Developing an Assessment Model for Evaluating Software Tools in Education

Vedant Jain  
*Missouri University of Science and Technology, vkqk7@mst.edu*

Hong Sheng  
*Missouri University of Science and Technology, hsheng@mst.edu*

Richard Hall  
*Missouri University Science and Technology, rhall@mst.edu*

Michael G. Hilgers  
*Missouri University of Science and Technology, hilgers@mst.edu*

Follow this and additional works at: [http://aisel.aisnet.org/amcis2009](http://aisel.aisnet.org/amcis2009)
Developing an Assessment Model for Evaluating Software Tools in Education

Vedant Jain  
Missouri University of Science and Technology  
vjqk7@mst.edu

Hong Sheng, Ph.D.  
Missouri University of Science and Technology  
hsheng@mst.edu

Richard Hall, Ph.D.  
Missouri University of Science and Technology  
rhall@mst.edu

Mike Hilgers, Ph.D.  
Missouri University of Science and Technology  
hilgers@mst.edu

ABSTRACT

This paper presents a research model to evaluate the usability of computer-mediated learning technology, user acceptance of the technology, and the effect of the technology on learning outcomes. We also introduce the use of innovative research techniques for the evaluation, including eye tracking and think-aloud protocol. Think-aloud protocol allows us to understand a user’s state of mind while he/she is interacting with the system. Eye tracking also provides additional insights into a user’s cognitive state while using the system. Both qualitative and quantitative data can be collected following the research model, thus providing more in-depth understanding on the effect of the system. We demonstrated the application of the research model and research techniques in evaluating visualization software designed in a project funded by the National Science Foundation.

Keywords

Usability, Technology Acceptance Model, Thinking aloud, Eye tracking

INTRODUCTION

Information technology has been widely used in education. It can help students to organize their learning activities (Gillet et al., 2005). It can also allow students access to more resources, which can facilitate and support acquisition of knowledge and information (Kumar et al., 2008). Information technology can also assist lecturers as a content transmission tool to transfer and distribute course materials and other related information (Benbunan-Fich, 2002).

Information technology, therefore, can not only enhance quality of learning, but also accommodate various learning styles (Carlson and Sullivan, 1998). In engineering education, Bissel (1996) categorized the application of IT as computer-aided learning, generic software applications such as Mathematica, and commercial engineering packages. He also pointed out the need for exploiting IT in development of specialized software packages for engineering educations (Bissel, 1996).

However, educational software can help its users in their work only if the users readily accept the technology. Previous studies have shown that most engineering students are found to be visual learners (Kolari and Savander-Ranne, 2004). This implies that the software used for engineering education should be visually enhanced in order to facilitate learning among the engineering students. Although previous studies have evaluated computer-mediated learning software and its usability (Gillet et al., 2005), a comprehensive model to aid the design of such software is yet to be established.
RESEARCH MODEL

This paper presents a research model for evaluating the effectiveness of computer-mediated learning software based on various constructs and models from prior literature. One such model, which has been widely applied in the prior literature, is the technology acceptance model (TAM). The TAM model was expanded to include new variables, such as learning styles and outcomes, in order to examine the connection between the user, the (possibility of) software being evaluated as a learning tool, and the learning outcomes after using the software. This model is also an extension of the LITE assessment model, which has been developed and improved upon over a period of time (Hall et al., 2006). Various methodologies and innovative techniques can be applied in the evaluation in order to triangulate the results. We applied this model along with a set of research techniques in evaluating specific educational software.

**Figure 1: Research Model**

**Technology Acceptance Model (TAM)**

The Technology Acceptance Model (TAM) is an extension to Ajzen and Fishbein’s (1975) Theory of Reasoned Action, which explains the relationship between attitude towards the technology and intention to use it. Theory of reasoned action suggests that if the person is to perform a certain action, it would depend upon his or her attitude toward that action and how others would see it, whether or not the action is performed (Davis, 1989; Ajzen and Fishbein, 1975).

