

Interactive Multimedia Intelligent Tutoring System

RESEARCH AND EVALUATION OVERVIEW

This report provides preliminary evaluation findings for the project *Science, Engineering and Educator Intelligent Tutor (SEE-IT)* which was awarded to Dr. Brian Butz, Intelligent Systems Application Center, College of Engineering, Temple University, Philadelphia, PA (DUE #9952291). This project consisted of the design and development of a multimedia expert system titled *Interactive Multimedia Intelligent Tutoring System (IMITS)* and a preliminary evaluation of the software program.

IMITS was created as a supplement to a two-semester introductory course on circuits for electrical engineering undergraduates. At present, it consists of three modules: Direct Current, (DC), Alternating Current (AC) and Transient Analysis. *IMITS* combines interactive multimedia with expert system technology. In addition to behaving like a domain expert, *IMITS* is designed to simulate the behavior of an intelligent human tutor. It teaches electrical engineering concepts, detects and analyzes student errors, corrects flaws in the student's logic, and clarifies student misunderstandings.

Goals of project were to:

- I. Design an effective, user-friendly, motivating intelligent tutoring system.
- II. Design an adaptable learning environment tailored to the needs of all students.
- III. Improve student learning of fundamental concepts and problem solving in introductory electrical engineering circuits courses through use of a problem-based intelligent tutoring program.
- IV. Implement integration of an intelligent tutoring system in traditional electrical engineering classes.
- V. Provide information on usability, implementation, and outcomes that would assist in the dissemination and adoption across institutions.

Dr. Susan M. Miller, the director of the Instructional and Learning Technology Program within Temple University's College of Education has served as the evaluator of the *IMITS* software and is the author of this research and evaluation overview. This report addresses the evaluation of *IMITS* in light of these goals. The purpose was to determine whether the program met these goals so that the developers could make informed decisions about maintaining or adjusting the intelligent tutoring system. This report provides evaluation findings through the end of the project award period¹. Although this report provides summative project evaluation findings, these results should be considered preliminary. Definitive conclusions about the effectiveness of the program will require large scale experimental studies (see Institute of Education Sciences, 2003 for recommendations regarding the relationship between the type of experimental design and certainty of results).

¹ Analysis from evaluation conducted during the Fall 2003 was not completed by the time this report was generated

IMITS was evaluated from two perspectives: Usability of the software (Goals I, II, IV) and the Effectiveness of the software (Goals III and IV). Different research designs were used to address these research questions. This Research and Evaluation Report is organized in two major sections titled: Usability and Effectiveness.

Usability

Usability data provides information on a program's functional effectiveness, efficiency, ease of learning, ease of use, motivational influence, as well as quality assurance (McIntyre, Uitdehaage, Geffen, Nyhof-Young, 2003). A variety of usability methods and types of users are appropriate for different stages of a software program's development and evaluation goals.

For this project, usability evaluation involved participation of end users (engineering students and faculty). 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). Usability activities documented in this report were carried out near the end of the production phase (beta level testing²). Data collection methods used in usability evaluation were:

- Survey User Feedback (students: quantitative and qualitative data)
- User log file analysis (students: quantitative data)
- Structured interview and discussions with instructors (faculty: qualitative data)

Effectiveness

Evaluation of the effectiveness of *IMITS* involved participation of end users (engineering students) in a quasi-experimental design. Three separate cohort groups were conducted at two institutions. Each cohort consisted of an experimental group (software group) and a control group (traditional instruction only). Data collection methods included:

- Classroom assignments and examinations (quantitative data)
- User log file analysis (quantitative data)

² This report does not include a description of alpha testing which consisted of one-on-one user feedback sessions conducted by program developers with individuals from the target population -engineering students. These sessions were conducted to modify discrete components of the system during the initial phase of software development.

RESEARCH AND EVALUATION METHODS

USABILITY EVALUATION

The usability evaluation addressed Goal I: Design an effective, user-friendly, motivating intelligent tutoring system; Goal II: Design an adaptable learning environment tailored to the needs of all students; and Goal V: Provide information on usability, implementation, and outcomes that would assist in the dissemination and adoption across institutions.

During Fall and Spring 2002, usability data were collected from students using a single group design. These students who were enrolled in introductory circuits courses at several institutions volunteered to use the software and complete the *IMITS Usability Questionnaire*. During the Fall 2002, usability information was also obtained from students who were involved in the effectiveness study. These students completed the usability questionnaire and provided their user log files for analyses.

Usability Sample

Prior to Spring 2002, faculty or administrators from six institutions (Howard, New Mexico State University, Duke University, Temple, Drexel, and Montgomery County Community College) either contacted the Principle Investigator to volunteer for the usability study or were contacted by the Principle Investigator and agreed to participate. Drexel University withdrew from participation early in the semester. At the remaining institutions, twenty-nine students volunteered to use the software program; 27 students completed the *IMITS Usability Questionnaire*. Participating institutions were: Howard University (n = 15), New Mexico State University (n = 2), Duke University (n = 6), Temple University (n = 4) and Montgomery County Community College (n = 2).

