

Iterative Usability Evaluation Methods Applied to Learning Technology Development

John M. Sullivan
Department of Information Science and Technology
University of Missouri – Rolla
United States
jms352@umr.edu

Richard H. Hall
Department of Information Science and Technology
University of Missouri – Rolla
United States
rhall@umr.edu

Michael G. Hilgers
Department of Information Science and Technology
University of Missouri – Rolla
United States
hilgers@umr.edu

Ronaldo Luna
Department of Civil, Architectural, and Environmental
Engineering
University of Missouri – Rolla
United States
rluna@umr.edu

Matt R. Buechler
Department of Computer Science
University of Missouri – Rolla
United States
buechler@umr.edu

William T. Lawrence
Department of Civil, Architectural and Environmental
Engineering
University of Missouri – Rolla
United States
lawre@umr.edu

Abstract: The goals of this project were to: a) carry out the evaluation of a learning system for teaching Civil Engineering students to use Geographic Information Systems software within the context of relevant problems and b) explore the efficacy of iterative usability testing as a tool for development of learning technologies. A series of three usability tests were conducted based on an initial proposed design of the learning system. In general, the results pointed to the role of the learning system as a support tool, and a consequent design that focuses on progressive scaffolding, flexibility, and logical consistency. The results also supported the efficacy of usability testing, in that iterations of the system improved as development progressed. Finally, the importance of scarce resources and education of the development team on the role of usability testing, were identified as central factors in determining the effectiveness of usability testing.

Usability Testing and Prototyping as Formative Evaluation Methods

Usability testing is the process of observing users interacting with and using a product to ascertain how well users function with the product. The iterative nature of usability testing comes when the testing is integrated into the design and development cycle resulting in incremental improvements. Characteristics that are generally found in usability testing (Dumas & Redish, 1999) are: 1) Each test has particular goals and concerns stated when designing the test; 2) The participants are representative of real users; 3) Participants work on real tasks; 4) Participants actions and words are recorded and observed; 5) Data is analyzed to diagnose problems and propose recommendations. Thus, in usability testing there is a controlled situation, recognizable with explicit goals, and nearly everything (times, comments, pauses, expressions, etc) is recorded.

Usability testing is generally considered to be different than a traditional experiment; the latter consisting of formal hypothesis testing, the manipulation of an independent variable, and the random assignment of participants to experimental groups while usability testing is closer in nature to observational or exploratory research, where variables are not manipulated, and hypotheses are not required. Questionnaires and interviews are often administered to determine prior knowledge as well as satisfaction with the product. Users are often asked to adhere to a “talk-aloud” protocol, the process of explicitly stating what one is thinking during a given task, to provide rich data about ongoing thought processes. Usability testing typically employs a small number of users, a phenomenon referred to as “discount usability testing,” because this results in a discount in both cost and time (Nielsen, 1994).

Due to the low number of participants, there is generally an emphasis on qualitative, rather than quantitative analyses, and inferential statistics with an emphasis on probability levels are not usually applied.

Usability Testing in Education

Some researchers who work within both the Human-Computer Interaction (HCI) and Learning Technology communities have suggested that the communication between the two is less than optimal, despite the fact that they share common subject matter (Dillon, 2000). Though usability testing of web-based learning technologies has become more common in recent years (Dean, 1999; Richardson, Swan, & Newman, 2000), it is still relatively rare.

There are a number of reasons why usability testing has not become a standard tool for learning technology researchers. First, usability testing does not fit well within the most common conceptions of qualitative research in education, which tend to focus more on naturalistic observation, with intense prolonged contact in a field setting (Miles & Huberman, 1994). Second, the small number of participants typically used in usability studies does not allow for much generalization of results, particularly if the technology being studied has a wide and diverse effect on users. Third and conversely, the use of single-subject observation and intensive data collection and analysis does not easily allow for the large number of participants required for any sort of inferential statistics to be applied. Fourth, and most fundamentally, usability testing is usually implemented with well-defined tasks and well-defined performance measures. Despite these issues, there are some specific cases where traditional usability testing can be very beneficial for the evaluation of educational technologies. In particular, when used as a supplement to more traditional quantitative experimentation, a detailed analysis of a small number of typical users interacting with instructional software can provide invaluable insights and explanations with respect to the quantitative results (Hall, Philpot, Hubing, Flori, & Yellamraju, 2004).