TAM suggests that perceived usefulness of the technology, as well as users’ perceived ease of use of the technology determine users’ attitude towards the technology, which in turn influence users’ intention to use the technology. According to Davis (1989), perceived usefulness is defined here as "the degree to which a person believes that using a particular system would enhance his or her job performance", whereas perceived ease of use is defined as "the degree to which a person believes that using a particular system would be free of effort." Attitude can be defined as the supportive or unsupportive feeling towards that particular behavior (Ajzen and Fishbein, 1975).

TAM has been widely applied to study the user acceptance of new technology and has been proven to be effective in predicting users’ intention to adopt or use a particular technology. For example, Davis (1989) showed that TAM can explain the use of information technology; while Lederer et al. successfully applied it with the World Wide Web (Lederer et al., 2000).

Learning Styles

Felder and Spurlin (2005) established various dimensions that define how students receive and process information, and the manner in which they learn.

A student’s learning style may be defined in the following dimensions (Felder and Letterman, 1988):

1. The kind of information that the student prefers to receive: sensory (external), such as sights, sounds, and physical sensations; or intuitive (internal), such as possibilities, insights, and hunches.
2. The sensory channel through which external information is most effectively perceived: visual, such as pictures, diagrams, graphs, and demonstrations; or auditory/verbal, such as words and sounds.
3. The organization of information with which the student is most comfortable: inductive, in which, facts and observations are given and underlying principles are inferred; or deductive, in which, principles are given and consequences and applications are deduced.
4. How the student prefers to process information: actively (through engagement in physical activity or discussion), or reflectively (through introspection).
5. The way in which the student progresses toward understanding: sequentially, that is, in continual steps; or globally, which is in large jumps and holistically.

The different ways of perceiving and understanding information influence students’ attitude towards a learning technology, which, in turn, is likely to affect the learning outcomes when using such technology. For example, a visual learner would presumably find it easier to learn a system using effective diagrammatic depictions while an auditory/verbal learner would find an oral instruction to be more helpful. Previous research has shown that information technology can support knowledge dissemination to people with different learning styles (Carlson and Sullivan, 1998). Therefore, learning systems should be designed to accommodate the different learning styles of the users.

Learning outcomes

We also want to assess how much the system facilitates learning. In the proposed model, learning includes perceived knowledge gained by the users with respect to various methods of studying/teaching, such as text, lab, and computer software. It also includes quantitative measures for related concepts such as motivation to learn, real world applicability, and metacognition (learning awareness) (Hall et al., 2006).

Usability

ISO/ FDIS 9126-1 define usability as “product quality as the capability of the software product to be understood, learned, used and be attractive to the user, when used under specified conditions” (Bevan, 2001). Importance of usability lies in the fact that the final software product should be easy-to-use for the end users and should meet the specific user requirements (Bevan, 2001; Bevan et al., 1991). These goals can be ensured by testing the software with the end users during the development life cycle of the software. Iterative evaluation is quintessential to ensure usability in software applications. (Harton et al., 2003)

RESEARCH METHODOLOGIES

The research model proposed in this paper can guide the evaluation process of educational software. Both qualitative and quantitative data can be collected, which will give us a more in-depth understanding of the effectiveness of software, users’ acceptance of the technology, as well as usability and design issues of the software.

A survey can be conducted to assess the users’ acceptance of the technology as well as the learning outcome. Qualitative data can be collected by interviewing the users, and observing them interacting with the software.

In addition to these research methods and techniques, we also introduce two innovative data collection techniques that could be used in usability testing, which are think aloud protocol and eye tracking.

Think Aloud Protocol

In the think aloud protocol (TAP), users are asked to talk aloud about whatever they are thinking while performing a task on the system (being evaluated), such as where they are within the system, what they should do next and the problems that they are facing (Someren et al., 1994; Lewis and Rieman, 1994).

