Evaluation of an Interactive Multimedia Intelligent Tutoring System

Susan M. Miller Brian P. Butz
Temple University, Philadelphia, PA. United States
susan.miller@temple.edu bpbutz@temple.edu

***Abstract-* This paper reports usability and effectiveness evaluations of the Interactive Multimedia Intelligent System (IMITS), a system designed to tutor second year electrical engineering undergraduate. Overall, students' responses were favorable. IMITS improved performance on at least one classroom achievement measure. Regression analyses revealed that the more students used IMITS, with usage defined as the number of the number of questions encountered on a particular engineering concept, the better they learned that concept.**

1. INTRODUCTION

This study evaluates the usability and effectiveness of Interactive Multimedia Intelligent System (IMITS). IMITS was designed by Dr. Brian Butz as a supplement to a two-semester introductory course on circuits for electrical engineering undergraduates. IMITS combines interactive multimedia with expert system technology so that IMITS simulates the behavior of an intelligent human tutor. In IMITS, expert understanding is programmed into the software, while the model of each individual student is constructed and modified by an algorithm as the student interacts with the software. During these interactions, the student's model is compared with the expert model. The expert system introduces remedial instruction when there is a significant disparity between the models (Butz, 2001, 2002).

IMITS incorporates authentic, contextually-based problems with a degree of learner flexibility unusual in an ITS environment. A story line was adopted so that the student could play a role of a junior engineer newly hired by the multinational IMITS Corporation and solve problems in a professional, albeit, virtual setting. The student selects a design team to join in the area in which the learner desires tutoring. Once a team member, the student has a virtual office and receives a series of contextually based problems from the virtual team leader via televideo and email communications. The student may complete the assignment using any of the virtual resources available: books, corporation notes and specifications, and a laboratory. Resource books contain highly interactive explanatory materials needed to perform the assignment and the file cabinet contains corporation notes and specifications about various corporation products. The student can access a virtual laboratory to build and test possible designs. The student can interact directly with the virtual team leader and the virtual skills advancement director who guide the student through the various assignments and monitor progress. In addition, the IMITS Corporation provides internal training on its own using case studies that serve as assignments. IMITS was evaluated from two perspectives described in the following sections: Usability of the software and Effectiveness of the software.

2. USABILITY STUDY

2.1. Sample and Procedures

Usability data provide information on a program's functional effectiveness, efficiency, ease of learning, ease of use, motivational influence, as well as quality assurance [3]. Usability data were obtained from (a) students in a single group design (usability data only) and (b) students who were part of the effectiveness study (quasi-experimental design). In all, 114 students completed a usability survey. Researchers communicated with faculty at geographically distant sites through email and telephone. Faculty members were provided written protocols for distributing the software and collecting information. They were able to obtain technical assistance from staff at Temple's Intelligent Systems Application Center and protocol guidance from evaluation staff.

2.2. Instrumentation

The IMITS Usability Questionnaire was used to gather information concerning students' reactions to the software. Not including demographic questions, the questionnaire included six yes/no questions, 22 questions using a 5-point Likert scale (where 1 = strongly disagree, 5 = strongly agree), and four open-ended questions. Also, usability data were obtained from log files generated by the software. As a student used IMITS, the expert system generated a log file that indicated each software feature selected or encountered by the student (e.g., assignments, bookshelf, virtual lab). For each location or feature in the software (e.g., Direct Currents assignment #5 or the virtual book chapter on Kirchhoff's Voltage Law), a student's performance on a set of learning objective was recorded.

2.3. Usability Results

A. Successful learning environment: Student rated IMITS as a useful learning tool, especially regarding acquisition of engineering concepts (mean = 3.58). They viewed the software as an expert source of information (mean = 3.36). Students were asked if they thought that the more they used the software the more they learned; about half agreed. Results of the effectiveness study confirmed their perceptions.

B. Motivational Value: Student enjoyed using the program. IMITS contains many motivational features including attention getting-strategies, confidence-building activities, and perceived learning value. Students gave IMITS high marks as a learning tool, which added to its motivational value. Although students are not as likely to agree they enjoy a learning product as they do a leisure one, we directly asked them this question (mean = 3.11).

C. Ease of Use: No matter how pedagogically well-designed, without a user friendly interface, students will not continue to use any program. Most students found IMITS easy to use (mean = 4.06) and intuitive (mean = 3.68). They gave the same high marks to using the different office components (mean = 3.77) and to using the virtual lab (mean = 3.15). They reported limited navigation within sections (mean = 3.12).

D. Perceived usefulness of software features: Students reported that the various components of the software program were helpful. Not surprisingly, given students' familiarity with textbooks, they rated the virtual bookshelf as the most helpful feature (58%). Eighty percent highly endorsed the clarity of the book materials. Ed, the virtual team leader, received the lowest endorsement.

E. Authenticity of virtual learning environment: Despite the technological sophistication of today's students, they viewed favorably the realism of several features of the virtual environment: student's role as a junior engineer (mean = 3.07), the company office (mean 3.05), virtual lab (mean = 3.04).

