A Model-Driven Multi-Year Assessment of a Software Development Project for Engineering Instruction

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

University of Missouri – Rolla

1. Abstract

This paper is a review of a series of evaluation studies that were utilized to inform and evaluate a large scale instructional software development project at the university of Missouri – Rolla entitled "Taking the Next Step in Engineering Education: Integrating Educational Software and Active Learning." This project was funded by the U.S. Department of Education Fund for the Improvement of Post Secondary Education (FIPSE), and was carried out over the last four years. The assessment was carried out under the auspices of UMR's Laboratory for Information Technology Evaluation (LITE), and guided by the LITE model for evaluation of learning technologies. The fundamental premise of the model is that evaluation should consist of the triangulation of multiple research methodologies and measurement tools. Five representative evaluation studies, consisting of eight experiments, are presented here. The studies range from initial research consisting of basic experimentation and usability testing; to applied research conducted within the class room; to a large multi-nation cross-cultural applied-dissemination survey conducted during the last semester of the project. The results indicate that the instructional multimedia developed in this project can have a substantial positive impact in enhancing fundamental engineering classes. Further, the research also indicates that the LITE model can be an effective tool for guiding a comprehensive evaluation program.

2. "Taking the Next Step in Engineering Education" Project

The University of Missouri–Rolla (UMR) recently completed a comprehensive three-year project sponsored by the U.S. Department of Education's FIPSE program [1-5] entitled "Taking the Next Step in Engineering Education: Integrating Educational Software and Active Learning" (#P116B000100). The project focused on enhancing learning in three core engineering courses—Statics, Dynamics and Mechanics of Materials—through the development and implementation of a suite of animated and interactive courseware modules. More than 250 computer-based instructional examples, problems, games, and theory modules and an extensive homework database administration system were developed. Examples of many of these modules are available at http://www.umr.edu/~bestmech.

The project evaluation was carried out under the auspices of UMR's Laboratory for Information Technology Evaluation (LITE). Several assessment and evaluation studies have been carried out from the beginning of the project, representing diverse methodologies and measurement tools, based on an assessment framework described below.

3. 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 [4, 6, 7]. 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.

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 reviewed in this paper, encompasses many parts of the model. Figure 1 is a graphical depiction of the LITE learning technology assessment model.

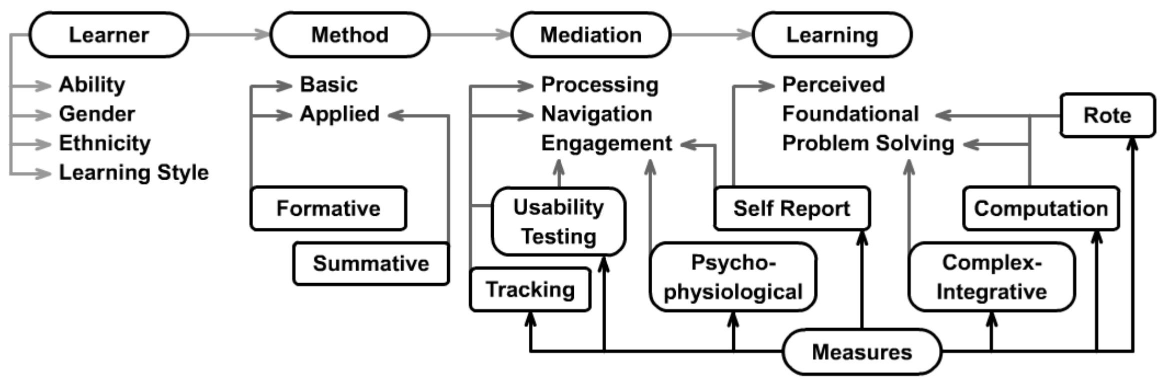

Figure 1. LITE Assessment Model

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 the multi-national survey we will discuss at the end of the paper, the nationality of the school (U.S. vs. international) had a significant impact in mediating the effect of the software.

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. We will present two examples of basic research studies, where students were randomly assigned to groups, and a controlled environment was used; a usability test where a small group of students were examined in great detail as they used the software; two applied research studies, conducted within the context of ongoing classes, and, finally a survey, which had very little experimental control, in comparison, but allowed for much more comprehensive generalization across learning contexts, nationalities, and cultures.

A mediation category was recently added to our model, which refers to the ways in which students use the technologies. Usability testing can be instrumental in this regard, in that it can provide important insight into the results of a larger scale experimental investigation as we'll see in the example study below. Tracking methodologies and psychophysiological measures are also currently being used in the lab to capture aspects of the process of software utilization in some projects. Though we did not use these methods in the experiments reported here, we have had success in using these types of tools for assessing the impact of learning environments in other projects [8].

