Supplemental Usability Testing as a Component of a Learning Technology Assessment Program

Richard H. Hall, Timothy A. Philpot, Nancy Hubing,
Ralph E. Flori, Pradeepkumar Subramanian, and Vikas Yellamraju

University of Missouri – Rolla

ABSTRACT

This paper discusses usability testing as it applies to the evaluation of learning technologies, with a particular focus on its use as a supplement to basic experimentation. The paper presents a comprehensive assessment model for evaluation of learning technologies, which includes multiple methodologies and measurement tools. An illustrative example of the application of supplemental usability testing in combination with basic experimentation, within the context of this model, is provided. This example includes a basic research experiment on the role of feedback in engineering education learning modules and a complimentary usability test. Few significant differences between feedback and non-feedback groups were found in the basic research experiment. The usability test provided explanations for these findings, such as minimal use of the feedback features and a large amount of variance in degree of engagement among participants. The usability test also indicated various ways for improving the software such as providing more guidance and structure, and subdividing modules into smaller parts.

USABILITY

The term usability has been commonly used within the field of human-computer interaction (HCI) for some time, but has become decidedly more popular in the past decade with the emergence of the World Wide Web. This is best represented by the popularity of the consummate “usability guru” Jakob Nielsen and his web site, useit.com (http://www.useit.com). According to Nielsen’s oft-cited definition, “Usability is a quality attribute that assesses how easy user interfaces are to use. The word ‘usability’ also refers to methods for improving ease-of-use during the design process” (http://www.useit.com/alertbox/20030825.html). According to Nielsen, usability consists of five basic components: a) *Learnability*: How quickly a user can learn to use the interface they first time they encounter it; b) *Efficiency*: How quickly users can perform tasks once they’ve learned the interface; c) *Memorability*: How easy is it to re-learn an interface once a user has quit using it for some period of time; d) *Errors*: What are the number and severity of errors, and how easy is it to recover; and e) *Satisfaction*: How pleasant and enjoyable is the interface for a given user (Nielsen, 2003). In general, definitions of usability focus on the degree to which a user can achieve specific tasks goals using criteria such as effectiveness, efficiency, and user satisfaction (Law & Hvannbert, 2002).

USABILITY TESTING

The term *usability testing*, not surprisingly, refers to a specific method for evaluating usability. Nielsen points to three basic characteristics that define user testing. First, representative users must be the study participants. Second, the user must perform representative tasks. Third, the users are observed in a non-invasive fashion as they perform these tasks, and the researcher’s job is to note where users succeed and where they have difficulties. In their classic book on usability testing, Dumas and Redish, identify these same basic characteristics, and add an analysis component, which states that the experimenter’s ultimate goal is to recommend changes to solve the problems that are identified (Dumas & Redish, 1999). Law and Hvannberg (2003) provide a more detailed definition of usability testing. In addition to reiterating these same defining characteristics, they point out that usability testing is just one methodology among a whole class of usability evaluation methodologies. They note that usability testing is time and labor intensive since testing often consists of single-subject observations of participants who are often asked to think aloud. Another characteristic is that sample size is small with as few as five participants (Barnum, 2002; Nielsen, 1994).

Though the term usability testing is sometimes used to refer to observations in the field (Barnum, 2002), its most commonly associated with laboratory testing. The prototypical usability lab setup includes recording devices so that the experimenter can dynamically capture the participant’s activity in the form of a video while a camera simultaneously records the participants’ facial expressions. This is typically mixed into a single picture-in-picture video of the participant’s activity on the screen with the video of facial expressions superimposed. These tapes are subsequently analyzed in great detail. The main emphasis is on evaluation of task performance, and this is typically evaluated both quantitatively and qualitatively. However, other factors associated with user frustration and unseen issues involving the interface are also noted. In addition to other classification schemes, a list of identified usability problems is created that details the number of participants who experienced the problem, the severity of the problem, and the problem’s scope (Barnum, 2002). The usability testing that was carried out in the example reported in this paper used a small number of participants, a typical usability lab setup, and detailed qualitative and quantitative analysis.

USABILITY TESTING AND EDUCATIONAL RESEARCH

In describing the basics of usability research, Dumas and Redish also point out that a defining characteristic of usability research is that it is not traditional experimental research (Dumas & Redish, 1999). Formal hypotheses are not proposed, an independent variable is not manipulated by the experimenter, and inferential statistics are not applied. Therefore, within the general context of educational and social science research, usability testing can be classified as qualitative research, in that it typically does not make use of traditional experimentation or inferential statistics. This is consistent with general definitions of qualitative research within social science fields (Strauss & Corbin, 1990). In addition, the research tends to involve “real life” tasks with subjects who would be normally be performing these tasks, which is also a fundamental characteristic of most qualitative research definitions (Miles & Huberman, 1994). However, usability testing normally occurs within a laboratory environment, and the user is expected to perform specific tasks provided by the experimenter, and though inferential statistics are not traditionally applied, quantification in the form of task performance measurement, is typically an important component of usability testing (Barnum, 2002). These characteristics are not consistent with traditional qualitative research which traditionally de-emphasizes quantification and tends to occur within a naturalistic environment, with minimal interference from the experimenter (Miles & Huberman, 1994; Strauss & Corbin, 1990). Therefore, usability testing is an interesting quantitative-qualitative hybrid.

