addition, studies that do not show statistical improvement might not be published, resulting in publication bias.

As discussed previously, we found that our subjects learned to edit faster using the final version. It would be interesting to know whether the previous and next buttons made users faster because it helped the users become more familiar with the application via scaffolding as an approach to

learning or whether these buttons were just a better navigation feature than the left handed navigation.

Another question that needs to be addressed is whether resident input is needed at all. The application could be redesigned so that the data is abstracted from an electronic health record by queries to a data warehouse. This modification would minimize the need for data entry, although the project would not be absolved of the need to incorporate user-centered design principles. Instead, the user-centered design would focus more on data interpretation and editing.

There also has been discussion about the need to rely on surgery residents to enter the data. If one queries the database, perhaps out of a data warehouse, one can find the data so certain fields can be populated with data that had already been entered. This will trim down the residents' role to that of fact checker. As these types of functions are required in many academic facilities, these may prove to be an offshoot of the EHR.

It is difficult to reproduce exactly the effect the real world has on subjects, so testing in a live environment should also be performed. Such observations often lead to further insights on how to improve the product and in usability in general.

CONCLUSION

Authors [10, 45-46] suggest that poor usability can contribute toward medical errors. The methodology described above offers healthcare organizations using an intranet the opportunity to improve the usability of their sites at minimal cost. Furthermore, many residency programs require various methods of tracking procedures, experiences and other items required by the RRC.

Inadequate consideration of user-centered design can decrease the use of applications, but good, inexpensive tools to improve usability exist and have been demonstrated to improve usability. [13-16] This study demonstrates the successful use of several of these tools. While performing this study we learned the importance of involving the end user in the development process, as well as the importance of incorporating good navigation and GUIs in the application. We were successful in developing a usable web site without having formal usability training or using more expensive usability experts.

APPENDIX A

REQUIREMENTS FOR CRITICAL CARE

INDEX CASE LOG [47]

Essentials in Critical Care Management

Select the patients who best represent all the essential aspects of intensive care unit management. Each resident is to develop a Critical Care Index Case (CCIC) log of at least 20 patients who best represent the full breadth of critical care management. At least two out of the seven categories listed below should be applicable to each chosen patient. The completed CCIC log should include experience, with at least one patient, in all seven of the following essential categories:

1. Ventilatory Management
    a. Etiology/Indications
    b. Ventilatory modes/techniques
    c. Long term vs short term intubation (days on the ventilator)
    d. Weaning method
2. Bleeding (non-trauma) greater than 3 units necessitating transfusion/monitoring in ICU setting
    a. Etiology
    b. Coagulopathy: Yes No
    c. Hypothermia: Yes No
    d. Autotransfusion: Yes No

3. Hemodynamic Instability

    a. Etiology
    b. Volume resuscitation
    c. Inotropic/pressure support: Yes No
    d. Mechanical assistance of cardiac failure Yes No

4. Organ Dysfunction/Failure (etiology/mode of management)

    a. Renal
    b. Hepatic
    c. Central Nervous System
    d. Endocrine
        i. Hypothyroidism
        ii. Adrenal insufficiency
        iii. Panhypopituitarism
        iv. Diabetes Insipidus
        v. SIADH

5. Dysrhythmias

    a. Etiology
    b. Drug Management
    c. Therapeutic interventions
    d. Monitoring

6. Invasive Line Management/Monitoring

    a. Arterial cannulation
    b. Pulmonary artery catheter
    c. Physiologic profile – directed management
    d. Complications

7. Nutrition

    a. Route (parenteral/enteral)
    b. Indications/contraindications
    c. Solution formulation
    d. Complication

APPENDIX B

NIELSON'S 10 HEURISTICS [24]

1. Visibility of system status	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2. Match between system and the real world	The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.
3. User control and freedom	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4. Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. Error prevention	Even better than good error messages is a careful design which prevents a problem from occurring in the first place.
6. Recognition rather than recall	Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. Flexibility and efficiency of use	Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
8. Aesthetic and minimalist design	Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.
9. Help users recognize, diagnose, and recover from errors	Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
10. Help and documentation	Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

APPENDIX C

INSTRUCTIONS TO PARTICIPANTS

Iterative Design

Study Protocol

Location: Richards Library (if available). Medical Informatics will be backup

The three residents who have already been selected will be given the "Introduction for Study Participants" and asked if they have any questions. After any questions have been answered they will be given their tasks in their test packets. As they perform these tasks they will announce which task they are performing. At the end of the tasks a debriefing interview will be conducted where the following questions will be asked:

- What did you like about the web log?
- Would you change anything in the web log?
- Did you have difficulty finding any of the categories (eg Endocrine, etc)?
- Any other changes or suggestions?

Other questions will be asked depending on the comments or actions taken by the resident during the testing.

APPENDIX D

EVALUATION FORM POST COMPARATIVE TEST

Participant #_____

Which web site did you prefer?