F. Quality Assurance: Overall, students found Interactive Multimedia Intelligent Tutoring System (IMITS) to be a high quality multimedia product including audio quality (mean = 3.99) and graphics (mean = 3.83).

2.4. **Usability: Log File Data**

Analysis of user log files is a standard usability data collection method for the evaluation of software [3]. Data were collected from user log files of students in the experimental condition of the Effectiveness Study (see next section).

A. Coverage: Two measures were used as indicators of software coverage: the number of learning objectives encountered and the number of visits (and re-visits) to major IMITS components (e.g., virtual books and assignments). The greatest number of objectives encountered by any student was 15. Thirty-five percent of students encountered 14 learning objectives (median = 11), indicating that students covered a substantial number of learning objectives. Another indication of coverage: Sixty percent of students visited or re-visited 15 or more chapters and assignments.

B. Content: The most frequently encountered sets of questions corresponded to: Ohm's Law (OHL, 541), Kirchhoff's Current Law (KCL, 461) and Kirchholff's (KVL, 421). The sets of learning objective questions least frequently encountered were Complex Algebra and Graph Reading. These findings are consistent with content in a first semester circuits course. Although Ohm's Law was the most frequently encountered set of questions, learners encountered almost twice as many individual questions related to the Power Objective as they did to Ohm's Law. This pattern occurred by virtue of the number of questions dealing with the Power Learning Objective in the chapter on Power. Students in Cohorts 3 and 4 covered less material than did students in Cohort 1 and Cohort 2. This type of variability, both in the amount of coverage and the type of material addressed, is expected as instructors differentially integrate IMITS in their courses.

C. Student Performance: The Tutor is a feature of IMITS that is automatically activated when a student's cumulative performance is under 60% on a given learning objective. The performance of 34% of the students (n = 28) was sufficiently high that not one tutor was evoked during the entire time they used IMITS. The performance of 25% dipped so that a tutor was activated between 3 and 8 times. The tutor for Ohm's Law was most frequently activated (32 times). This learning objective was also the most frequently encountered set of questions, so correspondingly frequent activation of tutors would be expected. KVL and KCL were next most frequently encountered set of questions, and tutors for these learning objectives were the next most frequently encountered (22 and 17). There were 17 activations of the tutor for Thevenin Equivalent (THE), although this learning objective was not frequently encountered. One explanation is that THE is a complex concept for students covering DC material. Given the complexity of this concept for novice learners, it is understandable that cumulative achievement on questions relating to this objective would remain low enough to activate the tutor.

3. EFFECTIVENESS STUDY

3.1. Research Design

A quasi-experimental design was used to examine the impact of IMITS on student learning. The study involved students enrolled in four introductory courses on circuits (at two institutions). Each course had two sections; the intervention was randomly assigned to one section. Students in the experimental group received a curriculum in which the instructor integrated IMITS. Students in the control group received the traditional course curriculum. Three instructors participated (one faculty member taught two different courses). Each corresponding experimental and control group is referred to as a cohort.

3.2. Sample and Procedures

Cohort 1 consisted of 39 students (experimental group, n = 24, control group, n = 15). Cohort 2 consisted of 39 participants (experimental group, n = 21, control group, n = 18). Cohort 3 consisted of 49 participants (experimental group, n = 27, control group, n = 22). Cohort 4 consisted of 44 students (experimental group, n = 18, control group, n = 26). Prior to the start of the study, faculty participants met with researchers at Temple University to review the software and discuss research protocol. On-going collaboration with instructors was facilitated through email, telephone and on-site visits. Written consent was obtained from student participants.

3.3. Instrumentation

Instructors obtained measures of student course performance (e.g., examinations) as part of regular class activities. Because each instructor uniquely implemented IMITS, each cohort had different classroom outcome measures. In addition, data were obtained from log files automatically generated by the software. For purposes of the effectiveness study, the dependent variables derived from the log file were gain scores. A gain score was determined by the cumulative performance (percent correct) on the last encountered set of questions minus the student's performance (percent correct) on the first encounter with a set of questions.

3.4. Research Method and Results

A. <u>Did students who used IMITS learn more than their counterparts in the control condition</u>? Between-subjects analyses using analyses of variance (with grade point average as a covariant for Cohorts 2 and 4) were used to answer this question. The dependent variables were student outcome measures. As these measures differed by instructor, analyses were conducted by cohort.

Cohort 1. Analyses were conducted using ANOVA. Statistically significant findings were found on two classroom measures: Lab assignment #4 and Lab assignment #5. For Lab #4, there was a statistically significant difference, $F(1,37) = 22.45$, $p. = .000$, between students in the software group (mean = 9.17, sd = 2.8) and those in the comparison group (mean = 3.33, sd = 4.8). For Lab #5, which embedded IMITS DC #5 problem, there was a statistically significant difference, $F(1,37) = 16.16$, $p. = .000$, between students in the software group (mean = 94.58, sd = 5.8) and those in the comparison group (mean = 87.80, sd = 4.7).