Some researchers who work within both the 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 methodological tool for learning technology researchers. First, as mentioned above, 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 wide and diverse effects on users. Third, the use of single-subject observation and intensive data 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 works best with well-defined tasks and well-defined performance measures. Usability tasks typically involve activities such as searching for a given piece of information, entering data, or manipulating a program in a specific manner. Educational researchers, on the other hand, are often interested in outcomes such as knowledge integration, problem solving, and critical thinking. These are fundamentally different from typical usability testing task performance measures.

TRIANGULATION

Despite these limitations, there are some 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. Such a multi-method approach is traditionally referred to as triangulation — a term that is popular within both the HCI and educational research communities. For example, educational researchers point out that triangulation can be used to check the validity and reliability of findings (O'Malley & Valdez, 1996), and that multiple forms of evidence will provide a more accurate picture of the student (Wiggins, 1998). HCI researchers Wendy Mackay and Anne-Laure Fayard point out that research effectiveness can be enhanced by triangulation across research methodologies and across academic disciplines (Mackay & Fayard, 1997). This paper provides an example of such a triangulation approach utilizing multiple methodological tools common in two disciplines. We refer to the specific approach applied below as supplemental usability testing.

LITE ASSESSMENT MODEL

The Laboratory for Information Technology Research (LITE) at the University of Missouri – Rolla has developed a comprehensive framework for the assessment of learning technology projects. This assessment model has evolved over the course of a number of projects, involving the evaluation of software tools for engineering and science education (Eller, Hall, & Watkins, 2001; Hall et al., 2002; Hall, Watkins, Davis, Belarbi, & Chandrashekhara, 2001). A fundamental assumption of the model is that conclusions and recommendations should be based on the triangulation of information gleaned from multiple methodological and measurement tools. We present a brief discussion of the model here since it serves as an example of a model that utilizes triangulation of various approaches, including usability testing of learning technologies. In addition, the illustrative example of supplemental usability testing below was carried out under the auspices of this lab within the context of this model.

The model is meant to serve as a guide for large-scale assessment projects, and also to provide some context for individual experiments. Therefore, any given experiment does not include all aspects of the model, but often an entire project, such as the project referenced below, encompasses many parts of the model. Figure 1 is a graphical depiction of the LITE learning technology assessment model.

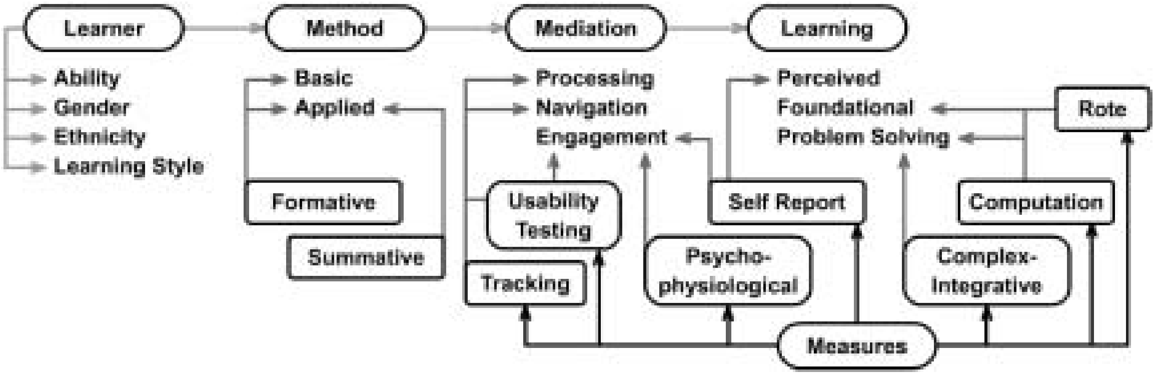


Figure 1. LITE Assessment Model

As indicated in the left portion of the Figure 1, learner variables are included in many of our experiments. Ability, as measured by students' grades or more basic measures such as visual skills, can often serve as covariates. Also, learner variables such as gender, ethnicity, and learning style often interact with factors under examination. For example, in the usability study discussed below there was a great deal of variance among the small number of student participants based on motivation and engagement. In terms of methodology, projects generally proceed from basic experimentation involving manipulation of specific variables to more comprehensive and applied experiments carried out in the context of ongoing classes. A mediation category was recently added, which refers to the ways in which students use the technologies. Usability testing can be instrumental in this regard. As we will discuss below, this information can provide important insight into the results of a larger scale experimental investigation. Tracking methodologies and psychophysiological measures are also currently being used to capture aspects of the process of software utilization in some projects. As for outcomes, the focus is mainly foundational knowledge and problem-solving measures, due to the nature of the content material in projects where this model has been applied, which have primarily been science and engineering. Attitudinal and perceived outcomes are considered as well. Finally, many measurement approaches are included, as illustrated in Figure 1 and most are used over the course of a large scale project. In the illustrative example below we used measures of rote knowledge, basic computation, and self report. Within the context of this model, usability testing is conceived as one of these measurement approaches, though it could just as easily be viewed a research methodology. The illustrative example presented below includes a basic research experiment, in the form of the manipulation and examination of a specific characteristic of a software module, and a supplemental examination of processing in the form of usability testing of the same software module that was used in the basic research experiment.