As for outcomes, the focus of objective outcomes is mainly foundational knowledge and problem-solving, 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 often considered as well. Finally, many measurement approaches are included, as illustrated in Figure 1 and most are used over the course of a project such as that reported here.

4. Study 1: Effect of Interactive Feedback (Basic Experimentation)

4.1 Rationale

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 [4]. There is evidence that providing learners with feedback as a component of instructional software systems can significantly enhance learning [9], particularly for novice learners [10, 11]. 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 was important in this project to establish, from the outset, whether adding these feedback components would be effective enough to warrant the extra resources required.

4.2 Experimental Procedures

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 brief 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. The study module consisted of animated movies that explained the relevant theory and specific aspects important in successfully applying the theory. For students in the feedback group, the study module also included a self-quiz component. Students in the non-feedback group were not given access to the self-quiz. The self-quiz component consisted of an interactive game feature in which the student was asked to respond to a series of questions, applying the information presented in the study module.

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*.

At the completion of each experiment, all students completed a series of short-answer questions, which included completion, multiple choice, and true/false (15 questions in study 1 and 10 in study 2). In terms of the LITE model above, these questions were principally measures of foundational knowledge. In addition, they responded to the following four questions using a 9-point Likert scale (1 = "strongly disagree" and 9 = "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)

They were also asked to provide open-ended comments to support their responses to each of these items and to provide additional comments on the software effectiveness and/or suggestions for improvement.

4.3 Results

4.3.1 Quantitative Analysis

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. In all cases, the mean for the feedback group was higher. The only statistically significant mean difference, however, was for the metacognition question in experiment 1.

4.3.2 Qualitative Analysis

Students' open-ended responses to each of the questions were examined with special emphasis on identifying differences between groups. Several themes emerged from the qualitative data. Three of the major themes and representative comments are provided. (A more detailed description of this analysis is provided elsewhere [4]).

(1) Students in the feedback groups were more positive about their perceived learning and especially about the degree to which it made them aware of their knowledge level (metacognition).
   - *(feedback):* Made me realize I didn't know anything.
   - *(non-feedback):* I am not sure what I did right and what I did wrong

(2) The modules could serve as a good supplement to class, especially for a student who was having difficulty.
   - I have some previous knowledge of the material but the program offered some insightful hints that will prove very beneficial as supplemental material.
   - If I were unable to comprehend what was going on in class this would be a good place to seek info.

(3) The animations, graphics, and self-paced nature of the modules added a dimension that aided learning in a way that other methods could not accomplish.
   - … it allowed you to go at your own pace.
   - Pictures and "movies" make it more interesting.
   - Nice to learn at your own pace and to go back and re-read.

4.3.3 Conclusions

In the quantitative results, the feedback and non-feedback groups differed significantly on only one of the ten outcome measures. Those in the feedback group in experiment 1 rated their software significantly higher in the item referring to impact on metacognitive awareness. Despite this lack of significant effects, it is important to note that students in the feedback group scored higher on both quizzes and rated the software more positive on all subjective measures, so the means were consistently in a direction favoring the software with feedback. In addition, student comments for those in the feedback groups were more positive, and specifically pointed to advantages associated with the interactive components of the software. For this reason, it was concluded that interactive feedback should be an important part of subsequent software design. We also suspected, based on informal observation during the experiments, that students were using the software in very diverse ways, so that the impact of the feedback components may be strongly mediated by the way in which the software was used. This suspicion led to a more detailed analysis of students' use of the software in the usability study that follows.

5. Study 2: Detailed Analysis of Software Use (Usability Testing)

5.1.1 Description

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.

5.1.2 Results

5.1.2.1 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.

5.1.2.2 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.

5.1.2.3 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.

5.1.2.4 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.

5.1.2.5 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.

5.1.2.5.1 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.

5.1.2.5.2 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.

5.1.3 Conclusions

The 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 tended 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.

6. Study 3: Using Games to Teach Statics (Applied: Class Context)

6.1 Rationale

Fundamental engineering courses such as Statics (one of the courses targeted in the project) seek to develop the student's ability to analyze basic engineering machines, mechanisms, and structures and to determine the information necessary to properly design these configurations. Fundamental calculations such as centroids and area moments of inertia are building blocks that students must employ to solve problems and develop designs in a variety of situations.

