Iterative Usability Testing: Ensuring a Usable Clinical Workstation

Janette M. Coble, MS[1], John Karat, PhD[3], Matthew J. Orland, MD[1,2], Michael G. Kahn, MD, PhD[1,2]
[1]Section of Medical Informatics, Washington University School of Medicine
[2]BJC Health System, St. Louis, Missouri
[3]IBM T. J. Watson Research, Hawthorne, New York

Once the users' needs are determined, how does one ensure that the resulting software meets the users' needs? This paper describes our application of a process, usability testing, that is used to measure the usability of systems as well as guide modifications to address usability problems. Usability testing is not a method to elicit opinions about software, but rather a method to determine scientifically a product's level of usability. Our application of usability testing is designed to determine the current usability level of a workstation designed for the clinician's use, determine specific problems with the Clinical Workstation's usability, and then evaluate the effectiveness of changes that address those problems.

BACKGROUND

Project Spectrum is a joint technology consortium consisting of Washington University School of Medicine, BJC Health System, IBM, Kodak, and SBC Corp.[1] The purpose of Project Spectrum is to provide users with comprehensive, longitudinal clinical information across all 15 hospitals in the BJC Health System. The target user for Phase I is a clinical physician in the field of general medicine or general surgery, including academic and community physicians.

Due to past (less than successful) experiences with introducing information systems for physicians into the BJC Health System, it was believed and emphasized that the resulting Clinical Workstation (CW) must truly meet the needs of the physicians in a highly usable manner. To ensure that this outcome would be the case, we knew we needed to start with, and focus on, the physicians. Contextual Inquiry,[2,3,4] a formal approach for analyzing the physicians' needs, was used resulting in a user requirements document with physician priority ratings which have been used to assist in guiding the direction of the CW.[5,6]

Although the user requirements identify the users' needs for the system, the requirements tend to be under-specified in meeting the developers' needs. User requirements describe what the user needs, but not how to meet those needs. To define the functions which were thought necessary to satisfy the user requirements, a problem analysis activity was undertaken resulting in a functional specification document. The developers are now implementing the CW from these functional specifications.

What can be done to ensure the resulting system will meet the needs of the users when the developers are using the functional specifications, not the user requirements?[7] This paper discusses one tool we have used, usability testing, to ensure that the resulting system continues to meet the needs of the users in a usable manner. Similar approaches at varying levels of formality have been used by others successfully.[8,9] On our project, we have employed a version of usability testing that has enabled us to determine the current level of the CW's usability, determine specific problems with the CW, and then evaluate the effectiveness of changes that address those problems.

USABILITY TESTING

Usability has been defined as the extent to which a system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. Usability is not just subjective, since effectiveness, efficiency, and satisfaction can be measured through a number of methods. Additionally, usability is not just a function of the technology in a system. To measure usability one needs to involve users who carry out real world tasks to achieve user-identified goals. Advances in the understanding of what usability is and how to improve this quality in the systems being built has led to the development of many techniques, collectively called usability testing. Usability testing measures the usability of systems and guides modifications to these systems based on careful interpretation of usability test results.[10]

Usability testing has its roots in formal experimental methods. Usability tests can range in formality from a classical experiment with large sample sizes and complex test designs to very informal qualitative studies with only one participant. The objectives of the various testing approaches differ, as do resource and time requirements.

To develop a usable CW, our project applied a version of usability testing that was rapid but allowed some measure of statistics and continuity of results across usability tests. It was felt that there was not enough time and resources to use the classic experimental method. However, a process void of statistical results also was not desired. Additionally, due to the nature of our project with the developer in a different geographic location than the customer and user, it was important to be able to document formally the system's current level of usability along with improvements or degradations as they occurred.

Our Approach

Taking these tradeoffs into account, a less formal approach to usability testing that incorporated both objective and subjective measures was chosen.