ILLUSTRATIVE EXAMPLE

"Taking the Next Step in Engineering Education" Project

The University of Missouri-Rolla is currently conducting a comprehensive multi-year project sponsored by the U.S. Department of Education's FIPSE program

(#P116B000100) (Flori et al., 2002; Hall et al., 2002; Hubing et al., 2002; Philpot et al., 2002; Philpot et al., 2003) entitled “Taking the Next Step in Engineering Education: Integrating Educational Software and Active Learning.” The purpose of the project is to enhance learning through the creation and integration of a suite of courseware modules into three engineering courses: Statics, Dynamics and Mechanics of Materials. Examples of many of these modules are available at http://www.umr.edu/~bestmech. The project evaluation is being carried out under the auspices UMR’s Laboratory for Information Technology Evaluation (LITE) mentioned above.

Impact of Self-Testing Feedback

One of the first studies conducted within this project was an examination of the role of interactive feedback as a component of the software modules. There is evidence that providing learners with feedback as a component of instructional software systems can significantly enhance learning (Barron, 1996), particularly for novice learners (Dillon & Gabbard, 1998; Shin, Schallert, & Savenye, 1994). However, incorporating interactive feedback into instructional modules can often be resource and labor intensive since it requires the production of significantly more content in the form of problems and solutions. Further, developers are required to create a more complex and interactive learning environment. Therefore, it is important to establish, from the outset, whether adding these feedback components within the context of a project will be effective enough to warrant the extra resources required for development.

Basic Research Experiments

Method

Two experiments were conducted using two courseware modules. In both of these experiments, students were recruited from ongoing Mechanics of Materials classes to study courseware modules pertaining to course concepts that had not yet been discussed in class. In both experiments, students completed these experiments in computer labs in controlled conditions, observed by experimenters. After studying the target modules for thirty minutes, all students completed tests on the subject matter and responded to subjective quantitative questions.

In both experiments, students were randomly assigned to one of two groups, either a feedback group or a non-feedback group. For those in the feedback group, the study module included a self-quiz component, and for those in the non-feedback group the module did not include the self quiz. In experiment 1, students studied information pertaining to stress transformation equations and in experiment 2, they studied information on the Mohr’s circle procedure for stress transformations. Thirty-five students participated in experiment 1 (15 in the feedback group and 10 in the non-feedback group). Twenty-eight students participated in experiment 2 (16 in the feedback group and 12 in the non-feedback group). The experiment directions and the modules for both experiments and both groups are available at: http://campus.umr.edu/lite/feedback

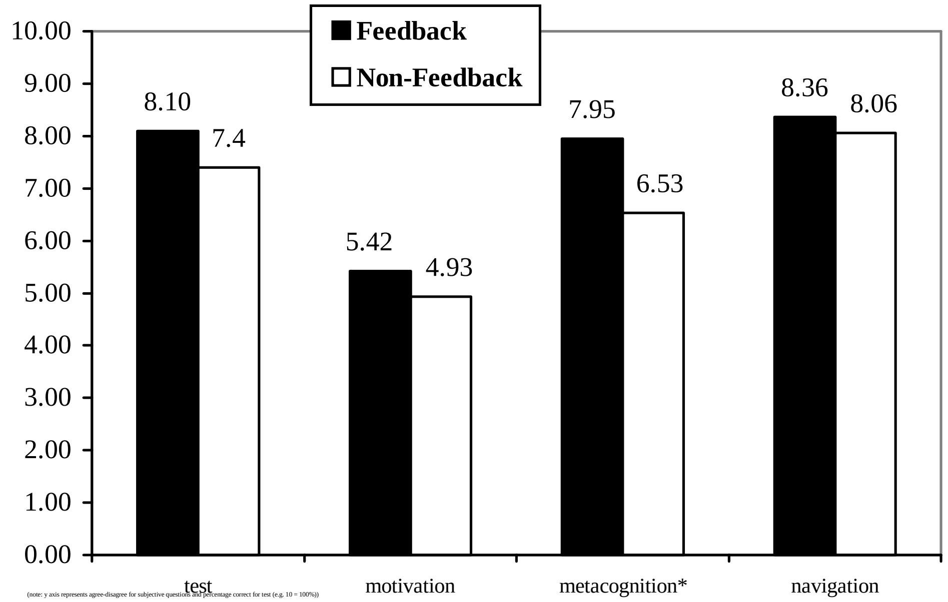
Outcome Measures

In each experiment, after studying the instructional software module, students completed a series of short-answer questions, which included completion, multiple choice, and true and false (15 questions in study 1 and 10 in study 2). In terms of the model above, these questions were principally measures of foundational knowledge. In addition, they responded to the following four questions using a 10-point Likert scale (1 = strongly disagree and 10 = strongly agree).

a. I learned a great deal of information from the multimedia tutorials. (learning)
b. I found the multimedia tutorials to be very motivational. (motivation)
c. The web tutorials were effective in aiding me in recognizing how much I know and don't know about this topic. (metacognition)
d. I found the navigational scheme for the web tutorials to be logical and easy to navigate. (usability-navigation)