It is often assumed that repetition leads to proficiency; however, few students relish working dozens of problems on a particular topic. To make the learning process more enjoyable, repetition and drill on a specific topic can be encapsulated in a game context. Games have been found to be an effective method of increasing motivation, enjoyment and learning for many math and science topics that may otherwise seem boring to students [12-15]. There is evidence that such tools can be a particularly powerful for learning engineering concepts where visualization is important, such as engineering graphics [16]. Through the challenge of the game, the student can receive the benefits of repetition without the sense of labor that they might feel otherwise. A game context provides students with a structure for learning and permits students to develop their skills at their own pace in a non-judgmental but competitive and often fun environment. Since the computer is a medium that is well suited for repetitive processes and for numeric calculations, computer-based games focused on specific calculation processes offer great potential as a new (or perhaps updated) type of learning tool for engineering mechanics courses. Therefore, a number of interactive games were developed as part of this project [17], and three experiments were conducted to examine the impact of this software within the context of ongoing classes. Two specific game modules, *The Centroids Game – Learning the Ropes* and *The Moment of Inertia Game – Starting from Square One*, were evaluated.

6.2 Experimental Procedures

In the 2002 academic year, the effectiveness of two games as teaching tools was assessed with two undergraduate Statics classes at UMR. In lieu of the customary lecture period, students were taken to a computer lab where a computer was available for each student. During the preceding class period, students had been introduced to the topic. At the start of the assessment class period, students were given a two-minute introduction to the relevant calculation procedure and then given approximately 40 minutes to play the game at their own pace. An instructor was present in the computer lab to answer questions about the topic and to clarify game procedures. Near the end of the class period, students were stopped and asked to complete a questionnaire in which students responded to Likert-type statements using a 9-point scale where 1 = "strongly disagree" and 9 = "strongly agree". The questionnaire used in assessing *The Centroids Game* is shown below, and a similar questionnaire was used for *The Moment of Inertia Game.*

1. After using *The Centroids Game*, I felt confident in my ability to calculate centroids for composite bodies.
2. After using *The Centroids Game*, I was able to visualize the procedure for calculating centroids.
3. After using *The Centroids Game*, I understood which cross-sectional dimensions to include in my calculations when working a centroids problem.
4. *The Centroids Game* helped me to recognize how much I know and don't know about the procedure for calculating centroids.
5. I found *The Centroids Game* to be motivational concerning the procedure for calculating centroids.
6. I liked playing a game to help me get better at calculating centroids.
7. I learned a great deal about the procedure for calculating centroids from *The Centroids Game*.
8. I learned a great deal about the procedure for calculating centroids from my Statics textbook (Spring 2003 only).
9. I thought the time spent playing *The Centroids Game* was a worthwhile use of my study time.
10. The procedure for playing *The Centroids Game* was easy to understand.
11. The number of questions and the number of rounds used in *The Centroids Game* seemed about right to me.
12. Give your overall evaluation of *The Centroids Game* on the procedure for calculating Centroids, using the 1…9 scale, with 1 being very poor and 9 being outstanding.

6.3 Results from Game Assessment

6.3.1 Survey

For both game assessments, the survey results are summarized in Table 1 for both Fall and Spring Statics classes. Students were also asked to comment on their overall evaluation of the games. Their comments were consistently positive, as characterized by representative comments such as:

- "I think it's a much easier way to do homework and I did 10 times as many problems as I normally do. I have this concept down very well."
- "Easy to understand. Helps to teach by progression…easy-to-hard."
- "It showed me everything I didn't know and allowed me to learn."
- "Most fun I've had while learning in a long time."

Two open-ended questions were included in *The Moment of Inertia Game* questionnaire to explore students' perceptions of instructional software in general, particularly after having just had an experience with the game.

1. Are there things you really dislike about instructional software? Do you think software is a waste of time or just no-good? What really bugs you about this stuff?
2. Are there things that you really like about instructional software? Have you tried instructional software? Are there any programs that you think are really good?

Students' responses to these open-ended questions were combined and categorized according to themes. Two themes that emerged from students' comments and some representative student comments are presented below:

Theme 1. Students felt very positive about instructional technology in general and *The Moment of Inertia Game* in particular. The principle advantages cited were (a) immediate feedback, (b) aid in visualization, and (c) increase in motivation and enjoyment.

(a) *Immediate Feedback*
- "It's a great way to do homework and it gives you the correct answers right away – that way I KNOW I'm doing it right, every time."

(b) *Aid in Visualization*
- "If it is good visually and outlines steps, it can be very helpful."

(c) *Increase motivation and enjoyment*
- "I enjoyed it thoroughly. I like the competitive view, try to get the better score."

Theme 2. It is important that instructional software is integrated with the class and instructor.

- "I like it in class if the prof is walking around helping."
- "I think it (instructional technology) is a good idea, but must be assigned in class."

6.3.2 Comparison with Textbook

To compare student ratings of the game with their textbook, survey statement 8 was added to the Spring 2003 questionnaire. The responses to statement 7 were compared with the responses to statement 8, using a within-subjects t-test. In both *The Centroids Game* assessment [$t(22) = 10.098$, $p < 0.001$] and *The Moment of Inertia Game* assessment [$t(22) = 6.86$, $p < 0.001$], this test was statistically significant. On a scale of 9, students' agreement with the statement that they learned a great deal from the game was more than two to four times higher than their rating of the same statement for the textbook.