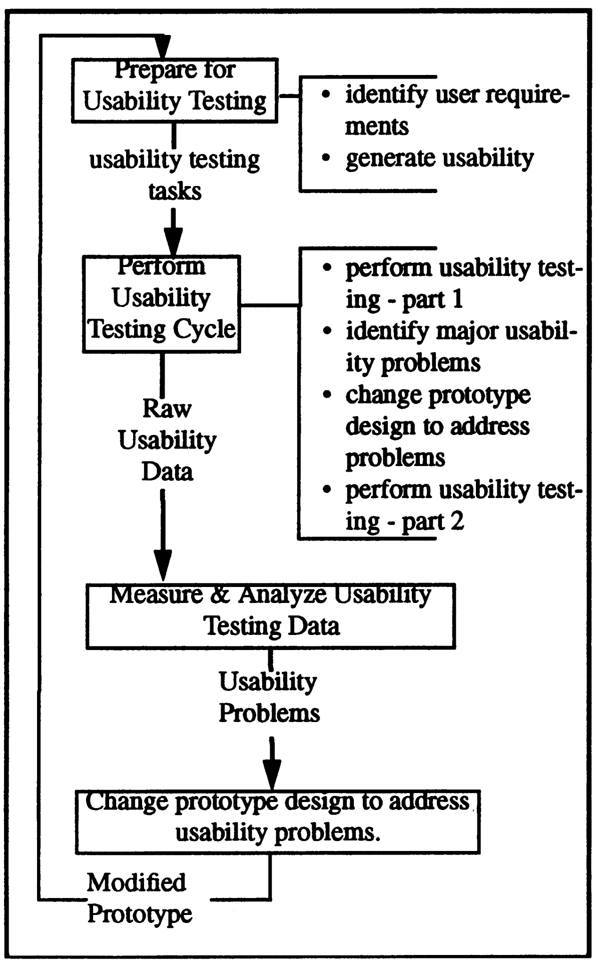

Figure 1: Usability Testing Process

With each new iteration of a prototype, a set of usability tests was performed. Our goals were to determine three things: (1) does the user understand how the system as a whole works, (2) did changes to the prototype which were meant to improve the usability of a feature actually improve the feature's usability, and (3) did the new functions added to this version of the CW obtain a satisfactory level of usability or are there problems that need to be addressed.

Our high-level process is documented in Figure (1). The process involves basically four steps: (1) prepare for usability testing by grounding the tests in the user requirements, (2) perform a usability testing cycle which includes making modifications to the prototype to quickly address the most glaring usability problems, (3) measure and analyze the usability tests identifying the usability problems, then (4) make changes to the prototype to address the identified usability problems. The cycle then repeats until a satisfactory level of usability has been achieved.

Prepare for Usability Testing. Properly preparing for the usability tests is one of the most important steps in usability testing. By grounding usability testing in the users' needs, the results of the usability test will show whether or not the users' needs are met in this iteration of the prototype and to what degree they're being met.

Jointly with the developer, the functions available in the prototype for the next usability test are identified. For each function in the prototype, the corresponding user requirements are identified. Finally, usability testing tasks are generated based on the user requirements identified.

For example, this iteration of the prototype incorporates the ability for the physicians to review Intake and Output measurements. First, the user requirements for reviewing intake and output measurements were identified. Two of the user requirements are:

- "The user must be quickly able to focus on the total intake and output for any given day."
- "The user must be able to determine cumulative or multiple day intake and output totals."

Secondly, usability testing tasks were generated directly from these user requirements:

- "Find the intake and output results. State aloud the daily net for the most recent day measured."
- "The patient had surgery on 10/3/97. Check the cumulative intake and output result since that date and state it aloud."

These tasks were then combined with other tasks to create clinical scenarios for usability testing.

The usability testing cycle. The typical cycle was:

1. Perform usability tests - part 1 (day 1)

In Project Spectrum, a usability test is one in which a number of physicians are given typical tasks and asked to use the CW prototype to accomplish the tasks. The physicians are asked to think aloud while they perform the tasks so observers can gain insight into the physicians' thought processes and identify more precisely with which features any problems occur. These sessions are video-taped to allow for additional analysis following completion of the tests.

Our usability tests were conducted with the physician and the usability specialist in one room. The observers were in a separate room viewing the usability test through two monitors. The first monitor was the output from the video camera which showed the physician and the computer screen from about 8 feet. The second monitor had a direct feed from the computer the physician was using so the observers could see exactly what the physician was doing.

During the tests the usability specialist and the observers took notes. At the end of part 1 of the usability tests these notes were consolidated either informally by scanning the notes or more formally by placing each identified problem on a post-it note and creating an affinity diagram (a logical grouping of the usability problems in a hierarchical tree organization).

2. Identify clear usability problems (day 2)

Once the notes were consolidated, the major usability problems were identified either by the number of users experiencing the same problem or by the severity of the problem.