Results

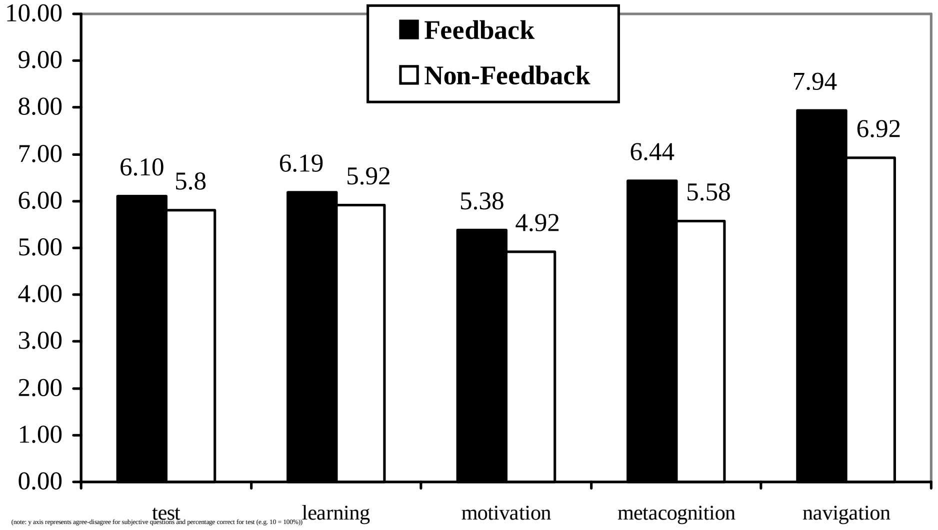
The quantitative analyses for each experiment consisted of a series of five between-subject t-tests with group (Feedback vs. Non-Feedback) serving as the independent variable. Test and Likert responses to each of the four subjective questions served as the dependent variables. The means for experiment 1 and 2 are listed in Figures 2 and 3 below. In all cases, the mean for the feedback group was higher. However, the only statistically significant mean difference was for the metacognition question in experiment 1.



(note: y axis represents agree-disagree for subjective questions and percentage correct for test (e.g. 10 = 100%))

*p < .05

Figure 2. Experiment 1 Mean test and subjective rating scores as a function of experimental group.



(note: y axis represents agree-disagree for subjective questions and percentage correct for test (e.g. 10 = 100%))

Figure 3. Experiment 2 Mean test and subjective rating scores as a function of experimental group.

Usability Testing

In order to gain insight into how students were using the software, we performed a usability study using the stress transformation module with feedback components. Ten students who had not participated in either of the two feedback experiments were recruited for the usability study. These students were given the same instructions as those students in the basic research experiments. They were also asked to complete the same qualitative measures using the same time schedule. Students carried out the task individually and in isolation in the LITE usability testing facility. Their activity on the screen was captured while their facial expressions were videotaped. These were mixed together to create one video for scoring that consisted of picture-in-picture of the video of the students facial expressions superimposed in the corner of the video screen capture of their activity.

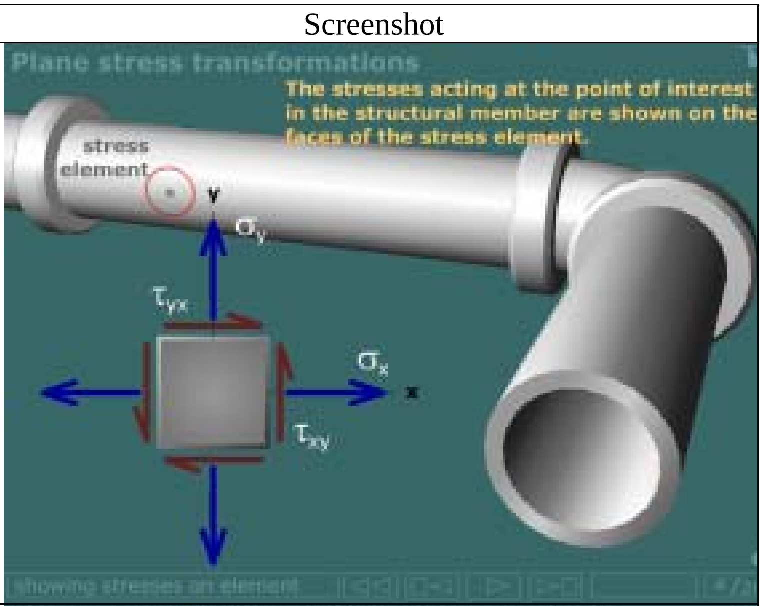
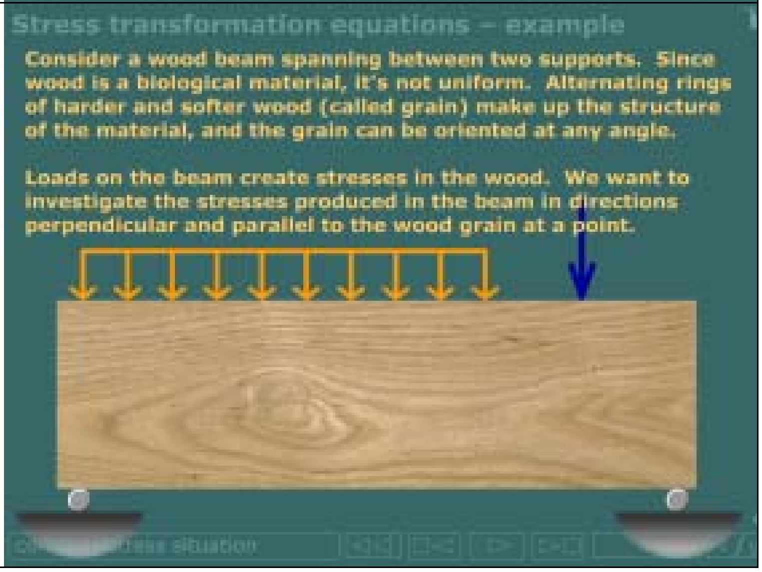
Scoring