6.3.3 Impact of Game on Learning

In the Spring 2003 assessment experiment, a single-problem quiz was administered to students at the end of the class period following completion of *The Centroids Game* exercise. To serve as a control group, students in four additional Statics sections were also given the same quiz. None of the students in the control groups had exposure to *The Centroids Game*. Students in the control group took the quiz either one or two class periods after the topic of centroids of composite areas had been discussed in lecture. Students in the control group, therefore, had some opportunity to review notes and work assigned homework problems in the days following their in-class exposure to this topic. Students in both the experimental and control groups, however, were not told about the quiz before the class period in which it was administered.

Students were asked to compute the vertical location of a centroid for a double-tee shape. For the purpose of this experiment, quizzes were marked either *100% correct* or *incorrect* (for any type of error). The results of the quiz are shown in Table 3.

An analysis was conducted to compare problem scores for students in the test group with those in the control group. Since these data consisted of dichotomous data, a Pearson Chi-Square was computed to test for significant differences in the distribution of correct and incorrect responses between the groups (test vs. control). This test was statistically significant, indicating that those in *The Centroids Game* group performed significantly better on the quiz problem than those in the control group.

To compare students who used *The Moment of Inertia Game* to those who learned in a traditional lecture, the test class was compared with a control group of three Statics classes that had not used the game. Students in both the test group and the control group were given a brief quiz at the beginning of the class period after moments of inertia for composite areas had been presented, either by the game or in a lecture. Students in both groups, therefore, had some opportunity to review notes and work assigned homework problems in the two days following their in-class exposure to this topic. This assessment differed from *The Centroids Game* assessment in that students in both the test and control groups were told in advance about the upcoming quiz. Students were asked to compute the area moments of inertia I_x and I_y for a tee-shape about both the horizontal and vertical centroidal axes. The vertical location of the centroid was explicitly given. Quizzes were graded and grouped into three categories: *100% correct* if the student correctly determined both I_x and I_y, *partially correct* if the student correctly determined either I_x or I_y or if they simply made a calculation error while performing the correct procedure, or *100% incorrect* if the student did not demonstrate understanding of the proper calculation procedure. The results of the quiz are shown in Table 4.

A Pearson Chi-square analysis was again used to test for statistical significance between the distributions of scores for those in the test group versus those in the control group. The Chi square test was statistically significant $X^2(2) = 10.71$, $p < .01$. The frequencies displayed in Table 3 indicate that the significant Chi square was due to the fact that virtually all of the students in the test group scored 100% correct on the quiz while over half of the students in the control group received partially correct or 100% incorrect.

6.3.4 Conclusions

This assessment was one of the strongest and most positive with respect to indicating the software's effectiveness. Students' responses to all questionnaire items were very positive in comparison to the scale midpoint of 4.5 (Table 2). In addition, students rated the game as significantly and overwhelmingly more effective than the textbook. Finally, student scores on objective measures of their learning were significantly higher for those who used the games in two different experiments.

The consistent and positive findings in this study are also indicative of improvements in the software design that were informed by initial research, and reflect the improving nature of the learning technologies informed by iterative evaluation.

7. Study 4: Instructional Multimedia as Support for a Traditional Lecture in Statics (Applied: Class Context)

7.1 Rationale

Studies conducted for the project such as the games experiments presented above provide strong evidence that the software can serve as an effective alternative to traditional lecture. This is consistent with other research [18, 19], which indicated that these tools can be effective substitutes, so long as they are well designed and used in the proper context [20, 21]. However, very few investigations have examined these tools as enhancements for a traditional lecture. In the classroom, software instructional aids can potentially free the lecturer from writing out individual problems step-by-step on the board, which allows time for higher-level interactive discussion. In addition, these types of modules have the potential to allow for elaborate multi-dimensional dynamic visualizations, which go far beyond a two-dimensional diagram or drawing, providing the student with a much more realistic and concrete representation of spatial concepts [22, 23]. Therefore, an assessment experiment was conducted which focused on the use of a multimedia module as support for a "traditional" classroom lecture.

7.2 Experimental Procedures

In the fall of 2002, one Statics instructor gave a lecture to two Statics classes on the topic of truss analysis using the method of sections. In the control class, the instructor used "traditional" tools (chalkboard) to support the lecture, and presented the theory followed by example problems. The experimental class also included a discussion of theory followed by example problems, but the examples were provided in the form of multimedia modules via a computer and projector. The multimedia modules allowed the instructor time to cover the examples and to cover multiple choice questions included in the module in an interactive format with the class. The class, as a whole, determined the answers to the multiple-choice questions through discussion lead by the instructor. In both classes, the same total amount of class time was devoted to the topic. There were 24 students in the control class and 27 in the experimental class.