Of those that completed the usability questionnaire, 67% were male and 33% were female. Twelve participants were African American, 9 were Caucasian, 5 indicated Other, and 1 was Asian American. Over half (59%) were between 17 and 20 years old, 22% were between 21 and 25 years old, 7% were 26 to 30 years old, 4% were 31-35 years old and 7% were over 35 years old. Students were also asked questions about their prior knowledge. The percent of students indicating completion of the following courses were: Precalculus, 76%; Calculus I, 79%; Calculus II, 72%; Introduction to Physics I, 72%; Introduction to Physics II, 66%; Differential Equations, 55%. Students rated their algebra skills as above average (mean = 4.19, where 1 = poor and 5 = excellent) and they rated themselves as experienced computer users (mean = 4.73)

During Fall 2002, Montgomery County Community College continued participation in the usability evaluation with 11 engineering students using *IMITS* and completing the *IMITS Usability Questionnaire*. Usability data were also collected from students at Rowan and Temple Universities who were using the software as part of the Effectiveness study (3 experimental groups, n = 75). (The experimental study is described in the next section.) These students

completed the usability questionnaire and, in addition, their user log files were analyzed to answer questions regarding usability. Most students completed demographic information. Combining groups for the purpose of usability analysis, the percentages of male and female participants were 85% and 14%. Sixty-seven percent of the students were Caucasian, 10% were African American, 9% were Asian American, 4% were Hispanic, 1% were Pacific Islander, and 9% indicated Other. Over three quarters (80%) were between 17 and 20 years old, 14% were between 21 and 25 years old, 3% were 26 to 30 years old, and 3% were 31-35 years old. On questions about prior knowledge, the percent of students indicating completion of the following courses were: Precalculus, 99%; Calculus I, 99%; Calculus II, 86%; Introduction to Physics I, 94%; Introduction to Physics II, 75%; Differential Equations, 40%. Students rated their algebra skills as above average (mean = 4.33, where 1 = *poor* and 5 = *excellent*) and they rated themselves as experienced computer users (mean = 4.23).

Procedures

Researchers communicated with faculty at geographically distant sites through email and telephone. Materials were distributed and returned via mail. Faculty were provided written protocol for distributing the software and collecting information. Each participating instructor made *IMITS* available to their students in their engineering circuits course. Faculty were able to obtain technical assistance from staff at Temple's Intelligent Systems Application Center and protocol guidance from evaluation staff.

Instruments

IMITS Usability Questionnaire

The *IMITS Usability Questionnaire* was used to gather information about the software including various software options and features, authenticity, and navigation. This questionnaire also included demographic questions as well as questions designed to elicit information about students' study habits and beliefs about learning. Usability questions asked students their views on aspects of the interface design (graphics and animation), quality of the software, ease of use, preferences for various components, authenticity of problem-solving scenarios, and impact on learning. Not including demographic questions, the initial questionnaire (Spring 2002) 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 that addressed various usability issues. In addition, there were 12 questions that asked students about their study habits and learning beliefs (using a 5 point Likert scale where 1 = *not at all true*, 5 = *very true*). Minor revisions to the questionnaire were made for subsequent use: the content of a study habits question was changed, and all items using a Likert-type scale were anchored with the wording 1 = *strongly disagree*, 5 = *strongly agree*.

User Log Files

Data were obtained from log files generated by the software. As a student used *IMITS*, the expert system generated a log file (in a manner invisible to the user) that sequentially indicated each software feature selected or encountered by the student (i.e., assignments, bookshelf, virtual lab).

As a student used *IMITS*, she/he encountered assignments within various realistic scenarios as well as questions embedded in chapters within the virtual books. These assignments and embedded questions were based on specific learning objectives. 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 in the user log file (i.e., a percent correct for that set of questions on a particular learning objective). This information proved useful for diagnostic purposes. Examining students' performances at each location yielded information on the effectiveness of the different features and helped to diagnosis elements that required further modification.

Faculty Engagement and Interviews

Evaluation staff interacted with classroom instructors throughout the usability and effectiveness studies. Observations and faculty comments were noted. Faculty in the effectiveness study completed an exit interview.

USABILITY RESULTS

Usability results are organized by data collection method. Specific evaluation questions are listed in each subsection.

QUESTIONNAIRE

Responses from the *IMITS Usability Questionnaire* were used to answer the following questions about the expert system:

- A. Did learners think that the software was useful in their learning?
- B. Did learners find the software motivating?
- C. Did the learners find the intelligent tutoring system easy to use?
- D. How did students perceive the usefulness of various software features?
- E. Did students "buy-into" the authenticity of problem-solving scenarios?
- F. What was the perceived quality of the software (quality assurance)?
- G. What additional features would users like?