Project Context

The learning system serves as a focus of the project is rooted in teaching civil engineering students how to use GIS. A Geographic Information System (GIS) is a computerized database management system that provides geographic access (capture, storage, retrieval, analysis and display) to spatial data. GIS can produce maps and pertinent information in a visual format regarding a specific area based upon the spatial data provided. While the industry sector of civil engineering has begun the process of integrating GIS itself, the academic world has been slower to respond. Since civil engineering is replete with uses for GIS functions, public agencies' (the civil engineer's primary employer) use of GIS technology is increasing rapidly. There exists a consequent need for civil engineers versed in GIS and able to apply GIS tools to civil engineering problems in innovative ways

The learning system developed for the civil engineering curriculum for this project focuses on a geotechnical application. The prototype consists of a comprehensive problem and an associated repository of learning objects organized using progressive scaffolding approach (Sullivan et al., 2004).

Method and Results

Identification of an Existing System

After establishing some initial design parameters, a review of existing systems was carried out with the goal to identify a learning system that was close enough in design, to provide useful and meaningful data to guide future development and iterative testing of prototypes of the proposed system. In this case the most important requirements for such an existing system were: a) The training tasks were similar in that the learner was required to learn a software tool, within the context of specific closed-ended problems, and the learning system was external to the target software to be learned; b) The interface was similar to the interface design proposed for the GIS system, which was based on past research and guidelines associated with design of learning technologies; c) The system was web based; d) The content could be easily conceived as learning objects; and e) The system was designed to supported multiple levels of scaffolding consistent with a progressive scaffolding model.

Microsoft's E-Learning Library (MELL) was identified as a software system that met these criteria (<http://www.microsoft.com/learning/mell/>). The main goal of the Microsoft software is to educate end-users on how

to use different Microsoft products in real-world situations. The training goals map well with those of our project, which is intended to instruct civil engineering students on how to use GIS software while reinforcing and applying skills learned in the traditional class. However, the tasks of the software are different in the spectrum of training versus educating.

It was decided that our system interface would be modeled on this system, dependent on the results of initial usability testing of the MELL interface, even with an understanding of its limitations. A MELL module that covers Microsoft Access®, a database implementation was selected for this experiment. Four students were recruited from an Information Systems class in an Information Science and Technology (IST) department where database concepts are taught and practiced, and were given a specific problem to solve using Microsoft Access. This testing mapped well to our project, since the students in the IST course had received instruction in the basic theory of databases, similar to our target subjects who will have had instruction in geotechnical theory. In essence, the test subjects for MELL and the intended end-users of the learner prototype had the pre-requisite understanding of their respective domains, but were being asked to implement it in an unfamiliar manner. In both cases the purpose of the learning system is to teach the students to use a software tool (i.e., Access or ArcView GIS), within the context of specific problems.

First Iteration Usability Testing: MELL Website

Overall, users found the MELL system quite usable as exemplified by 75% of the users completing all required tasks. This is impressive when given that all but one user rated themselves as having no prior knowledge about forms within MS Access. While Subject B was unable to complete any of the tasks, the problem did not appear to be system oriented. Both Subjects A and D found that the MELL website greatly helped their completion of tasks as indicated in their exit questionnaire. Subject B seemed to find the website useful despite being unable to complete the tasks with its aid. Subject C, who did complete all the tasks successfully, did not feel that the website greatly helped the completion of tasks. This correlates well with his pattern of use for the site, which was to only use the site if he could not figure out the next step on his own first.