Think aloud protocol is used mainly to provide insight into the cognitive behavior of a user while he/she interacts with the system. This method was first used to analyze problem solving and, more recently, has been used to collect data during usability testing for product development (Someren et al., 1994). Previous literature suggests that TAP is particularly useful as part of the iterative development cycle of software (Roberts and Fels, 2006). While users interact with the system and make errors, the think aloud protocol can provide vital clues about where and why the errors were made (Lewis and Rieman, 1994). The TAP has also been effective in indicating the source of problems or errors (Someren et al., 1994). Previous studies have shown that TAP is more effective for identifying problems during interface evaluation when compared with questionnaires or interviews (Donker and Markopoulos, 2002).

Eye tracking

An eye-tracker records a user’s eye-movements during his/her interaction with a system. Eye-tracking has been used for evaluating websites like e-commerce websites. For example, Tzanidou et al. (2005) summarized studies using eye-tracking on e-commerce websites. Eye tracking has also been applied to study users’ visual preferences for color (Djamasbi et al., 2007).

The eye tracking data can be analyzed and the results are usually presented as gaze plot or hot spot graphs. The Gaze Plot displays a static view of the gaze data for each image of the stimuli and is a useful tool for visualization of scan paths. Each fixation is illustrated with a blue dot where the radius represents the length of the fixation - the larger the dot the longer the user fixated on that particular point. Hot Spots are basically heat maps that depict the areas where the users focus. The colors in the map differ depending on how long the user focused on a given area, with darker colors representing more focus, and lighter colors representing less.

The gaze data can be quantified and further analyzed by dividing the stimuli into different parts as areas of interest. Area of interest is defined as an area of a display or visual environment that is of interest to the research or design team and is thus, defined by them (not by the participant) (Jacob and Karn, 2003). The gaze data for areas of interest consists of fixation and gaze % (proportion of time) on each area of interest.

In evaluating educational software, we would like to examine how users adapt to an unfamiliar design layout and how quickly they learn where to look for certain information (Tzanidou et al., 2005). Therefore, eye tracking is suitable for this kind of evaluation.

APPLICATION OF RESEARCH MODEL AND METHODOLOGIES

Project Background

We utilized this model for a pilot project wherein we conducted usability testing on the Rapid Development System (RDS), which is a learning system under development, funded by the National Science Foundation. A primary goal of the system evaluation is to provide feedback for future design to the development team. RDS is used in teaching materials on control design/insertion in Mechanical Engineering. RDS is a graphic user interface (GUI) for easy controller, interpolator, and model implementation with a mini-CNC system (see Figure 2). The graphic user interface of the RDS was created using GUIDE in the Matlab 7.1 (R14) software. The current GUI has three mode selection options: simulate, emulate, implement. Simulate allows the users to test the controller on the local PC and tune the controller after initial design. Emulation allows the user to run the controller on the target PC (machine PC) to determine if the communication time, computation time, and send out time can all be accomplished within the desired step time. This will allow the user to further tune the controller. Implementation is the final test and allows the user to test the controller in actual operation of the mini-CNC. Production of parts is possible in this mode.

The GUI has the option of selecting a user-created controller or selecting one of two premade controllers, Proportional and Proportional Integral Derivative (PID). Once the user has selected the mode and the controller a
push button is selected and the Simulink interface model is created. The goal of the project was to evaluate how well the users come to accept the RDS technology, the amount of control design/insertion learning that RDS promotes, and understand issues/problems with the RDS interface and how to overcome them in the next development phase.

![Figure 2: Rapid Development System](image)

**Research Procedure**

Three participants were recruited for the evaluation. When a participant arrived in the lab, he/she was shown the setup. The participant was asked to first sign the consent form. The participant was then given a very brief introduction of the RDS and the think aloud protocol. The list of the tasks was then given to the participant. The list comprised of 3 tasks, which covered most of the functionality of the RDS phase-I.