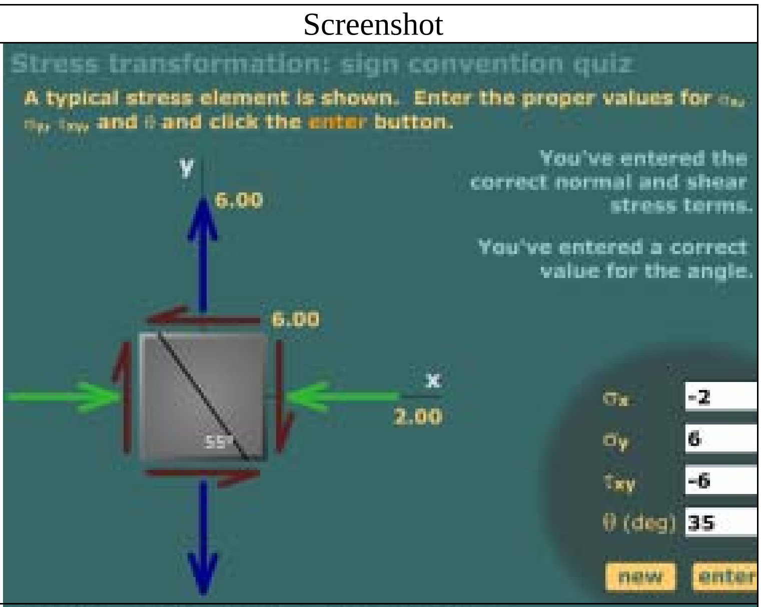
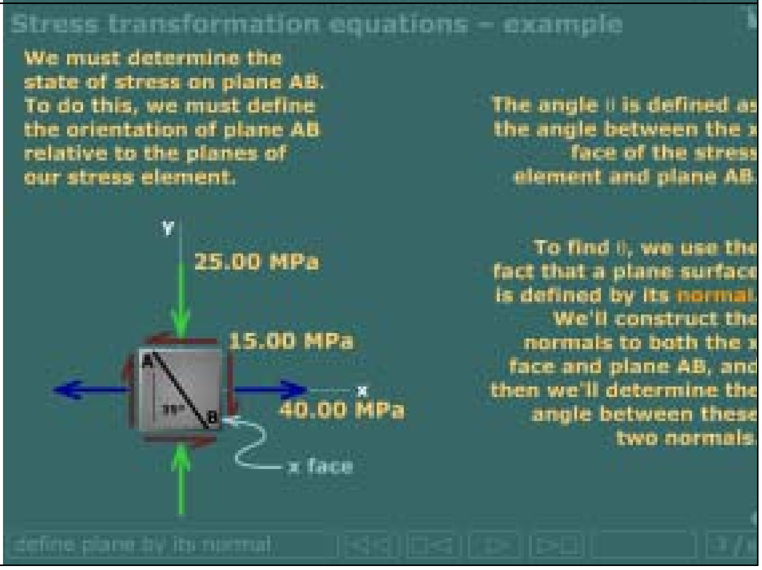
It is important to note that the instructional software module consisted of multiple movies, and most movies consisted of multiple scenes. The scoring scheme consisted of the following steps.

1. The time that students spent on each movie and each scene was recorded.
2. The videotape of each participant was viewed by two experimenters. Each experimenter recorded comments with a particular focus on (a) navigation, (b) level of engagement, and (c) level of frustration. The main goal was to gain insight into how students used the software.
3. Viewing duration, comments, quiz scores, quantitative ratings, and students' comments about quantitative ratings were collated as a summary for all students and for each individual student.

4. These data were reviewed and conclusions were drawn based on different aspects of the data.

General descriptions of the six movies and sample screen shots are presented in Table 1 below. Also, the movies along with the experimental protocol can be viewed at: http://campus.umr.edu/lite/feedback.

Movie	Description	Screenshot
1	This is a movie detailing the theory pertaining to stress transformations. The movie contained 30 scenes describing the importance and relevance of the topic, the terminology used, and a complete derivation of the equations.	 Plane stress transformations The stresses acting at the point of interest in the structural member are shown on the faces of the stress element.
2	This movie is a basic example problem that illustrates the application of stress transformation theory in the context of a familiar, everyday object such as a wood beam.	 Stress transformation equations – example Consider a wood beam spanning between two supports. Since wood is a biological material, it's not uniform. Alternating rings of harder and softer wood (called grain) make up the structure of the material, and the grain can be oriented at any angle. Loads on the beam create stresses in the wood. We want to investigate the stresses produced in the beam in directions perpendicular and parallel to the wood grain at a point.