The *IMITS Usability Questionnaire* was modified slightly between the Spring and Fall data collection periods. Thus, findings from Spring 2002 (N= 27) and Fall 2002 (N = 71) samples are reported separately. Qualitative responses are reported in aggregate.

A. IMITS as a successful learning environment: Did students think that using the software was valuable for knowledge acquisition? Did they perceive the software as a learning tool?

Since the purpose of *Interactive Multimedia Intelligent Tutoring System* is to improve student learning, items that addressed this category are important. Responses give an indication of students' perceived value of *IMITS* as a learning environment. Students rated the *IMITS* as a useful learning tool, especially regarding to acquisition of concepts. They viewed the software as

an “expert” source of information. Students in the Fall group were asked if they thought that the more they used the software the more they learned. About half of the students agreed. Results of the effectiveness study confirmed their perceptions; the amount that the student used the software was a determinant of successful performance on embedded learning objectives.

The software helped me to learn engineering concepts.

	Spring 2002	Fall 2002
Strong Disagree	0%	7%
Disagree	0%	15%
No Opinion	22%	6%
Agree	70%	70%
Strongly Agree	7%	3%
Mean (SD)	3.85 (.53)	3.45 (1.0)

I felt I could trust the software to properly inform me.

	Spring 2002	Fall 2002
Strong Disagree	0%	17%
Disagree	11%	24%
No Opinion	7%	1%
Agree	67%	54%
Strongly Agree	15%	4%
Mean (SD)	3.85 (.82)	3.04 (1.3)

The longer I spent using the software the more information I felt I retained. (This item included only in the revised Questionnaire).

	Fall 2002
Strong Disagree	8%
Disagree	32%
No Opinion	9%
Agree	42%
Strongly Agree	9%
Mean (SD)	3.1 (1.2)

I needed to respond quickly to the assignment questions.

	Spring 2002	Fall 2002
Strong Disagree	0%	3%
Disagree	37%	39%
No Opinion	33%	9%
Agree	30%	45%

Strongly Agree	0%	4%
Mean (SD)	2.93 (.83)	3.08 (1.1)

I spent too much time on each assignment.

	Spring 2002	Fall 2002 N = 71
Strong Disagree	0%	1%
Disagree	26%	34%
No Opinion	33%	7%
Agree	37%	48%
Strongly Agree	4%	10%
Mean (SD)	3.19 (.88)	3.31 (1.1)

While working on the software, I solved assignments somewhere other than in front of the computer screen.

	Spring 2002	Fall 2002
Strong Disagree	11%	6%
Disagree	37%	38%
No Opinion	11%	4%
Agree	41%	48%
Strongly Agree	0%	4%
Mean (SD)	2.81 (1.1)	3.07 (1.1)

Did you use the software outside of class assignments? If so, describe the capacity you used it in.
This was an open ended qualitative question.

Thirty students responded to this question. Most reported that they used the software only as part of course assignments or activities, and some noted that it helped them learn material. A few students indicated that they made an effort to use the material to reinforce or better understand concepts they were learning in class. Here are select student comments.

I read topics in the bookshelf that we learned in class.

I only used it on class assignments.

Yes, at home for homework.

Yes, for completing assignments.

No, only when required to do so.

I used it as a study aid by using the bookcase for reviews and practice problems. Yes, because in areas I was unsure of it helped to explain.

Yes, we were assigned various assignments for homework. These assignments helped to understand the given material more.

I showed it to my dad, and used it to review concepts.

B. Motivational Value: Did students enjoy using the program?

For any learning product, there are several characteristics that contribute to making it motivational for the learner. Some of these characteristics are: attention-getting features, relevancy, confidence-building activities, and perceived learning value. *IMITS* contains these motivational features. Learners are immersed in a simulated office environment, given tasks similar to those faced by electrical engineers, and provided guidance and feedback. Various media are used for presentation and problems are embedded in exciting scenes (e.g., emergency room rescue). Issues of production quality and authenticity addressed in the following questions bear on the perceived motivational value of the program. These characteristics are necessary but not sufficient qualities for a motivational environment. However, as noted above, students gave *IMITS* high marks as a learning tool, which adds 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. In addition, students in the Spring semester were asked if they “lost track of time ” as an indication of engagement.

I enjoyed using the software

	Spring 2002	Fall 2002
Strong Disagree	0%	20%
Disagree	11%	31%
No Opinion	26%	3%
Agree	48%	41%
Strongly Agree	15 %	6%
Mean (SD)	2.81 (1.1)	3.07 (1.1)

I lost track of time while using the software.