**TASK 1**: Simulate and select the controller. Generate model.

**TASK 2**: Emulate and select the controller. Generate model.

**TASK 3**: Go back and implement a pre-existing controller.

The participant was asked to think aloud when they were using the RDS to complete the first two tasks. The participant was occasionally prompted during the process, with questions like “what are you thinking right now?”, “what is going on in your mind”.

After the participant finished working on the first two tasks, eye tracking equipment was used to track the eye movement of the participant while he/she worked on the third task. When the participant completed all three tasks, he/she was given a questionnaire and asked to complete it. The questionnaire included questions on learning style (as a moderating variable in the model, and was adapted from Felder et al.), and questions from the TAM, including perceived usefulness and perceived ease of use (adapted from Davis, 1989), attitude towards the technology and intention to use (adapted from Shirley and Todd, 1995, Gefen et al., 2003). Also, learning outcome was measured with questions adapted from Hall et al. (2006).

After the participant completed the questionnaire, we proceeded with an interview, where we asked open-ended questions about the participant’s experience using RDS, application of RDS, his/her learning experience, strengths and weaknesses of the RDS, as well as suggestions for improvement. The interview was recorded for data analysis.

**RESULTS**

We summarize the results from evaluating RDS in the following sections.

**Task Performance**

Only one out of three participants was able to complete all the three tasks. One of the three participants could not complete any of the three tasks. Table 1 summarizes the task performance of each participant.

*Proceedings of the Fifteenth Americas Conference on Information Systems, San Francisco, California August 6th-9th 2009*
Since our study had only 3 participants, only descriptive statistics were considered. Results of the learning style questionnaire verified that these students, like most engineering students (Kolari and Savander-Ranne, 2004), are visual learners. Two out of the three participants had preference on visual learning. Participant 3’s response showed that he was both a visual and verbal learner.

**Quantitative Analysis**

**TAM**

Figure 3 summarizes the results on the technology acceptance model (TAM) survey items.

a. **Perceived usefulness** –According to the interviews with the participants, they vary in terms of previous experience with the subject area (e.g. control design/insertion). Both participant 1 and participant 3 were experienced, whereas participant 2 did not have any previous practical experience with control design/insertion, and very little theoretical exposure in courses taken in the previous semester (spring 08). We saw that both the experienced participants perceived that the RDS would be useful with the average response tending towards positive. The average response of the inexperienced participant (participant2) showed that he neither agreed nor disagreed that the RDS would be useful. The average perceived usefulness is 4.44/7.

Table 1: Task performance

<table>
<thead>
<tr>
<th>Participant</th>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Completed</td>
<td>Completed</td>
<td>Not completed</td>
</tr>
<tr>
<td>2</td>
<td>Not completed</td>
<td>Not completed</td>
<td>Not completed</td>
</tr>
<tr>
<td>3</td>
<td>Completed</td>
<td>Completed</td>
<td>Completed</td>
</tr>
</tbody>
</table>

**Figure 3: Technology Acceptance Model**

b. **Perceived ease of use** – Results showed that the experienced participants felt that the RDS was quite easy to use, while the inexperienced participant did not. The results can also be validated from the task performance of the participant where the participant who completed all the three tasks had the highest response to the perceived ease of use and the participant who could not complete any of the three tasks had the lowest response. The average perceives ease of use is 4.61/7.

c. **Attitude towards RDS** – The findings show that regardless of experience and task performance all the participants had positive attitude towards RDS. Overall, the average attitude towards RDS is 5.0/7.

d. **Intentions to use RDS** – The results indicate an overall positive intention for future use of the RDS. Overall, the average intention to use is 5.0/7.

**Learning Outcomes**

We also measured and compared the perceived information learned about control design/insertion, motivation to learn, real-world applicability of RDS; comparing the system with class lectures and text. Figure 4 summarizes the results for various aspects of learning outcomes.