Movie	Description	Screenshot
3	This interactive movie focuses on a single trouble spot: proper sign conventions. Students are shown a typical situation, and they are required to enter the appropriate numeric values. The movie provides feedback for each answer, and students can continue the quiz until they are confident in their understanding of the proper sign conventions.	 Stress transformation: sign convention quiz A typical stress element is shown. Enter the proper values for σ_x, σ_y, τ_{xy}, and θ and click the enter button. You've entered the correct normal and shear stress terms. You've entered a correct value for the angle. σ_x -2 σ_y 6 τ_{xy} -6 θ (deg) 35
4	This interactive movie also focuses on a single trouble spot: the correct value for a particular angle used in the stress transformation equations. This movie uses few words but lots of animation. The tone of the movie is mildly humorous.	 Finding θ for stress transformations **Top** Start at the topmost corner of the inclined plane. **Sweep the Clock** At the top corner of your right triangle, sweep the angle from the vertical line to the hypotenuse. Start at 6 o'clock and turn the angle to the inclined plane. **Drop** Drop a vertical line from the topmost corner and make a right triangle with the inclined plane as the hypotenuse. **Sign of θ** If the angle sweeps in counterclockwise direction, θ is positive. If the angle turns in a clockwise direction, θ i negative.
5	This movie is a basic example problem that illustrates the application of stress transformation theory.	 Stress transformation equations – example We must determine the state of stress on plane AB. To do this, we must define the orientation of plane AB relative to the planes of our stress element. The angle θ is defined a the angle between the face of the stress element and plane AB. To find θ, we use the fact that a plane surface is defined by its normal. We'll construct the normals to both the face and plane AB, and then we'll determine the angle between these two normals.

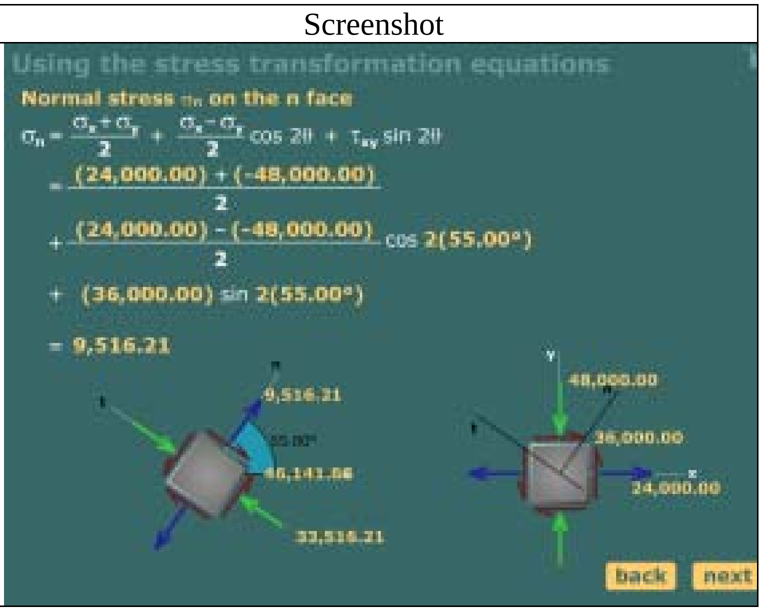
Movie	Description	Screenshot
6	This movie is a learning tool. Students enter a set of stress values, and the movie illustrates how to calculate various properties using the stress transformation equations. The application and solution process is animated, both with equations and with pictures.	 Using the stress transformation equations Normal stress σ_n on the n face $\sigma_n = \frac{\sigma_x + \sigma_y}{2} + \frac{\sigma_x - \sigma_y}{2} \cos 2\theta + \tau_{xy} \sin 2\theta$ $= \frac{(24{,}000.00) + (-48{,}000.00)}{2}$ $+ \frac{(24{,}000.00) - (-48{,}000.00)}{2} \cos 2(55.00°)$ $+ (36{,}000.00) \sin 2(55.00°)$ $= 9{,}516.21$

Table 1. Movie Summary and Sample Screen Shots.

Summary of time spent on movies

Movie	Average Time Spent	Number of Subjects
1	12.16	10
2	4.08	4
3	9.21	4
4	1.26	4
5	0.43	4
6	1.38	3

Table 2. Average time and total number of students on each of the movies

Notes and Conclusions

- Most users went only to movie 1. (Theory of Stress Transformations Including Derivations of Key Equations).
  - This behavior suggests that students have a strong tendency to proceed in a linear fashion and/or they tend to focus on the equations.
- Movie 3 was second most popular. This movie is interactive with feedback and focuses on a common trouble spot for students, which is determining sign conventions.
  - This behavior suggests that students recognize the effectiveness of interactive feedback and recognize specific problem areas.
  - This behavior is also further indication that students tend to proceed in a linear fashion since this was the first movie they would encounter that presented a specific problem area along with feedback.

Percent of time spent on movies as a function of student.

Student	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5	Movie 6	Quiz
1	22.64	25.62	28.71	5.16	2.92	14.96	86.67
2	77.13	2.37	13.09	4.94	2.47		66.67
3	17.46	2.31	77.43	.85	.12	1.82	73.33
4	100.00						100.00
5	100.00						73.33
6	100.00						100.00
7	100.00						93.33
8	62.81	12.36	12.42	8.31	3.93	.17	100.00
9	100.00						93.33
10	100.00						86.67

Table 3. Percent of time spent on each movie and
quiz score (also percentage) for each subject.