First___ Last___

Final Comments:

_____

_____

_____

_____

_____

_____

_____

_____

APPENDIX E

QUIS QUESTIONNAIRE [39]

OVERALL USER REACTIONS

Overall reaction to the system	Terrible								Wonderful	
	1	2	3	4	5	6	7	8	9	
	Satisfying								Frustrating	
	1	2	3	4	5	6	7	8	9	
	Difficult								Easy	
	1	2	3	4	5	6	7	8	9	
	Flexible								Rigid	
	1	2	3	4	5	6	7	8	9	

SCREEN

Layout of summary page	Helpful								Unhelpful	
	1	2	3	4	5	6	7	8	9	
Navigation between sections	Difficult								Easy	
	1	2	3	4	5	6	7	8	9	
Selection of menu choices	Inadequate								Adequate	
	1	2	3	4	5	6	7	8	9	

TERMINOLOGY AND SYSTEM INFORMATION

Messages which appear on screen	Confusing								Clear	
	1	2	3	4	5	6	7	8	9	NA
Phrasing of error messages	Pleasant								Unpleasant	
	1	2	3	4	5	6	7	8	9	NA

LEARNING

Learning to operate the system	Difficult								Easy	
	1	2	3	4	5	6	7	8	9	NA

LEARNING

Learning to operate the system	Difficult									Easy
	1	2	3	4	5	6	7	8	9	NA
Tasks that can be performed in a straight forward manner	Always									Never
	1	2	3	4	5	6	7	8	9	NA

REFERENCES

1. Fischer JF. *The impending disappearance of the general surgeon.* JAMA. 2007; 298(18): 2191-3.

2. ACGME. *Requirements for a Critical Care Log for General Surgury Residents*. The American College of Graduate Medical Education Web Site. http://www.acgme.org/acWebsite/RRC_440/440_SurgCCIG.asp. Accessed November 14, 2009.

3. Winslow EH, Nestor VA, Davidoff SK, Thompson P, Borum JC. *Legibility and completeness of physician's handwritten administration orders.* Heart Lung. 1997; 26(2): 158-64.

4. Holman J. *Personal Communications from Dr Holman*. 2003.

5. Katz-Haas R. *User centered design*. 1998; http://www.stcsig.org/usability/topics/articles/ucd%20_web_devel.html. Accessed November 14, 2009.

6. Virzi RA. *Refining the test phase of usability evaluation: How many subjects is enough?* Human factors and ergonomics society annual meeting proceedings. Safety. 1992; 34(4): 457-68.

7. Gould J, Lewis C. *Designing for usability: Key principles and what designers think.* Communications of the ACM. 1985; 28(3): 300-11.

8. Rubin J. *Handbook of usability testing*. New York. John Wiley and Sons; 1994

9. Stead WW, Lin HS, eds. *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions*. Washington DC. The National Academies Press; 2009.

10. Han Y, Carcillo JA, Venkataraman ST et al. *Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system.* Pediatrics, 2005; 116(6): 1506-12.

11. Coiera E, Westbrook J, Wyatt J. *The safety and quality of decision support systems.* Yearb Med Inform. 2006; 20-5.

12. Ash JS, Sittig DF, Seshadri V, Dykstra RH, Carpenter JD, Stavri PZ. *A cross-site qualitative study of physician order entry.* J Am Med Inform Assoc. 2003; 10(2): 188-200.

13. Johnson CM, Johnson TR, Zhang J. *A user-centered framework for redesigning health care interfaces.* J Biomed Inform. 2005; 38(1): 75-87.

14. Boyington AR, Wildemuth BM, Dougherty MC, Hall EP. *Development of a computer-based system for continence health promotion.* Nurs Outlook. 2004; 52(5): 241-7.

15. Taylor DP, Bray, BE, Staggers N, Olsen RJ. *User-centered development of a Web-based preschool vision screening tool.* AMIA Annu Symp Proc. 2003; 654-8.

16. Wachter SB, Agutter J, Syroid N, Drews F, Weinger MB, Westenkow D. *The employment of an iterative design process to develop a pulmonary graphical display.* J Am Med Inform Assoc. 2003; 10(4): 363-72.

17. Atack L, Luke R, Chien E. *Evaluation of patient satisfaction with tailored online patient education information.* Comput Inform Nurs. 2008; 26(5): 258-64.

18. Nahm ES, Preece J, Resnick B, Mills ME. *Usability of health Web sites for older adults: a preliminary study.* Comput Inform Nurs. 2004; 22(6): 326-34.

19. Nielson J. *Discount usability engineering*, 1994. http://www.useit.com/papers/guerrilla_hci.html. Accessed July 11, 2009.

20. *The new citizen centric, user-friendly FEMA Website*. http://www.fema.gov/media/site_case_study.shtm#5. Accessed July 29, 2009.