![Figure 4: Learning Outcomes](image)

Results suggest that students perceived class lectures to be particularly strong on all three aspects. The text and RDS system appear to be largely equivalent, with the exception of real-world-applicability, where the RDS system was rated as high as lecture.

**Qualitative analysis**

This section is divided on the basis of the different analysis techniques used.

**Observation**

Users’ facial expressions, voice, and screen recording were captured using Morae recorder. Several key themes emerged from the observation, as well as the data generated from the think aloud protocol. Some of the major themes are as follows:

- Since the participants were using the RDS for the first time, it was necessary for them to use the HELP file. But two out of the three participants did not notice the HELP button on the RDS interface.

*Proceedings of the Fifteenth Americas Conference on Information Systems, San Francisco, California August 6th-9th 2009*
• Participant 2, after being unable to complete task 1, exclaimed when starting task 2, “Oh, there is your help function!”

• All three participants had confusion about the function of Student Controller and which controller to choose in the first task.
  • Participant 1: “Student controller is mine, I am guessing”
  • Participant 3: “I am just guessing that Student Controller is the one that I have to select”
  • Participant 2 went to the wrong controller (in task1) altogether. After opening the HELP file in task2, participant 2 exclaimed, “Oh, I was supposed to select Student Controller”

• All the participants were reluctant to use the HELP file. Even when they started using HELP, all the participants found it extremely difficult to find the right information from the HELP document because it was badly organized.
  • Participant 1: “I don’t read HELP files ever. I am anti-help files actually”
  • Participant 2: “Don’t tell me I have to read the HELP file again”; “I am sorry but this is a really poor HELP file”

Eye-tracking

We used the eye-tracker to record the users’ eye-movements and activity on the screen during task 3 in order to see how well the users learn/accept the RDS technology after completing the first two tasks.

Figure 5, 6 and 7 show the Gaze plots for Participant 1, 2 and 3 respectively.

![Figure 5: Participant 1 Gaze Plot](image)

The window in the centre is the RDS interface. We found that participant 1 focused on the RDS as well as at the window in the background, but not on the HELP button.
Participant 2 was the only user who could not complete any of the three tasks. Analyzing the Gaze plot indicates that this participant was not able to find the relevant information on the interface of the RDS to complete the tasks. For example, although there is just the desktop in the background and no other windows were opened at the time, the participant’s eyes were all over the screen, showing signs of confusion and nervousness in using the RDS interface. Also, the participant did not notice the HELP button at all.
Participant 3 was successful in completing all three tasks and the Gaze plot indicates that the participant knew what to do and where to go as most of the dots were concentrated on the RDS interface.

Further, we analyzed the Hot Spots of each participant. The Hot Spot graph for Participant 3 provided interesting insight into the actual areas of focus of the user on the RDS.

Figure 8 is the hot spot graph for participant 3.

![Figure 8: Hot Spot Graph for Participant 3](image)

There were 4 major areas on the RDS interface:

a. Drop down Menus
   i. Select Mode
   ii. Select Controller

b. Buttons
   i. Generate Model
   ii. Help

We found from the Hot Spot analysis that the user focused mainly on three of the four areas. There were three distinct circles around the drop down menus and the Generate model button but none around the HELP button, which was equal importance.

This was further reinforced when we studied the collective gaze data of all the participants in the form of areas of interest.

**Area of Interest**

To examine the gaze data further, we determined the proportion of time associated with the following areas of interest.

The RDS was divided into four areas of interest (see Figure 9):

*Proceedings of the Fifteenth Americas Conference on Information Systems, San Francisco, California August 6th-9th 2009*
1. AOI1: Mode
2. AOI2: Controller
3. AOI3: Generate Model
4. AOI4: Help

Apart from the above AOI’s, we also considered the area on the screen, which were outside the RDS.