Students completed a quiz over the material presented in the lecture, following the lecture. This quiz consisted of three multiple-choice questions, each worth three points. After completing the quiz, students responded to the following seven questions, using a 10-point Likert scale (1 = strongly disagree and 10 = strongly agree).

1. I learned a great deal of information from today's lecture/presentation on the method of sections for analyzing trusses. *(learning)*
2. I found today's lecture/presentation helped me to better visualize how to apply the method of sections to trusses. *(visualization)*
3. I found today's lecture/presentation helped me to better understand which equation to choose when working a truss problem with the method of sections. *(understanding)*

4. Today's lecture/presentation on the method of sections helped me to recognize how much I know and don't know about this topic. *(metacognition)*
5. I found today's lecture/presentation on the method of sections for analyzing trusses to be motivational. *(motivation)*
6. Today's lecture/presentation on the method of sections helped me to see how this technique has "real world" engineering applications. *(application)*
7. Give your overall evaluation of today's lecture/presentation on the method of sections, using the 1 … 10 scale, with 1 being very poor and 10 being outstanding. *(overall)*

Students were also asked to provide open-ended explanations for their ratings to each of these questions.

7.3 Results

7.3.1 Quantitative Analysis

An examination of the distribution of quiz scores indicated that there was very little variance, with the majority of students (40) receiving a perfect score (9) and the other 11 students receiving a score of 6. Given the strong ceiling effect, it was not surprising that further analysis revealed no statistical difference between the experimental and control classes.

A series of seven between-subjects t-tests were computed to compare the groups on questionnaire ratings. Group (experimental vs. control) served as the independent variable in each t-test while one of the seven questionnaire items served as the dependent variable. The means for these three items are displayed in Figure 5.

The groups differed significantly in their ratings for Question 7, $t(48) = 2.51$, $p < .05$, which was the rating of overall effectiveness. In addition, the differences were marginally significant on question 2, $t(48) = 1.65$, $p = .10$, which asked students to rate the effectiveness of the lecture to help them visualize, and question 6, $t(48) = 1.79$, $p = .08$, which asked students to rate the applicability of the lecture to "real world" engineering. In each case, the experimental group's mean was higher.

7.3.2 Qualitative Analysis

Students' open-ended responses were examined, labeled, and categorized. Categories that emerged with a substantial number of comments were considered themes. Representative comments from students in the experimental group are presented below.

Theme 1: Those in the experimental group were more positive, overall, about the lecture.

- I learned a lot today.
- The lecture technique for me is very beneficial.
- I understood the lecture very well and it helped me understand the method of sections.

Theme 2: Those in the experimental group found that the lecture to aid in visualization of the concepts.

- The computer animation helped me see how to section a truss …
- Computer diagrams are neater and easier to see what is going on.
- The animation made it easy to see how the truss is broken apart and analyzed.

Theme 3: Those in the experimental group found that the interactive nature of the lecture encouraged them to remain engaged.

- Making us take a quiz at the end kept me awake despite how tired I am.
- I think I may have found other lectures more helpful if I had been forced to pay attention.
- It is more entertaining, easier to stay awake, and just as informative.

Theme 4: Those in the experimental group felt the lecture helped them to recognize the problem more multi-dimensionally and globally.

- On the computer it was clearer because it showed several wrong ways also.
- Discussing the different choices made it easier to see why some are better than others.
- This lecture helped me see which areas and clearly understand those that are harder.

7.4 Conclusion

These results support the contention that the multimedia learning tools discussed here can enhance a "traditional" classroom lecture. Although the groups did not significantly differ in quiz scores, due to a ceiling effect, those in the experimental group gave significantly higher ratings in their overall assessment of the lecture. These results were also consistent with the qualitative analysis of students comments, in which students in the control group had a greater number of positive comments about the lecture overall, as compared to those in the control group. These results extend this previous research by indicating that the software developed as part of this project can also significantly enhance a traditional classroom lecture.

A closer examination of the data provide us with some insights into the aspects of the multimedia lecture that lead to the more positive views of the students. It appears that two basic factors lead to the effectiveness of the multimedia-enhanced lecture. First, students found the tools as important aids in visualization and second, the tools supported a more engaging and interactive lecture style. It is not surprising that the dynamic three-dimensional representations of the content allowed students to more readily visualize some of the concepts in Statics, which involve complex multi-dimensional relations. A quick examination of the modules developed as part of this project (*http://web.umr.edu/~bestmech*) provides ample evidence that these modules can provide aspects of visualization that the most proficient instructor could never reproduce via a two-dimensional drawing, or through verbal description. Such tools can presumably be particularly powerful when combined with a professors' commentary within the lecture format.