	Spring 2002
Strong Disagree	7%
Disagree	56%
No Opinion	15%
Agree	19%
Strongly Agree	4%
Mean (SD)	2.56 (1.0)

C. Ease of Use: Did students find IMITS easy to use? Was the interface intuitive? Did they feel comfortable as they encountered the various IMITS components?

Central to any usability study is the question about the degree to which a program (or any educational tool) is user friendly. No matter how pedagogically well-designed, without a user friendly interface, students will not continue to use any program. Most students found Interactive Multimedia Intelligent Tutoring System (*IMITS*) easy to use. In the Spring, 67% agreed or strongly agreed it was easy to use; 80% did so in the Fall. They gave the same high

marks to using the different office components and to using the virtual lab. The only limitation was that students reported limited navigation within sections (37% and 54%). The wording of the question did not permit students to indicate in what manner they felt limited, but responses to the qualitative question that asked students to report problems provided some indications. Students wanted to be able to exit from a given location and then return to the same location or be able to skip previously seen elements. Students were given an opportunity to report problems with the software. Most responses dealt with interface issues, so students' responses are reported in this section.

The software was easy to use.

	Spring 2002	Fall 2002
Strong Disagree	15%	3%
Disagree	15%	14%
No Opinion	7%	3%
Agree	63%	63%
Strongly Agree	4%	17%
Mean (SD)	3.63 (.84)	3.77 (.99)

I was comfortable using the various office components.

	Spring 2002	Fall 2002
Strong Disagree	0%	3%
Disagree	0%	13%
No Opinion	15%	6%
Agree	85%	66%
Strongly Agree	0%	13%
Mean (SD)	4.0 (.55)	3.73 (.94)

I was comfortable using the virtual lab.

	Spring 2002	Fall 2002
Strong Disagree	4%	10%
Disagree	19%	24%
No Opinion	19%	31%
Agree	52%	30%
Strongly Agree	7%	5%
Mean (SD)	3.41 (1.0)	2.99 (1.1)

How to use the software was intuitive.

	Spring 2002	Fall 2002
Strong Disagree	0%	3%

Disagree	4%	14%
No Opinion	22%	6%
Agree	56%	70%
Strongly Agree	10%	7%
Mean (SD)	3.89 (.75)	3.65 (.91)

I was limited in my ability to navigate within sections.

	Spring 2002	Fall 2002 N = 70
Strong Disagree	0%	6%
Disagree	26%	36%
No Opinion	37%	4%
Agree	30%	43%
Strongly Agree	7%	11%
Mean (SD)	3.19 (.92)	3.19 (1.2)

D. Perceived usefulness of software features: To what degree did students think that the various components of the software program were helpful or well constructed?

Overall, students endorsed various components of the *Interactive Multimedia Intelligent Tutoring System*. Not surprisingly, given students familiarity with reading textbooks they rated the virtual bookshelf as the most helpful (48% and 63%). They highly endorsed the clarity of the book materials (82% and 74%), which is significant given that students have a great deal of experience reading textbooks. Ed, the virtual team leader received the lowest endorsement. Ed fared well in the Spring semester where 59% agreed or strongly agreed that they found interactions leader helpful, compared to 9% who disagreed or strongly disagreed. However, students in the Fall semester were more evenly divided with 54% finding the team leader helpful and 44% not. There were also a small number of negative as well as positive comments about the Ed, the virtual team leader in the open-ended questions.

Students using *IMITS* had the option of entering and using the virtual lab. About half of the students in the Spring sample (56%) agreed or strongly agreed that they used the lab to verify answers, but only 30% in the Fall semester did so. About one-third of the students each semester voiced no comment about the lab.

Which feature in the software was the most helpful?

	Spring 2002	Fall 2002
Video	22%	6%
Bookshelf	48%	63%
Lab	11%	6%
File Cabinet	15%	16%
Team Leader	4%	10%

On the whole, the material found on the bookshelf was clear.

	Spring 2002	Fall 2002
Strong Disagree	0%	4%
Disagree	17%	21%
No Opinion	11%	1%
Agree	67%	60%
Strongly Agree	15%	14%
Mean (SD)	3.89 (.75)	3.58 (1.1)

The interactions with the team leader were helpful.

	Spring 2002	Fall 2002
Strong Disagree	4%	13%
Disagree	5%	31%
No Opinion	22%	3%
Agree	44%	45%
Strongly Agree	15%	9%
Mean (SD)	3.52 (1.0)	3.06 (1.3)

The virtual lab helped me verify results.