Given the over-all positive nature of the results, the team felt confident in moving forward with the proposed interface design for the GIS system, which was similar to the MELL interface. However, this initial testing also pointed to weaknesses that could be addressed in development of the GIS prototype. The following specific recommendations were provided: a) Model overall GIS interface after MELL interface; b) Increase the contrast/size of the mouse cursor on the example animations; c) Increase the size of the next step button and place it on both the top and bottom of the window; d) Make the plain HTML page contain the same instructions as given as the sidebar on the animation page; and e) Support both static images as well as the dynamic animations (allow users to initiate playback for individual sub-tasks)

The testing also provided useful insights into how the user went about using the learning system in performing their tasks. Both the students' questionnaire responses and observation of their behavior indicated that they used the low level scaffolds to a greater degree. Further, the reliance on low level scaffolds increased as they progressed through the tasks, and the reliance on lower level scaffolds was greatest with those who were most skilled. All of these general findings are consistent with the series of progressive scaffolding studies discussed earlier (Hall, Digennaro, Ward, Havens, & Ricca, 2002), which were carried out with learning systems designed to teach different content. This pattern of use is consistent with a user who recognizes the goal of performing specific tasks with the target software, while limiting interaction with the learning system to the minimum necessary to perform these tasks. In general the learning system is then truly a peripheral tool, which is only used when required, and only at the minimum level necessary for task performance. Therefore, in general, design practices that support this type of learning activity will be most effective.

Second Iteration of Usability Testing: Initial Prototype

An initial prototype of the GIS learning system was developed, integrating the recommendations from the MELL assessment. Usability testing of this initial prototype was carried out in order to provide direction and improve the development of the GIS learner system. This prototype represented the initial design choices for layout and only included a portion of the content to be incorporated in the final prototype. This study consisted of four (4) undergraduate students (3 male, 1 female) who were enrolled (Summer 2004) in undergraduate civil engineering courses at the University of Missouri-Rolla. After filling out experiment consent forms, the subjects were given a

brief questionnaire to determine their familiarity, enthusiasm and/or comfort level with ArcView GIS, geotechnical engineering.

In terms of overall effectiveness of the interface design, the results were generally positive, indicative of a design that effectively supported user activities. All users completed the objective to produce the detailed map in ArcView GIS, but with varying levels of quality. The fact that all users completed the task is still noteworthy, given that the pre-questionnaire indicated that no users felt either comfortable or confident with ArcView GIS software. The post questionnaire responses were consistent with the high completion rate, in that users found the completion of tasks easy when facilitated with the learning module. The website and its design also rated highly among the participants. The student's use of the various scaffolds, again, provided interesting information about their use of the tool. Users unanimously preferred the high scaffold of animation compared to lower scaffold of text. These findings are seemingly inconsistent with the results of the MELL testing, which was interpreted as indicating that users tended to utilize the minimal scaffold necessary to perform the task. The problem noted with the lower scaffold of text was that it was not truly developed in this prototype. Users were forced to use the higher scaffold of video, which was complete for the tasks required.

However, even with the necessity of videos being used, these findings still provided more detailed insight into issues that were associated with the tendency to avoid or become frustrated with the higher-level scaffolds. First, participants had issues with the speed of video playback across the board. When watching videos that showed rather simplistic tasks, users became bored and frustrated with a video that was too elementary. On the other hand, when working on a complex problem and trying to follow the accompanying video, users became flustered with video whose speed exceeded their ability to work.

The issues associated with the use of multiple scaffolding levels, in particular the video, lead to some more general design principles, which are again consistent with the view of a user who is using the learning system as a flexible tool when necessary, while focusing on a specific task. First, it is important to provide the user with as much control as possible, particularly with the higher level scaffolds, to provide the flexibility necessary for her to focus on specific needs, without spending an inordinate amount of time or effort. Second, it's important to provide consistency, as exemplified by user frustration with the inconsistent level of detail. Consistency allows the learner to more readily form a mental model, which represents the whole system, facilitating quicker scanning and search for specific information necessary to perform tasks.

The following is a list of recommendations distilled from both observations and user suggestions: a) Review dialogue prompts and strive for consistency; b) Add static text on how to complete the tasks; c) Make an overall "Introduction" section to the lab; d) Make movies open in same window; e) Add a title to video web-page that correlates with the section heading; f) Improve video navigation; g) Increase visual contrast of cursor for added feedback; and h) Add redundant next and previous links in the main frame

Final Iteration of Usability Testing: Final Prototype

The usability of the final ArcView GIS Learner System prototype was tested in order to provide recommendations for final improvements prior to implementation in a classroom setting. This study consisted of five (5) senior undergraduate students (5 male) who were currently enrolled (Fall 2004) in civil engineering at University of Missouri-Rolla. Three (3) of the participants worked individually while the remaining two (2) worked cooperatively.