For example, while a patient's chart was being pulled there was no indication to the user that the system was processing this request. Every physician tested thought they had done something wrong so they started clicking on other buttons to fix their "mistake." This was a case of a "major" usability problem because a large percentage of users experienced it.

Another example showed that some physicians did not understand our visual indicator that a result or measurement had been deleted, a strikethru over the measurement. This problem was not experienced by every physician but was such a severe usability problem that it was considered "major."

3. Change prototype to address problems. (day 2)

Once the major usability problems were identified, design alternatives that were thought to address the usability problems were identified. The design alternatives were prioritized based on their anticipated impact on the usability of the system and the amount of time available to make changes before part 2 of the usability tests were conducted.

4. Perform usability tests - part 2 (day 3)

Part 2 of the usability tests were conducted in the exact same manner as part 1. The same tasks were used but with the modified prototype and different physicians.

Measure and consolidate usability results. At the completion of the usability testing cycle, notes and video tapes from all the usability tests in this cycle were gathered for measurement and analysis of the usability objectives.[11] The following usability criteria were measured for each task:

- the time it took the physician to complete the task,
- the percent of the task the physician successfully completed without assistance,
- the number of "problems" the physician encountered,
- The amount of assistance the physician requested for a given task, and
- the physician's satisfaction level using the prototype to perform the given task.

The time it took the physician to complete a task did not necessarily indicate usability problems at the task level. Some physicians got more involved in the clinical aspects of the task than others. Consequently, this measurement was used only for informational purposes.

The percent of the task the physician successfully completed without assistance is a useful measure for identifying where major conceptual problems with a function are occurring. These percentages are averaged across all physicians tested in this usability testing cycle to determine a high-level success rating.

The number of "problems" the physician encounters is one of the most effective measurements determining a function's usability. It's also one of the easiest ways to track the function's progress toward an acceptable level of usability. The number of problems the physicians experienced on a given task are averaged together to give an overall rating for this task.

It is a very good indicator when physicians ask for assistance on a given task that there is need for improvement on that feature. These questions and the context with which they were asked are used to improve the feature. The questions are also used when developing help documentation to ensure the physicians are supplied with useful help documentation.

At the completion of each task, the physicians were given a questionnaire consisting of one question, "How satisfied were you with using the Clinical Workstation to accomplish this task?". They rated their satisfaction level on a scale from 1 (Very Dissatisfied) to 4 (Neutral) to 7 (Very Satisfied). The results from all physicians were averaged together to give an overall satisfaction rating for the task.

Analyze usability results. Once the results from each physician were measured, the results could be analyzed. Typically, the various usability measurements were averaged across all physicians within this testing cycle to determine the prototype's overall usability level. It was also useful to look at the results by physician category. For example, to look at the differences experienced by medicine physicians vs. surgery physicians or community physicians vs. academic physicians. By splitting the results in this way, important questions could be answered:

1. Did the changes made to improve the usability of a feature actually improve its usability?

If changes were made to a feature between part 1 and part 2 of the usability tests, the average usability measurements for the physicians tested in part 1 were calculated separately from the usability measurements for physicians tested in part 2. For example, in the task that exercised the Intake and Output feature in the prototype, the results from the physicians participating in part 1 of the usability test were averaged together separately from the results of the physicians participating in part 2. Figure 2 shows the actual results from a recent usability test cycle of the Intake and Output Function of the CW.

From these measurements we could tell that the design changes did indeed improve the design, but more improvements are still necessary.

Measuring Methods	Usability Test, part 1 results	Usability Test, part 2 results
Time to complete task	8:22	6:11
Percent of task complete without assistance	80%	82.5%
Number of problems	3.6	2.3
Amount of assistance requested	1.0	0.6
Satisfaction level	5.0	6.0

Figure 2: Usability Test Results

2. Is there a difference between the usability experienced by a computer novice vs. an experienced computer user?

Each physician fills out a background questionnaire the first time he/she participates in usability testing. Based on the results of the background questionnaire, the physicians were categorized as novice or experienced computer users. By averaging results for the physicians categorized as novice computer users separately from physicians categorized as experienced computer users, we have found that the latest version of the CW has a bigger gap between the two types of users than was expected. As more features have been quickly incorporated into the CW during version 2, more "computer-eze" has also been incorporated into the interface. This indicates some design changes may be necessary to address the problem.