	Spring 2002	Fall 2002
Strong Disagree	4%	10%
Disagree	11%	21%
No Opinion	30%	39%
Agree	52%	30%
Strongly Agree	4%	0%
Mean (SD)	3.41 (.88)	2.89 (.94)

What did you use the virtual lab for? This was an open ended qualitative question:

Seventy-two students provided written answers to this question. Of those reporting, 43%, (n = 31) provided positive responses regarding their use of the virtual lab; only five students (7%) reported finding it difficult to use. For those answering this question, 51% noted they did not use the lab. Of those using the lab, most indicated they used the lab to verify to test their answers or to aid understanding of specific problems. The following were representative comments:

*I used the virtual lab to confirm a few questions asked. I also built various circuits just for interest's sake.
Overall, I felt the laboratory was a very instructive and engaging component of the software.*

It helped me apply circuit theory and practical application.
I used it to build a circuit, and I gained experience similar to a real lab.
For the 1st project, I didn't need the virtual lab, but I used it just to build my own circuit and see if I could get some correct values. I built a small circuit taken from my textbook and it gave me correct values.
To enhance my understanding of the problem.
To test a circuit before I entered my answer.
Understanding frequency concepts. Calculating small problems.
Play around with connecting circuits to see if it worked.
Help with node voltage and mesh current problems.
Test answers where possible. Did trial and error solving for one problem so I would know what value I was supposed to be getting.
To double check my theory and information I learned while doing the problems.
I used the virtual lab to confirm my answers.
To help become familiar with the information.
Help on certain topics.

E. Authenticity of virtual learning environment: Did students “buy-into” the virtual world of the IMITS Corporation that was created through the narrative story line, with the student’s role as a junior engineer, the company office, contextual videos, and related problem assignments?

Overall, students in the Spring usability group endorsed authenticity higher than did students in the Fall group.

The interaction with the software was personal and I felt as though someone in the environment was talking directly to me.

	Spring 2002	Fall 2002
Strong Disagree	0%	14%
Disagree	19%	37%
No Opinion	19%	3%
Agree	56%	41%
Strongly Agree	4%	6%
Mean (SD)	3.52 (.89)	2.87 (1.2)

While using the software, I felt as though I was playing the role of an engineer within an organization.

	Spring 2002	Fall 2002
Strong Disagree	0%	13%
Disagree	15%	40%
No Opinion	15%	4%
Agree	67%	41%

Strongly Agree	4%	3%
Mean (SD)	3.59 (.80)	2.82 (1.2)

The office gave the impression of a real office.

	Spring 2002	Fall 2002
Strong Disagree	4%	13%
Disagree	15%	40%
No Opinion	33%	4%
Agree	41%	40%
Strongly Agree	8%	4%
Mean (SD)	3.33 (.96)	2.83 (1.2)

The lab seemed or "felt" like a real lab.

	Spring 2002	Fall 2002
Strong Disagree	7%	11%
Disagree	19%	34%
No Opinion	15%	23%
Agree	48%	31%
Strongly Agree	11%	1%
Mean (SD)	3.37 (1.2)	2.77 (1.0)

F. Quality Assurance: Is the software program of good quality? Does it have high production value?

Overall, students found *Interactive Multimedia Intelligent Tutoring System (IMITS)* to be a high quality multimedia product.

The quality of the audio was good.

	Spring 2002	Fall 2002
Strong Disagree	0%	1%
Disagree	0%	9%
No Opinion	7%	%
Agree	75%	68%
Strongly Agree	19%	20%
Mean (SD)	4.11 (.50)	3.96 (.83)

The graphics/video were of good quality.

	Spring 2002	Fall 2002
--	-------------	-----------

Strong Disagree	0%	6%
Disagree	7%	10%
No Opinion	15%	1%
Agree	52%	70%
Strongly Agree	26%	13%
Mean (SD)	3.96 (.85)	3.75 (1.0)

The software ran trouble free

	Spring 2002	Fall 2002
Strong Disagree	0%	17%
Disagree	33%	23%
No Opinion	7%	1%
Agree	44%	45%
Strongly Agree	15%	14%
Mean (SD)	3.41 (1.1)	3.17 (1.4)

Are there any problems you would like to report? This was an open ended qualitative question.

Slightly over half of the students (58% or $n = 59$) answered this open ended question. Of those, 12 reported no problems. Some of these responses were:

The software ran smoothly without any problems!

No, it was overall a good program.

The software was very helpful. No problems.

Overall it was a good program to use, very helpful when needed.

No, I thought the program was useful and clear.

*I think this was a great software. It actually helped where what we learn in class comes to play (transient
→hospital)*

Of the 47 students who reported problems, 13 reported a that the software was not accepting their correct answer.

I'm not sure, but it seemed like the program was sometimes not accepting answers that were right.

In submitting solutions to the design problems, I became somewhat frustrated by my responses being marked as incorrect when I submitted answers in fractions and with a different number of significant digits than the software's answer.

An equal number ($n = 13$) reported a problem running some feature of the software. There did not appear to be a pattern to these reported problems.

It needs a save button. I lost my work numerous times.

I think it should not take up that much space.

The calculator did not work.

The videos would not play on my computer and it would often freeze up. My pc has better than all of the minimum requirements.