Notes and Conclusions

- No relationship was apparent between the quiz score and either the number of movies watched or the amount of time spent on certain movies.

Average time spent on individual scenes within movies.

Scene/ Movie	1	2	3	4	5	6	7	8	9	10
1	1.50	1.27	.93	.75	2.08	1.33	1.16	0.81	1.22	2.26
2	1.56	.58	2.26	.75	.46	.75	2.31	1.74	0.46	-
3	32.47	-	-	-	-	-	-	-	-	-
4	.46	.46	.93	.69	.75	.75	0.93	0.23	-	-
5	.52	.58	.41	.35	.35	.41	0.46	0.12	-	-
6	5.67	-	-	-	-	-	-	-	-	-

Scene/ Movie	11	12	13	14	15	16	17	18	19	20
1	1.27	1.97	2.43	2.60	5.32	2.66	4.51	4.05	1.45	3.01

Table 4. Average percentage time spent on scenes within movies. (Dashes indicate the given movie did not have these scenes. Note that only movie 1 had more than 10 scenes).

Notes and Conclusions.

- Scene 15, 17, & 18 on movie 1 were the most popular. These movies integrated equations with the corresponding graphical representations.

Individual students.

Users in this study could be grouped into two archetypes: the engaged student and the non-engaged student. As mentioned above, experimenters provided notes about navigation, frustration, and other reactions for each student. These data along with quantitative information for two dissimilar students are presented below to illustrate these archetypes.

Example of Engaged Student:

Student	**Movie 1**	**Movie 2**	**Movie 3**	**Movie 4**	**Movie 5**	**Movie 6**	**Quiz (Points)**	**Quiz (Percentage)**
8	18.38	03.40	03.41	02.28	01.10	00.03	15	100

Navigation Summary:

The user started with Movie 1, went through the scenes until scene 14 in a sequential order, spending and equal amount of time on each scene, then came back to scene 5 in descending order of the scenes and then went to scene 20 in ascending order. The user started the movie again from scene 1 and went to scene 20 in order. In the process he spent a large amount of time on 16, 17 & 18. Then the user proceeded to Movie #2 and spent fair amount of time on each scene. Then the user went to Movie #3 and entered values 6 times, mostly correctly. Then went to Movie #4 and again spent an even amount of time on all scenes, a little more time on scene 7. Then, he went to Movie #5 and watched it quickly.

Notes and Conclusions:

- User proceeded systematically, varying speed when necessary.
- Focused on aspects of module that involved integration of graphics and equation (scenes 16, 17, & 18)
- Completed quiz after reviewing important information.

Example of Non-Engaged Student:

Student	**Movie 1**	**Movie 2**	**Movie 3**	**Movie 4**	**Movie 5**	**Movie 6**	**Quiz (Points)**	**Quiz (Percentage)**
3	04.47	00.38	21.13	00.14	00.02	00.30	11	73.33

Navigation Summary:

The user started by reading the instructions and then proceeded to the Movie 1. The user came back to scene 3 from scene 4; otherwise he followed the order of the movie. The user spent time on scenes 10 & 15, but generally he watched the movie pretty quickly, averaging 13 seconds on each page of the movie. Then, the user went to Movie #2, watching that movie in a sequential manner and spending less than 5 seconds on each scene. Then, the user went to Movie #4 and spent only 2 seconds on each scene of the movie. The user watched the Movies #5 & #6 also very quickly spending 2 seconds and 30 seconds respectively. The user went to Movie #3 after completing all the movies and spent almost 22 minutes on this movie. Then, finally he took the quiz entering numbers quickly over and over usually getting the answer incorrect.

Notes and Conclusions.

- User clicked through the scenes largely in order.
- User spent little time on any one scene.
- User did not complete quiz until the end without much apparent thought.

CONCLUSIONS

The supplemental usability study provided a number of insights about how students' used this software. First, they tended to proceed in a linear fashion, despite the fact that a menu allowed for non-linear navigation. Second, students tended to focus on the first movie for much of the allotted time. In fact, the majority of those who participated spent all of their time on this movie. Third, those who did go beyond the first movie tended to gravitate to the movie that included an interactive component and focused on a specific problem that tends to be a trouble spot for students. This indicates that a minority, but still substantial number, did exhibit a strategic awareness. These students recognized the potential learning efficacy of the interactive components of the areas where they needed the most practice. Fourth, there was not an apparent relationship between the number of movies watched or the time spent on any given movie and performance. This indicates that the factors in processing of the software that accounted for the best performance were more subtle and complex than simply the number of movies watched or time spent on a given movie. Fifth, the scenes within the first movie that were most popular were those that combined equations with three-dimensional graphical images. This provides some evidence that the graphical content within movies such as these can provide a scaffolding context for the equations used to solve engineering problems. Sixth, there was a dramatic degree of variance evidenced among these ten participants. In particular, engaged and non-engaged learning patterns emerged. Though both types of students tended to progress linearly, the engaged users varied their speed as necessary and focused particularly on the sections involving integration of graphics and equations. On the other hand, the non-engaged learner tended to quickly advance through the movies, spending very little time on any particular scene.