This final iteration of testing revealed a prototype that appeared to be substantially improved from the original, with all users completing the required tasks as much as technically possible (the webpage required for "soil testing" results was down during some users testing and they had make decisions without that information). Not only were all objectives completed by the users within the two-hour time block, but they deemed the experience as generally favorable and improvements focused around content clarifications. Users felt that they had learned a great deal about borrow site selection from this lab. In regards to its motivational factor, they again rated it above average, implying this lab is more motivating than the average civil engineering lab. They indicated that the activity was highly indicative of "real world" engineering. Overall, they felt that their knowledge on borrow sites increased rather dramatically. Qualitatively, the prototype implementation proved to be a success as well. Users were able to make good use of the tool provided, and minimal questions were asked of the researcher during the study. Users were able to adjust to the ArcView application quickly and transfer skills/functionality learned in an early module section to accomplish a task required for later sections.

The navigation within the learner module also appeared sufficient based on users' behavior, in that none got lost, and based on self-report, in that users did indicate that this was a problem.

Despite the fact that the videos had been changed based on recommendations from the initial prototype testing, all users still had some level of discomfort with watching the videos. The control provided was still insufficient which made for movies that were hard to speed up or slow down.

The learners' use of various scaffolds went smoothly with users finding the level of scaffold necessary with minimal effort. As with the previous testing, with the exception of the under-developed initial prototype and its text scaffold, users preferred the more minimal level of scaffolds, and the process of users beginning with minimal scaffolds and moving to a higher level if the minimal failed, was even more evident in this research.

Recommendations revolved around content issues, with the exception for the future recommendation of dynamic videos.

Discussion and Conclusion

The usability testing began with the testing of the MELL system in an attempt to investigate a system with design and functionality similar to the proposed GIS prototype. The four undergraduates who participated provided us with useful insights into how end-users interact with and manipulate a learning technology towards accomplishing learning goals as well as the strengths and weaknesses of the MELL design. Three of the four participants were able to make good use of this prototype with most users completing all required tasks with a high degree of competency. The students made use of the multiple layers of assistance. Text based help was rated to offer more assistance than its video counterpart, but the video scaffold allowed users to acquire more in-depth help when text proved insufficient, and also served as a tool for checking their work. This initial evaluation of an analogous system also indicated the need for some design modifications, including issues with video and navigational components such as video playback speed, visual contrast of actions, and navigational tools being out of sight.

The first round of testing for the project teams' initial design, again, consisted of four undergraduates, but since this prototype was developed with target population in mind, civil engineering undergraduates were excellent candidates to test the system. With the intended audience testing the system, new insights were garnered. First, video speed was again found to be a problem, with users complaining that the videos were slower than they needed to be, once again taking time away from their opportunity to interact with the target software, Second, dialogue boxes, which were added to the videos, to provide users with additional information, were inconsistent in the amount of information they contained, some times extremely detailed and at other times very general – this inconsistency clearly bothered the users. With regard to navigation, users had trouble with the tree structure provided and at times became lost in the system. While the SCORM interface did contain embedded next and previous buttons, most users did not use them. Finally, even after being given a general overview of the tests purposes and the prototype, users still had trouble understanding what their true objective was.

Despite the large number of recommendations for improvement, over-all users found the system to be quite usable. The exit questionnaire results showed that, not only did they find the tasks to be relatively easy, but found the website easy to use as well.

The prototype was again tested by four undergraduates in civil engineering after recommendations were implemented and the rest of the content was developed, but prior to a larger applied experiment, which was later carried out within the context of a Civil Engineering class. Though over-all the system performed much better than the previous iteration, issues in need of modification were identified. Video was again a culprit as users still found it too slow to aid the completion of all tasks. Navigation functioned much more effectively than previous iterations, presumably more consistent with users' mental models. Users still began by over-estimating task difficulty and using the higher scaffold of video, but quickly reverted to the lower scaffold of text for the majority of steps. There were some frustrations in users when text descriptions became too long and they consequently had more trouble with those sections. Content disparities were noted as well, with an eye towards making a seamless prototype. Most recommendations focused on content, indicating that most significant interface issues had been adequately addressed.