We analyzed the fixation duration of all the participants in each area of interest. While interacting with the RDS interface, we found that on an average the participants mostly fixated outside the RDS – at the desktop or the Simulink window or the HELP document in the background. Table 2 summarizes the gaze % (proportion of time) on each area of interest.

<table>
<thead>
<tr>
<th>Area of Interest ID</th>
<th>Content</th>
<th>Gaze %</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Background: Desktop/ Simulink/ Help document</td>
<td>45.77</td>
</tr>
<tr>
<td>1</td>
<td>Mode</td>
<td>16.44</td>
</tr>
<tr>
<td>2</td>
<td>Controller</td>
<td>15.78</td>
</tr>
<tr>
<td>3</td>
<td>Generate Model (Button)</td>
<td>20.16</td>
</tr>
<tr>
<td>4</td>
<td>Help (Button)</td>
<td>1.85</td>
</tr>
</tbody>
</table>

Table 2: Gaze % (proportion of time) on each area of interest

We also noted that AOI’s 1, 2 and 3 reported fairly equal amount of gaze distribution while the HELP button had the least and by a very high margin. Jacob and Karn (2003) state that the proportion of time looking at a particular display element or area of interest could reflect the importance of the area [Jacob and Karn, 2003]. Therefore, we can conclude that the users perceived the HELP to be less important as compared to the other areas of the interface.

**Interview**

The interviews were conducted with open-ended questions, such as the participants’ previous experience with the subject area (control design and insertion), differences in their experience with RDS during lab activity (experiment), application of RDS, learning experience, strengths and weaknesses of the RDS and suggestions for improvement.

Three themes emerged from the interview, and a discussion of each participant in the context of these themes follows.

1. **Past Experience**
   Participant 1 was very experienced with control design/insertion.
Participant 2 did not have any previous practical experience with control design/insertion, and very little theoretical exposure in courses taken in the previous semester (spring 08).

Participant 3 reported that he was highly experienced with control design/insertion.

2. **Strengths of the RDS**

Participant 1 felt that the RDS was very convenient as it kept users from rewriting a lot of code every time. He thought that it would be useful in the academic environment.

Participant 2 said that the HELP file was self-explanatory and felt that RDS would be useful as a learning tool.

Participant 3 felt that RDS would be a good tool for first time exposure to control design/insertion, stating, “It would be nice if I first got exposed to it.” He also said that RDS is a good prototyping tool in the “real-world” and it is good that there is a template which one knows that works; all he/she has to do is make slight modifications to the template as opposed to starting from scratch. He thought that RDS would be useful in industry and introductory controls courses. Overall, he enjoyed the experience with the RDS and thought it to be a “Nice development system for very complex programs.” Like participant 2, participant 3 also felt that the HELP file was self-explanatory.

3. **Weaknesses of the RDS**

Participant 1 did not like Simulink as a programming language. He said, “RDS limits the controllers that you are able to make with it because one might want to differentiate past values of the control signal.”

Participant 2 felt that working on the RDS was very confusing. Also, the HELP file was not at all useful, because it was not arranged in order of the tasks.

Participant 3 stated that RDS helps to see the overall picture but one does not learn as much as he/she is not able to see the full aspects of how it is working. Also, he felt that more information about RDS could be incorporated in the HELP file.

All the participants noted that the computer that they were working on was very slow.

**CONCLUSION**

Based on the qualitative data and quantitative data collected from the pilot project, we concluded that RDS overall is an effective learning tool for students, though this may be mediated by experience. We also identified some usability issues with the design of the interface. We have made recommendations to the design team based on the evaluation results and recommendations will be implemented in the next version of the software.

The pilot study also demonstrated the use of two innovative techniques; think aloud protocol and eye tracking, in evaluating the educational software. These two techniques combined gave us a more comprehensive understanding of users’ cognition and behaviors when interacting with the software, thus helping to explain the users’ perception and attitude toward the technology.

**REFERENCES**