Change the design to improve usability. The final step remaining was to change the design of the prototype to address the identified usability problems. First the results had to be communicated to the actual developers. We used three different methods to accomplish this:

1. Have developers observe the usability tests. There is no better way to explain a usability problem than for developers to see multiple physicians experiencing the exact same problem with a feature. This is the preferred method for communicating the results, but there is overhead in time and

money for the developers to travel to the customer/user site.

2. After analyzing the results, prepare a report describing the usability results and a highlight video. The highlight video contains clips of physicians experiencing usability problems. As each usability problem in the report is being discussed, video clips showing the physicians experiencing the problem are shown. This is the next best method because the developers still see the users experiencing the problems. This approach, however, is a very time-intensive process which causes a delay in the results being communicated.
3. After analyzing the results, prepare a report describing the usability results. The report contains not only the usability problems but also includes quotes from physicians verbalizing the problems. The results were communicated in person while talking to each developer individually about the usability problems. During the communication of the results, the actual scenarios the physicians performed during the usability tests were described. When the developers could not attend the usability tests, this approach is the one chosen most often on our project.

RESULTS / REFLECTIONS

The usability testing process described in this paper has enabled our project to determine quickly areas of the CW that had usability problems, determine if changes in designs intended to improve the usability of the system actually resulted in improving its usability, and determine if the overall level of the CW's usability was acceptable.

NEXT STEPS

Usability testing has been a valuable tool to improving the CW's usability and increasing its likelihood of success in the "real world." However, the usability tests are not conducted in the "real world" but are usually conducted in a central, but convenient location to the physicians. It is anticipated that the physicians are more tolerant of problems and the product's performance because they are not in the context of their real world pressures.

Consequently, our future plans will incorporate additional usability testing of physicians doing actual clinical decision-making in their work environment. We anticipate this will provide a more accurate measure of the CW's acceptability in the real world for certain usability dimensions like system performance.

Acknowledgments

Michael Kahn is supported by grants R29-LM05387 and U01-LM05845, both from the National Library of Medicine.

References

1. Fritz, K. and Kahn, M. "Project Spectrum: An Eighteen Hospital Enterprise-Wide Electronic Medical Record." in C. Peter Waegemann (ed), Proceedings, Toward an Electronic Patient Record '95. Volume 2, Medical Records Institute, Newton MA, pgs 59-67, 1995.
2. Wixon D., Holtzblatt K., and Knox S. "Contextual Design: An Emergent View of System Design." *Proceedings of CHI '90.* New York: ACM. April, 1990.
3. Wixon, D. and Raven, M. "Contextual Inquiry: Grounding your Design in User's Work," *CHI '94 Tutorial.* April, 1994.
4. Holtzblatt, K. and Jones S. "Contextual Inquiry: Principles and Practice," Digital Equipment Corporation document: DEC-TR 729, October, 1990.
5. Coble, J.M., Maffitt, J.S., Orland, M.J., and Kahn, M.G. "Contextual Inquiry: Discovering physicians' true needs," *Proceedings of SCAMC '95.* October, 1995.
6. Coble, J.M., Maffitt, J.S., Orland, M.J., and Kahn, M.G. "Using Contextual Inquiry to Discover Physicians' True Needs." In Wixon, D., Ramey, J., (eds.), *Field Methods Casebook for Software Design.* J. Wiley, New York, 1996.
7. Coble, J.M., Karat, J., and Kahn, M.G. "Maintaining a Focus on User Requirements Throughout the Development of Clinical Workstation Software," *Proceedings of CHI '97.* March, 1997.
8. Kushniruk, A., Patel, V., Cimino, J.J., and Barrows, R.A. "Cognitive Evaluation of the User Interface and Vocabulary of an Outpatient Information System," *Proceedings of AMIA*, 1996.
9. Wu, S., Smith, J.W., and Swan II, J.E. "Pilot Study on the Effects of a Computer-Based Medical Image System," *Proceedings of AMIA*, 1996.
10. Rubin, J. *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests.* John Wiley & Sons, Inc. 1994.
11. Whiteside, J., Bennett, J., and Holtzblatt, K. "Usability Engineering: Our experience and evolution." In M. Helander (Ed.) *Handbook of human-computer interaction.* Elsevier, Amsterdam, 1988.