Installation was a pain. I would like for it to prompt on installation which directory you want it installed not just install it right to that last drive (c:1) I like to organize stuff on my PC and usually put programs on my D:1 drive.

About 11% of those reporting problems (n = 5) addressed navigation, especially entry and exit issues.

Should be able to stop partway into the chapter or something else, and still access where you were.

You cannot exit the books freely and you cannot exit easily in a lot of ways. Sometimes it freezes

The commentary at the beginning of each problem was very long. I would have liked to have skipped it.

G. What additional features or learning situations would students like to see in the software?

Students were asked a series of questions about additional features that would enhance the software. For one item, students in the Spring 2002 semester indicated they would like to see incentives added as a feature of the story line (i.e., occurring as part of the simulated environment). When added for the Fall 2002 implementation, this feature was not highly endorsed.

Would sample exams in the software be useful?

	Spring 2002	Fall 2002
Yes	100%	66%
No	0%	34%

Would actual lab exercise be useful?

	Spring 2002	Fall 2002
Yes	96%	58%
No	4%	42%

Would you, as an employee of the IMITS Corporation, like to receive incentives (e.g., promotions, new office, golf trips, etc.)? For the Fall 2002 group, this item read "Did you like to receive....?"

	Spring 2002	Fall 2002
Yes	73%	30%

No	27%	70%
----	-----	-----

Would you like to be able to ask Ed, the team leader, questions and have him respond?

	Spring 2002	Fall 2002
Yes	85%	83%
No	15%	17%

Did you use the user's manual?

	Spring 2002	Fall 2002
Yes	18%	14%
No	82%	86%

Would you have preferred to use the software on your own (not part of a class)?

	Spring 2002	Fall 2002
Yes	78%	55%
No	22%	45%

While working with the software, is there anything you would have liked to do that you were unable to do? (e.g., While working with the software, I would have liked to enter the lab half-way between the chapter to do work.) This was an open ended qualitative question

Not really for the most part the program was designed well. It helps teach the topic and that is the most important thing.

I was unable to quit a question session in the middle of the session.

Return and redo a problem better.

Be able to ask Ed questions and have him respond.

I would have liked to ask my team leader questions. Or they should have had some kind of hint option, where you can keep asking for hints throughout the project.

More examples and problems.

Receive useful information from other engineers.

Summary of Questionnaire Findings:

Analysis of the *IMITS Usability Questionnaire* yielded the following key findings:

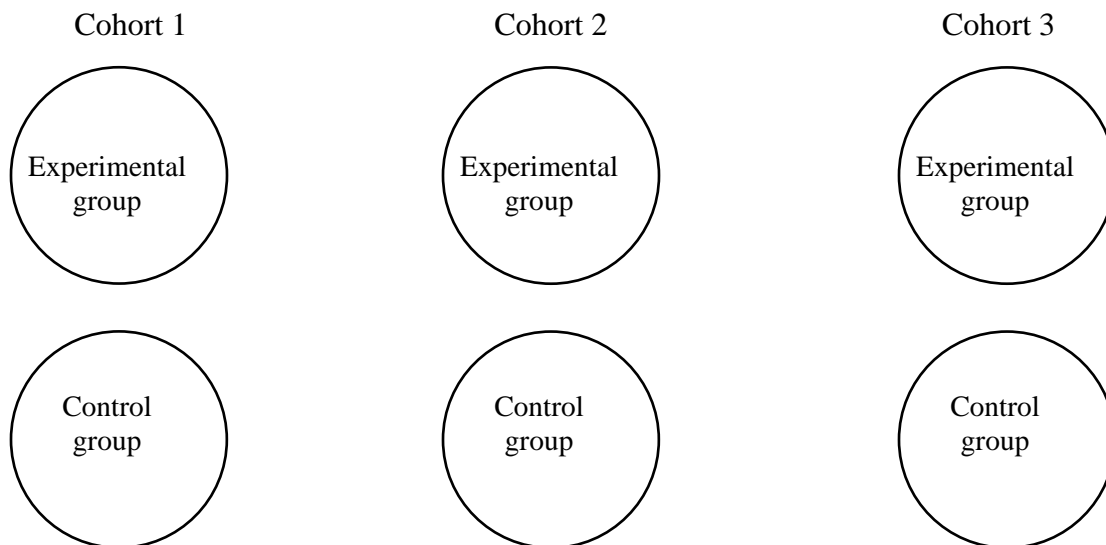
- A. Students saw *IMITS* as a valuable learning tool. Three fourths of the students reported that it helped them learn engineering concepts. Most students used *IMITS* as part of class assignments; some indicated that they used it to review particular content areas.
- B. Students found the software moderately engaging.
- C. Overwhelmingly, students reported that *IMITS* was a user friendly program. They were quite comfortable using its components (e.g., bookshelf, videos, lab).
- D. Student ranked the bookshelf as their favorite software feature, although they highly endorsed other components as well. About half of the students did not explore the virtual lab, which was an optional feature. Of those that did, many reported using it to verify their answers or to understand concepts.
- E. Students in the Spring sample endorsed the capability of the software to mimic or create an authentic working environment higher than did students in the Fall sample.
- F. Students gave *IMITS* very high marks in terms quality assurance (i.e., quality of graphics and audio, and trouble-free execution). When asked to report problems, there were three categories that emerged (a) navigation, especially exit and entry strategies, (b) required precision of the match (e.g., number of decimals/fraction) between entered and pre-programmed answers, and (c) unique execution problems.
- G. Students indicated that they would like the program to include more practice problems or sample examinations, more feedback, and student initiated questions of the “expert” (i.e., team leader).