A subtle, yet more fundamental, characteristic of the multimedia-enhanced lecture was the transformational effect that it had on the instructors' lecture. A number of researchers have pointed out that learning technologies can only be effective if they are used in conjunction with good pedagogical practice, such as increasing student interactivity and engagement [20]. This appears to be exactly what happened in this case. Since the instructor was not required to work through the examples on the board step-by-step, she was free to spend more time discussing different aspects of the problems, which led the students in the multimedia-enhanced class to view more aspects of the problem in a more global fashion. Further, multiple-choice questions were built into the multimedia modules, and the time saved by displaying examples contained within the instructional modules allowed time to cover these questions, which the instructor did, in an interactive discussion format.

8. Study 5: Evaluation of Software in Multiple Instructional Environments (Applied: Dissemination Evaluation)

8.1 Rationale

The purpose of this study was to extend the project evaluation by examining the effect of the software in a number of instructional settings, representing different types of universities and different cultural contexts. Professors from a number of U.S. and international colleges and universities were contacted during the fall of 2003 and offered an opportunity to access four different interactive instructional modules. The modules covered four subjects: Centroids and Moments of Inertia; Stress Transformations; Mohr's Circle Stress Transformations; and Structural Analysis (Trusses and Frames). The professors were told that they could integrate the modules into their classes in any way they chose, and they were requested to encourage students to complete Internet-based surveys over the modules that they used. Students of nineteen different faculty responded to the survey, representing sixteen different schools. Ten of the schools were located in the U.S. and six were located outside of the U.S.

Results from the Centroids and Moments of Inertia (CMI) survey will be discussed here. The conclusions from this survey are representative of the three other studies. For the CMI study, students from four U.S schools (Ohio State University, Columbus, Ohio; Texas Tech University, Lubbock, Texas; Virginia Western Community College, Roanoke, Virginia; MCCC, Monroe County Community College, Monroe, Michigan) and three schools from outside of the U.S. (Tec de Monterrey, Monterrey Mexico; Instituto da Technologia da Amazonia, Brazil; University of Sarajevo, Zenica, Bosnia and Herzegovinia) participated in the survey.

8.2 Experimental Procedures

In addition to a number of general survey statements, the CMI survey included a series of six questions that asked students to rate their knowledge on an important aspect of the topic covered in the module *before* and *after* they completed the module. On each of these questions students were asked to rate their degree of agreement from 1 to 9, with 1 representing "strongly disagree" and 9 representing "strongly agree". The six before/after questions for the CMI module were:

1 & 2 **Before/After** using the centroid and moment of inertia review, I was confident in my ability to determine the centroid location for composite shapes…

3 & 4 **Before/After** using the centroid and moment of inertia review, I was confident that I could correctly determine the moments of inertia (about both the horizontal and vertical centroidal axes) for composite shapes...

5 & 6 **Before/After** using the centroid and moment of inertia review, I was confident that I could correctly determine the moments of inertia (about both the horizontal and vertical centroidal axes) for shapes consisting of standard steel shapes.

A "before" and "after" composite score was created for the survey by averaging the responses to the three before and three after questions, respectively.

8.3 Comparison of Pre vs. Post Knowledge Ratings

8.3.1 Analysis

In order to examine the effect of location (U.S. vs. International) a location variable was created, with students from U.S. schools classified as *U.S.* and students from schools outside the U.S. classified as *International*. A two-way mixed analysis of variance (ANOVA) was computed with Time (before vs. after) serving as a within-subject independent variable, Location (U.S. vs. International) serving as a between-subject independent variable, and ratings serving as the dependent variable.

In the CMI analysis, significant effects were found for Time $F(1,132) = 90.17$, $p < .001$; Location $F(1,132) = 12.36$, $p < .01$; and Time X Location $F(1,132) = 21.52$, $p < .001$. Figure 6 displays the means associated with this ANOVA.

8.3.2 Interpretation

The significant main effects for Time and Location indicate that, overall: 1) students rated their knowledge after using the software as significantly higher than before and 2) the international students rated their knowledge of the topics higher. However, both of these main effects are better explained by the significant interaction. In both cases, the international students rated their knowledge before as being substantially greater than the U.S. students, but this gap was, for the most part, closed in the after-software rating. As a consequence, the large pre-to-post rating gain was mainly the result of the U.S. students' increase, while the increase was not so dramatic for the international students.

This effect may be interpreted in a number of ways. The most obvious, is that the U.S. students benefited more from the software. It is also possible that the high initial ratings for the international students simply left little room for improvement on the post ratings. It's also possible that the U.S. students are simply less confident initially, and become more confident as a result of their experience with the software. In summary, the ratings increased significantly across groups, but the effect was exhibited much more dramatically with the U.S. students.