Usability testing provided clear potential explanations for the results of the basic research experiment. One of the most dramatic insights was that most students did not even get to the scene that included the quizzes. If one assumes that the students in the basic research and usability studies used the software in a similar manner, the impact of feedback in explaining group differences should be much less than originally assumed. This finding provides a potential explanation for why those in the feedback group scored higher on all measures, but not significantly higher on any. The students who did not get to the quiz portion of the software may have diluted the higher scores for those in the feedback group. Also, the usability study indicated that the variance among students was fairly dramatic with respect to the students' motivation and engagement. This variance may have created a large degree of within-group error in the experiment, again contributing to the non-significant effects.

The usability study also provided information that has helped to guide further software development and design, and in design of subsequent experiments. First, the large amount of variance in usage may be partly due to the lack of guidance and control integrated into these software packages, which implies that some degree of navigational guidance may well encourage more engagement and focus among the users. Students who performed the best tended to focus on scenes that involved the integration of equations and visual information. The navigation could be designed to encourage all students to focus more on these areas. Second, it appears that the material included in this module included far more information than most of the students could study thoroughly in 30 minutes. As for software design, this finding implies that it may work best to break this module down into smaller modules to maximize learning. With respect to experimental design, it implies that smaller modules should be used for these types of experiments, or more time should be allotted.

REFERENCES

Barnum, C. M. (2002). *Usability Testing and Research*. New York, NY: Allyn & Bacon.

Barron, A. E. (1996). Design Guidelines for the World Wide Web. *Journal of Interactive Instruction Development, 8*(3), 13 - 17.

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.

Dillon, A., & Gabbard, R. (1998). Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research, 68*, 322 - 349.

Dumas, J. S., & Redish, J. C. (1999). *A Practical Guide to Usability Testing*. Exeter, England: Intellect.

Eller, V. E., Hall, R. H., & Watkins, S. E. (2001). Multimedia web-based resources for engineering education: The media design and assessment laboratory at UMR. *Proceedings of American Society for Engineering Education Conference*, http://www.asee.org/conferences/search/00713_02001.pdf.

Flori, R. E., Oglesby, D. B., Philpot, T. A., Hubing, N., Hall, R. H., & Yellamraju, V. (2002). Incorporating Web-Based Homework Problems in Engineering Dynamics. *Proceedings of American Society of Engineering Education Conference*, http://www.asee.org/conferences/caps/document2/2002-2076_Paper.pdf.

Hall, R. H., Philpot, T. A., Oglesby, D. B., Flori, R., Hubing, N., Watkins, S. E., et al. (2002). A model for the evaluation of innovative engineering courseware: Engineering an assessment program. *Proceedings of American Society for Engineering*

Education Conference, http://www.asee.org/conferences/caps/document2/2002-1697_Paper.pdf.

Hall, R. H., Watkins, S. E., Davis, R. L., Belarbi, A., & Chandrashekhara, K. (2001). Design and Assessment of Web-Based Learning Environments: The Smart Engineering Project and the Instructional Software Development Center at U.M.R. In L. R. Vandervert & L. V. Shavinina (Eds.), *Cybereducation: The Future of Long Distance Education* (pp. 137 - 156). New York: Mary Ann Liebert, Inc.

Hubing, N., Flori, R., Hall, R. H., Oglesby, D. B., Philpot, T. A., & Yellamraju, V. (2002). Interactive Learning Tools: Animating Statics. *Proceedings of American Society for Engineering Education Conference*, http://www.asee.org/conferences/caps/document2/2002-1533_Paper.pdf.

Law, L., & Hvannbert, E. T. (2002). Complementarity and Convergence of Heuristic Evaluation and Usability Test: A Case Study of UNIVERSAL Brokerage Platform. *Proceedings of NordiCHI*, 71-80.

Mackay, W. E., & Fayard, A. (1997). HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines. *Proceedings of Conference on Designing Interactive Systems*, 223-234.

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.

Nielsen, J. (2003). Usability 101. *useit.com, Aug. 2003*, http://www.useit.com/alertbox/20030825.html.

O'Malley, J. M., & Valdez, P. L. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. Reading, MA: Addison-Wesley.

Philpot, T. A., Flori, R., Hall, R. H., Hubing, N., Oglesby, D. B., & Yellamraju, V. (2002). Interactive learning Tools: Animating mechanics of materials. *Proceedings of American Society for Engineering Education Conference*, http://www.asee.org/conferences/caps/document2/2002-1527_Paper.pdf.

Philpot, T. A., Hall, R. H., Hubing, N., Flori, R., Oglesby, D. B., & Yellamraju, V. (2003). Animated instructional media for stress transformations in a mechanics of materials course. *Computer Applications in Engineering Education, 11*, 40-51.

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.

Shin, E., Schallert, D., & Savenye, C. (1994). Effects of learner control, advisement, and prior knowledge on young students' learning in a hypertext environment. *Educational Technology Research and Development, 42*, 33 - 46.

Strauss, A., & Corbin, J. (1990). *Basics of Qualitative Research*. London: Sage.

Wiggins, G. (1998). *Educative Assessment*. San Francisco, CA: Jossey-Bass.

Acknowledgement

This work was supported in part by a grant from the United States Department of Education Fund for the Improvement of Post-Secondary Education (FIPSE #P116B000100).