USER LOG FILES (USABILITY ANALYSIS)

Research Design

A quasi-experimental design was used to examine the impact of *IMITS* on student learning. The study involved students enrolled in three introductory courses on circuits (at two institutions). Each course had two sections; the intervention was randomly assigned to one class. Two instructors participated (one faculty member taught two different courses). For purposes of this report, each course, consisting of an experimental and control group, is referred to as a Cohort.

Students in the experimental group received a curriculum in which the instructor integrated *IMITS*. Students in the control group received the traditional course curriculum.



Sample and Sampling Procedure

An invitation to participate in a research study was mailed to the chair or department head of Electrical or Electrical and Computer Engineering Departments in the Delaware Valley as well as in New Jersey and Maryland. Criteria for participation included (a) the program offered multiple course sections of an electrical circuits course, (b) the department program and the instructor agreed to provide differential curriculum treatments to these sections, and (c) the instructor agreed to attend a training workshop as well as collaborate with researchers for the duration of the study. Two institutions that met these criteria volunteered to participate in the study: Rowan University (New Jersey) and Temple University (Pennsylvania).

At Rowan University, the curriculum was designed so that two separate courses were offered for different student populations: mechanical engineering students (Cohort 1) and electrical engineering students (Cohort 2). Thirty-nine students in Cohort 1 completed the course (experimental group, $n = 24$, control group, $n = 15$). There were 30 male (79%) and 8 female (21%) students in this cohort. Most students (82%) were under 21 years old; the remainder were

under 26 years old. Eighty-two percent Caucasian, 5% were Asian-American, 5% were Hispanic, 3% were African-American, and 3% Pacific Islander, and 3% indicated Other. On questions about prior knowledge, the percent of students indicating completion of specific courses were: Precalculus, 97%; Calculus I, 100%; Calculus II, 95%; Introduction to Physics I, 100%; Introduction to Physics II, 95%; Differential Equations, 89%. Students rated their algebra skills as above average (mean = 4.21, where 1 = *poor* and 5 = *excellent*) and they rated themselves as experienced computer users (mean = 4.38).

Cohort 2 consisted of 39 participants (experimental group, n = 21, control group, n = 18). There were 37 male (95%) and 2 female (5%) students in this cohort. Most (95%) were under 21 years old; 15% were between 21 and 25 years old, and 3% were between 31 and 35 years old. Most were Caucasian (87%), 5% were Asian-American, 3% were African-American, and 5% recorded Other. On questions about prior knowledge, the percent of students indicating completion of the following courses were: Precalculus, 92%; Calculus I, 92%; Calculus II, 87%; Introduction to Physics I, 100%; Introduction to Physics II, 58%; Differential Equations, 16%. Students rated their algebra skills as above average (mean = 4.05, where 1 = *poor* and 5 = *excellent*) and they rated themselves as experienced computer users (mean = 4.16).

There were 49 participants in the third cohort at Temple University) (experimental group, n = 26, control group, n = 23). There were 43 male (88%) and 6 female (12%) students in Cohort 3. Sixty-three percent were between 17 and 20 years old, 28% were between 21 and 25 years old, 6% were between 26 and 30 years old and 2% were between 31 and 35 years old. Twenty-five percent were Caucasian, 25% were African American, 15% were Asian American, 6% were Hispanic, 2% were American Indian, 2% were Pacific Islander, and 25% recorded Other. On questions about prior knowledge, the percent of students indicating completion of the following courses were: Precalculus, 86%; Calculus I, 92%; Calculus II, 69%; Introduction to Physics I, 76%; Introduction to Physics II, 49%; Differential Equations, 20%. Students rated their algebra skills as above average (mean = 4.14, where 1 = *poor* and 5 = *excellent*) and they rated themselves as experienced computer users (mean = 4.26).