8.4 Males vs. Females

8.4.1 Analysis

Change in pre to post test rating was also examined as a function of gender in a mixed analysis of variance (ANOVA) for the CMI survey. Time (before vs. after) again served as a within-subject independent variable, Gender (male vs. female) served as a between-subject independent variable, and ratings served as the dependent variable. In reporting the results below we will not discuss the main effect for Time since that effect is redundant with respect to the previous ANOVA.

In this analysis, there was no significant main effect for Gender, but there was a significant Gender X Time interaction, $F(1,132) = 6.26$, $p < .05$. The means associated with this interaction are displayed in Figure 7.

8.4.2 Interpretation

The significant interaction that was found in the CMI analysis indicates that, while both males and females increased in their rating, the increase was greater for the males. They initially rated their knowledge lower, and subsequently rated their knowledge slightly higher, following their experience with the software. This indicates that, to the extent that the positive impact of the software differed as a function of gender, this impact was greater for males. However, it's important to note that ratings increased for both males and females.

8.5 Conclusions

The survey results provide further support for the effectiveness of this instructional multimedia. This support was almost as strong as it could be given the subjective questionnaires that were used. Students rated their knowledge as greater after using the software than before and demonstrated strong agreement with a number of other statements indicating the effectiveness of the software.

There are, however, two important qualifications. The positive impact was manifested more for students from U.S. and for males, as opposed to non-U.S., institutions and females. Before addressing potential explanations for these effects, it's important to note that ratings increased for both nationality and gender groups, the effect was simply more pronounced within the U.S. and male groups.

With respect to the differential effect of nationality, there are a number of reasons why such an effect would occur, most of which center around the fact that the software was created at the University of Missouri – Rolla, a U.S. University. In fact, all of the content and most of the software design was provided by professors who were born and raised in the United States. All of the written information in the software was in English written by these professors, and for many of the students in the international universities, this was most likely not their first language. This may have been particularly relevant in comparing the software and text books, where the text may not have been written in English. Beyond language, there may also be cultural difference in the way that students in U.S. vs. non-U.S. schools view and most effectively learn engineering concepts.

The significant interaction involving gender is more difficulty to explain. One potential explanation is that males and females may differ to some degree in the ways in which they process spatial displays, or more accurately in their preference for such displays (Hall & Hickman, 1999). Another, perhaps more important, explanation is that this effect was somewhat of an anomaly, in that it does not occur consistently in these types of context. In fact, a more detailed analysis of these surveys including all modules was carried out and published elsewhere (Hall, Hubing, Philpot, Flori, & Yellamraju, 2004), and this indicated that gender did not interact with pre vs. post test with the other three modules examined (though the nationality interaction effect was consistent for all modules).

9. Overall Conclusions

In terms of the project, this set of studies provide strong evidence that this instructional software can be effective across a variety of contexts in comparison to different types of criteria, such as pre-knowledge, class text books, and control groups, based on both objective and subject phenomenon. The progression of the research also indicates that initial systematic research on software components lead to more effectively designed software.

In terms of the evaluation model, this project provides an example of how the triangulation of diverse methodological and measurement tools can be combined to provide a rich an integrative picture of a large scale instructional software project, informing and evaluating development.

10. References

[1]Philpot, T.A., R. Flori, R.H. Hall, N. Hubing, D.B. Oglesby, and V. Yellamraju. *Interactive learning Tools: Animating mechanics of materials*. in *American Society for Engineering Education Conference*. 2002. Montreal, Canada.

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

[3]Hubing, N., R. Flori, R.H. Hall, D.B. Oglesby, T.A. Philpot, and V. Yellamraju. *Interactive Learning Tools: Animating Statics*. in *American Society for Engineering Education Conference*. 2002. Montreal, CA.

[4]Hall, R.H., T.A. Philpot, D.B. Oglesby, R. Flori, N. Hubing, S.E. Watkins, and V. Yellamraju. *A model for the evaluation of innovative engineering courseware: Engineering an assessment program*. in *American Society for Engineering Education Conference*. 2002. Montreal, Canada.

[5]Flori, R.E., D.B. Oglesby, T.A. Philpot, N. Hubing, R.H. Hall, and V. Yellamraju. *Incorporating Web-Based Homework Problems in Engineering Dynamics*. in *American Society of Engineering Education Conference*. 2002. Montreal, CA.

[6]Eller, V.E., R.H. Hall, and S.E. Watkins. *Multimedia web-based resources for engineering education: The media design and assessment laboratory at UMR*. in *American Society for Engineering Education Conference*. 2001. Albuquerque, NM.

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

[8]Hall, R.H., L. Wilfred, M.G. Hilgers, M.C. Leu, C.P. Walker, and J.M. Hortenstine, *Virtual Terrorist Attack on the Computer Science Building: Design and Evaluation of a Research Methodology.* Presence-Connect, 2002. **4**(4): p. http://www.presence-connect.com.