21. Atkinson NL, Saperstein SL, Desmond SM et al. *Rural eHealth nutrition education for limited-income families: an iterative and user-centered design approach.* J Med Internet Res, 2009; 11(2): e21. Accessed July 29, 2009.

22. Yao P, Gorman P. *Discount usability engineering applied to an interface for web-based medical knowledge resources.* Proc AMIA Symp, 2000; 928-32.

23. Allen M,Currie LM, Bakken S, Patel VL, Cimino JJ. *Heuristic evaluation of paper-based Web pages: a simplified inspection usability methodology.* J Biomed Inform, 2006; 39(4): 412-23.

24. Nielson J. *Ten Usability Heuristics*. 2005; http://www.useit.com/papers/heuristic/heuristic_list.html. Accessed November 22, 2009.

25. Wong R. *8 Golden Rules of Interface Design.* 2008; Available from: http://www.84bytes.com/2008/08/20/8-golden-rules-of-interface-design/. Accessed November 22, 2009.

26. Staggers N, Kobus D, Brown C. *Nurses' evaluations of a novel design for an electronic medication administration record.* Comput Inform Nurs, 2007; 25(2): 67-75.

27. Kushniruk A, Patel V, Cimino J. *Usability testing in medical informatics: Cognative approaches to evaluation of information systems and user interfaces.* Journal of Biomedical Informatics, 2004; 37(1): 56-76.

28. Lai TY, *Iterative refinement of a tailored system for self-care management of depressive symptoms in people living with HIV/AIDS through heuristic evaluation and end user testing.* Int J Med Inform, 2007; 76 Suppl 2: S317-24.

29. Nielson J. *Intranet Usability Shows Huge Advances*. 2007; http://www.useit.com/alertbox/intranet-usability.html. Accessed September 19, 2009.

30. Kushniruk A. *Evaluation in the design of health information systems: application of approaches emerging from usability engineering.* Comput Biol Med, 2002; 32(3): 141-9.

31. Lewis C, Rieman J. *Task-centered user interface design*. 1994; http://grouplab.cpsc.ucalgary.ca/saul/hci_topics/tcsd-book/chap-5_v-1.txt. Accessed June 14, 2009.

32. *Content analysis*. UsabilityNet website. 2006; http://www.usabilitynet.org/tools/r_content.htm. Accessed July 11, 2009.

33. Leung E. Intranet Usability. 2010; https://sites.google.com/site/intranetusability/file-cabinet. Accessed October 23, 2010.

34. Ritter F, Schooler L. *The Learning Curve*, in *In International Encyclopedia of the Social and Behavioral Sciences.* Pergamon: Amsterdam. 2002: 8602-5.

35. Coulton DA. *Relaxing the homogeneity assumption in usability testing.* Behavior & Information Technology 2001; 20(1): 1-7.

36. Piaget J. *Genetic epistemiology*. 1968; http://www.marxists.org/reference/subject/philosophy/works/fr/piaget.htm. Accessed July 11, 2009.

37. Wood D, Bruner JS, Ross G. *The role of tutoring in problem solving.* J Child Psychol Psychiatry, 1976; 17(2): 89-100.

38. Staggers N, Kobus D. *Comparing Response Time, Errors, and Satisfaction Between Text-based and Graphical User Interfaces During Nursing Order Tasks.* J Am Med Inform Assoc, 2000; **7**: 164-176.

39. Harper B, Slaughter L, Norman K. *Questionnaire administration via the WWW: a validity and reliability study for a user satisfaction questionnaire*. In: WebNet 97. Nov 1-5,1997; Toronto, Canada. http://www.lap.umd.edu/webnet/paper.html. Accessed July 11, 2009.

40. Zar JH. *Biostatistical Analysis*. 1999, Upper Saddle River: Simon & Schuster.

41. Hall R. *Within subjects design*. 1998; http://web.mst.edu/~psyworld/within_subjects.htm. Accessed July 11, 2009.

42. Nielson J. *Characteristics of Usability Problems Found by Heuristic Evaluation*. http://www.useit.com/papers/heuristic/usability_problems.html. Accessed July 11, 2009.

43. Dorup J. *Experience and attitudes towards information technology among first-year medical students in Denmark: longitudinal questionnaire survey.* J Med Internet Res, 2004; 6(1): e10.

44. Link TM, Marz R. *Computer literacy and attitudes towards e-learning among first year medical students.* BMC Med Educ, 2006; 6: 34.

45. Connoly C. Doctors cling to pen and paper. *The Washington Post*, Mar 25, 2009:A01.

46. Koppel R, Metlay JP, Cohen A et al., *Role of computerized physician order entry in facilitating medication errors.* JAMA, 2005; 293: 1197-1203.

47. *General Surgury Residency Program*. University of Utah School of Medicine website. http://utahhealthsciences.net/pageview.aspx?id=16897. Accessed August 6, 2009.