Cohort 2. Analysis of variance with grade point average as a covariant was used for these analyses. Statistically significant findings were found on two classroom measures: Lab assignment #4 and Examination #3. For Lab #4, there was a statistically significant difference, $F(1,34) = 5.17$, $p. = .029$, between students in the software group (mean = 79.9, sd = 36) and those in the comparison group (mean = 91.8, sd = 23.1). For Examination #3, there was a statistically significant difference, $F(1,34) = 4.90$, $p. = .034$, between students in the software group (mean = 78.79, sd = 19.8) and those in the comparison group (mean = 65.94, sd = 18).

Cohort 3. Analyses were conducted using ANOVA. Statistically significant findings were found for two achievement measures: IMITS DC #3 assignment and IMITS DC #4 assignment. For DC #3, there was a statistically significant difference, $F(1,47) = 21.74$, $p. = .000$, between students in the software group (mean = 3.9, sd = 3.9) and those in the comparison group (mean = 7.96, sd = 2.0). For DC #4, there was a statistically significant difference, $F(1,47) = 17.50$, $p. = .000$, between students in the software group (mean = 3.59, sd = 3.9) and those in the comparison group (mean = 7.74, sd = 2.9).

Cohort 4. Analysis of variance with grade point average as a covariant was used for these analyses. A statistically significant difference, $F(1,43) = 5.79$, $p = .000$, was found on the final examination. Students in the software group outperformed students in the control group (means 75.67, 73.39 respectively).

B. <u>Did performance on the IMITS embedded learning objective questions improve the more the software was used</u>? An ordinary least squares (OLS) regression was used to capture the direct effect of the amount of software usage on achievement of each learning objective. The dependent variables were gain scores. To determine effectiveness of software usage on achievement of learning objectives, user log data from the four cohorts were combined. Seventeen learning objectives were covered by students. Analysis of student performance on Complex Algebra was not conducted as too few students encountered questions on this learning objective.

Sixteen separate OLS analyses were run, one for each learning objective gain score under investigation. In addition to the independent variable of interest - the number of questions encountered on that learning objective - other theoretically related variables were regressed on the gain score. These variables were the student's grade point average, race (recoded as White, non-White), and School. School was included in the model only 11 of the 16 analyses, as there were either no or too few students at both institutions addressing five objectives. For 12 learning objectives, the Number of questions encountered by a student was directly related to the learning gain score. For the following learning objectives: Graph Reading, Algebra, Node Voltage, Mesh Current Analysis, Resistance, Ohm's Law, Sources, Voltage Divider, and Thevenin Equivalent, the Number of questions encountered was the only statistically significant predictor variable. For Basic Concepts, the Number of questions encountered, Race and School predicted learning gain. For Power, the Number of questions encountered and School predicted learning gain. For Equivalent Circuits, the Number of questions encountered and Race predicted learning gain. None of these variables proved statistically significant for KVL, KCL, CD or TC. Details on the regression analyses are available from the authors.

4. SUMMARY

Usability analysis revealed that students viewed IMITS as a valuable educational tool for learning engineering concepts. Students found the software moderately engaging and overwhelmingly reported that IMITS was a user-friendly program. Students ranked the bookshelf as their favorite software feature, although they highly endorsed other components as well. Students across cohorts varied in their endorsement of the software to create an authentic working environment. Students gave IMITS very high marks in terms quality assurance (i.e., quality of graphics and audio, and trouble-free execution). Amount of use varied by cohort. Material covered was consistent with material covered in a first semester course on circuits. Areas in which students had the most difficulty were Ohm's Law, Kirchhoff's Current Law, Kirchhoff's Voltage Law and Thevenin Equivalent. There was support for the effectiveness of IMITS on student achievement. For each cohort, the software group outperformed the control group (at a statistically significant level) on at least one measure. There was strong evidence that IMITS works to improve learning engineering concepts in introductory circuits courses. The more students used the software, the better they performed. All things being equal, that is, taking into account grade point average, race, and school, for 12 learning objectives, the number of questions encountered predicted learning gain on that objective.

ACKNOWLEDGMENT

Partial support for this work was provided by the National Science Foundation's Division of Undergraduate Education through grants DUE #9752417 and DUE #9952291.

REFERENCES

[1] Butz, B. P. (2001). The learning mechanism of the Interactive Multimedia Intelligent Tutoring System (IMITS). *Journal of Engineering Education*, *90*, 543-548.

[2] Butz, B. P. (2002, June). *Interactive Intelligent Tutoring System (IMITS): A progress report*. Paper presented at the Annual Conference and Exposition of the American Society for Engineering Education, Montréal, Quebec Canada.

[3] McIntyre, S., Uitdehaage, S., Geffen D. Nyhof-Young, J. *Evaluation multimedia teaching software: how to design and implement usability studies*. Slice of Life, Philadelphia, PA, June 24. 2003.