The prototype navigation worked well because it did not dictate how a user would be forced to use the system. Users were allowed to select the navigational model that fit best with their personal mental models. This choice afforded users the ability to focus in on tasks and less on how to achieve the tasks through the navigation provided.

Taken together, these results point to the importance of a learning tool as a flexible support system – the tool is not the task. There is no need to require users to engage all parts of the learning system, when all they need to do is to accomplish their tasks. This is why the navigation of the final prototype had minimal problems during

testing. Users were able to apply the navigational scheme that most facilitated task completion. Across all experiments learners tried to minimize their interaction with the tool by relying on minimal scaffolds as much as possible. Again, users do not want to spend more time on additional tasks than necessary. The lower scaffolds, despite being more abstract, offer the quickest interaction. Students were also most critical of, and performance was most hampered by, situations where they were unable to quickly locate the important information which would allow them to perform their primary task of learning the software. When users were not given the opportunity to work through the system deftly, their performance suffered and agitation arose. Not only do users like quick learner system interaction to be able to move to the text task, a lack of quick interaction encourages poorer performance. This leads to at least three general design principles, which apply in systems such as that developed in this project. First, the presence of multiple scaffolds is helpful since it allows the user to select the level of scaffold that is needed at any given time. Second, it is important that the learning system be flexible, including multiple types of navigation schemes (i.e., hierarchical and linear), and allowing for easy access of the necessary scaffold at any given time. Third, it is particularly important that the content be structured consistently and logically, allowing the learner to quickly develop an effective mental model, since the user needs to be able to quickly identify and acquire specific content.

This project yielded a number of important insights about the efficacy of iterative usability testing as a tool for enhancing the effectiveness of learning technologies. Over-all, the approach did indeed appear to result in a more pedagogically sound and usable system. Which each step of the evaluation process the system improved based on observations of users and their feedback. The MELL website tested well and was acceptable to the users. The initial prototype, despite drawbacks and limitations, tested well also. The final iteration tested significantly better than the previous two. This growth of user acceptance is one indicator that usability testing integrated into the development process leads to more effective learning technologies. In addition to indicating that the method can be effective over-all, other more subtle issues that mediate this effectiveness were identified in this project. The two factors were: a) resource constraints and b) the need to educate team members on the role and importance of usability testing.

References

- Dean, J. (1999). The Functional Web: User Testing Your Site. *Proceedings of World Conference of the WWW and Internet*, 1684.
- Dillon, A. (2000). Designing a Better Learning Environment with the Web: Problems and Prospects. *CyberPsychology and Behavior*, 3(1), 97-102.
- Dumas, J. S., & Redish, J. C. (1999). *A Practical Guide to Usability Testing*. Exeter, England: Intellect.
- Hall, R. H., Digennaro, A., Ward, J., Havens, N., & Ricca, J. (2002). Development of a web-based learning system for teaching web design and development: A problem-based, progressive-scaffolding approach. *Proceedings of the AACE E-Learning Conference*, 1562-1565.
- Hall, R. H., Philpot, T. A., Hubing, N., Flori, R., & Yellamraju, V. (2004). A Model-Driven Multi-Year Assessment of a Software Development Project for Engineering Instruction. *Proceedings of the American Society for Engineering Education*, http://www.asee.org/acPapers/2004-1651_Final.pdf.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis* (2nd ed.). London: Sage.
- Nielsen, J. (1994). Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. In R. G. Bias & D. J. Mayher (Eds.), *Cost-Justifying Usability* (pp. 242-272). Boston, MA: Academic Press.
- Richardson, J. C., Swan, K., & Newman, D. (2000). Evaluating Software for the New Millenium: An Example from Project Links. *Proceedings of World Conference on Educational Multimedia, Hypermedia, & Telecommunications (EdMedia)*, 1451-1452.
- Sullivan, J. M., Hall, R. H., Luna, R., Hilgers, M. G., Taylor, A. J., & Buechler, M. R. (2004). Degree of scaffolding as learning object metadata: A prototype learning system design for integrating GIS into a civil engineering curriculum. *Learning Technology*, 6(2), 60 - 64.

Acknowledgements

This work was supported in part by a grant from the National Science Foundation's Curriculum, Course, and Laboratory Improvement program (Award No. DUE-0341016).