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

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

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

[12]Smith, S. *An Interactive Tool for Math and Science Learning.* in *International Conference on Mathematics / Science and Technology*. 1999.

[13]Jih, H., *Promoting Interactive Learning Through Contextual Interfaces on a Web-Based Guided Discovery CAL.* Journal of Computers in Mathematics and Science Teaching, 2001. **20(4)**: p. 367-376.

[14]Westbrook, J. and J. Braithwaite, *The Health Care Game: An Evaluation of a Heuristic, Web-Based Simulation.* Journal of Interactive Learning Research, 2001. **12**: p. 89 - 104.

[15]Amory, A., *Building and Educational Adventure Game: Theory, Design, and Lessons.* Journal of Interactive Learning Research, 2002. **12**: p. 249-263.

[16]Crown, S. *The Development of a Multimedia Instructional CD-ROM/Web Page for Engineering Graphics*. in *World Conference on Educational Mutlimedia, Hypermedia and Telecommunications*. 1999.

[17]Philpot, T.A., N. Hubing, R.H. Hall, R. Flori, D.B. Oglesby, and V. Yellamraju. *Games as Teaching Tools in Engineering Mechanics Courses*. in *American Society of Engineering Education Conference*. 2003.

[18]Merino, D.N. and K.D. Abel, *Evaluating the Effectiveness of Computer Tutorials Versus Traditional Lecturing in Accounting Topics.* Journal of Engineering Education, 2003. **92**: p. 189-194.

[19]Steif, P.S. and L.M. Naples, *Design and Evaluation of Problem Solving Courseware Modules for Mechanics of Materials.* Journal of Engineering Education, 2003. **92**: p. 239-247.

[20]Kadiyala, M. and B. Crynes, *A Review of Literature on Effectiveness of Use of Information Technology in Education.* Journal of Engineering Education, 2000. **89**: p. 177-189.

[21]Babu, S.V., I.I. Suni, and D.H. Rasmussen, *Development of a CD-ROM in Thin Film Technologies: Design, Usability Assessment, and Challenges.* Journal of Engineering Education, 1998. **87**: p. 583-589.

[22]Crown, S.W., *Improving Visualization Skills of Engineering Graphics Students Using Simple JavaScript Web Based Games.* Journal of Engineering Education, 2001. **90**: p. 347-355.

[23]Sorby, S.A. and B.J. Baartmans, *The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students.* Journal of Engineering Education, 2000. **89**: p. 301-307.

RICHARD H. HALL

Richard H. Hall is an Associate Professor of Information Science and Technology at the University of Missouri-Rolla. He received his BS degree in Psychology from the University of North Texas, and Ph.D. degree in Experimental Psychology from Texas Christian University. He is the co-director of UMR's Laboratory for Information Technology Evaluation, and his research focuses on web design, and evaluation with a focus on web based learning technologies.

TIMOTHY A. PHILPOT

Timothy A. Philpot is an Assistant Professor in the Basic Engineering Department and a Research Associate for the Center for Techology Enhanced Learning at the University of Missouri–Rolla. Dr. Philpot received a Ph.D. degree from Purdue University in 1992, an M.Engr. degree from Cornell University in 1980, and a B.S. from the University of Kentucky in 1979, all in Civil Engineering. Dr. Philpot teaches Statics and Mechanics of Materials and is the project director of the U.S. Department of Education grant that supported this work. Dr. Philpot is the author of *MDSolids – Educational Software for Mechanics of Materials*.

NANCY HUBING

Dr. Hubing is an Associate Professor in the Basic Engineering Department at the University of Missouri–Rolla. Prior to joining the BE department in August 2000, she was on the faculty of the Electrical and Computer Engineering Department at UMR from 1989 to 1999, and taught high school physics in 1999-00. She completed her Ph.D. in ECE at N.C. State University in 1989. Dr. Hubing enjoys research involving educational methods and technology in the classroom.

RALPH E. FLORI

Dr. Ralph E. Flori was educated as a petroleum engineer (UM-Rolla Ph.D. '87). As an associate professor in the Basic Engineering Department at the University of Missouri–Rolla, he teaches Dynamics, Statics, Mechanics of Materials, and a freshman engineering design course. He is actively involved in developing educational software for teaching engineering mechanics courses. He has earned fourteen awards for outstanding teaching and faculty excellence.

VIKAS YELLAMRAJU

Vikas Yellamraju is a Software Support Analyst at the University of Missouri–Rolla. He received a M.S. in Industrial Engineering from the University of Oklahoma in 2001 and a B.Tech. in Mechanical Engineering from Nagarjuna University, India in 1995. His present work involves research on multimedia and online learning technology for engineering applications. He is responsible for designing, developing, supervising, and implementing online education tools.