Procedures

On-going collaboration was facilitated through email, telephone and on-site visits. The number of contacts varied per site, with an average of one to two weekly contacts. Topics discussed included implementation and maintenance of research protocols, and technical questions regarding software. At the start of the study, the research staff visited each site, presented a description of the study and obtained written consent from those students who agreed to participate in collection of student data. At that time, researchers administered three instruments: (a) a Demographic Form, (b) Learning Styles Inventory, and (c) Group Embedded Figures Test. For the experimental groups, students were provided the software with directions on how to use the software and information on how the software was to be used as part of class activities. The instructors obtained measures of student course performance (i.e., assignments, examinations) as part of regular class activities. At the conclusion of the course, students in the experimental groups were asked to complete the *IMITS Usability Questionnaire* and upload or provide user log files generated by their interaction with the software. All data were coded to ensure confidentiality.

An intelligent tutoring system models an expert's understanding of a subject area and a student's understanding of that area (Anderson, Boyle, & Reiser, 1985). In *IMITS*, the expert's understanding is programmed into the software in advance. The student's model is constructed and modified by an algorithm as the student interacts with the software. Periodically, during the student's interaction with the system, the model of the student's performance is compared with the expert model. As the student interacts with the software, user log files are created. These files are used for modification of the interface of the program with the student. These files also provide useful information regarding each student's use of the program. The user file sequentially records each software location (feature) the student accessed and each interaction that occurred.

Analysis of user log files is a standard usability data collection method for the evaluation of software. Examining log files can provide information on usage patterns and functional effectiveness of a software program. One value of this approach is that quantitative data is automatically collected and stored in a statistical-friendly fashion. This data collection method causes no visible intrusion on the user, and information can readily be obtained from a large number of learners who are using the software under different conditions. Another advantage of this method is that log files reflect a user's actual interaction and performance with the software, thus eliminating bias that is inherent in self-reporting techniques. The disadvantage of this approach is that it does not provide information on users' cognitions, so it is generally used with another method such as questionnaires (as was done in this study).

Analysis of *IMITS* user log files provided information about the degree to which students in the experimental study (N = 75) used *IMITS*. There were 7 missing or unusable log files (2 for Cohort 2 and 5 for Cohort 3).

The following questions guided analysis of *IMITS* user log files:

- A. To what extent did students use *IMITS*?
- B. What content areas did students cover?
- C. How well did students perform?
- D. Was there a difference between high and low ability learners?

Variables included in a user log file (or variables that were derived from information in the log file) that were used for analysis included:

- Number of learning objectives encountered
- Number of books and assignments
- Number of sets of learning objective questions
- Number of questions for each learning objective
- Number and kind (by learning objective) of tutors activated

A. Coverage: To what extent did students use IMITS? Did students in the cohorts differentially interact with IMITS?

One variable not obtainable from *IMITS* user files was the amount of time that students spent on various activities within the software (i.e., entry and exit to a specific location or software event). Only dates and times of entry were recorded. Although there is some advantage to recording date and time of log-on and log-off, or to change of location within the software, time as a measure of usage can be misleading. It is not possible to tell from a time indicator if the student was engaged with the software during the period indicated by the log file.

We used two measures 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).

IMITS consists of three modules for use in two introductory circuits courses. Material was developed based on 36 learning objectives that would typically occur during these courses. These learning objectives are arranged hierarchically; lower objectives are prerequisite to higher objectives. Each software assignment and each question in the software is linked to a single underlying learning objective. It is a student’s response to these learning objective questions that constitutes his/her knowledge model.

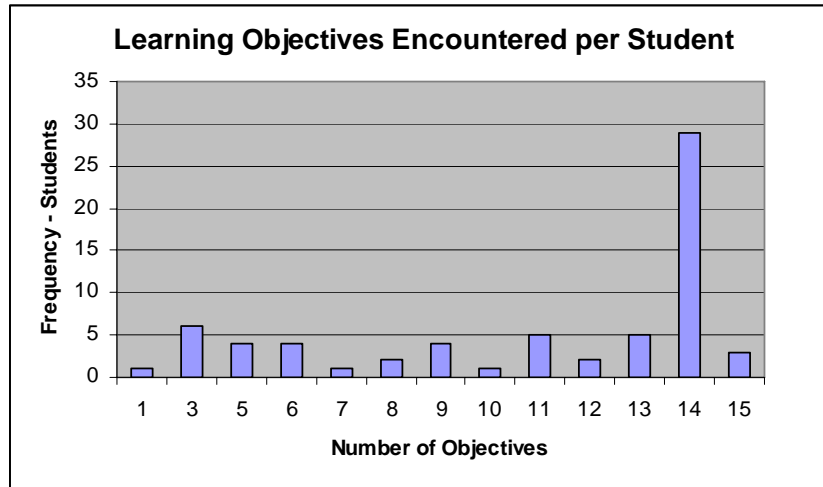
Students in the evaluation study were taking their first semester of introductory circuits, so it was expected that they would not address the entire range of 36 objectives, but would focus on material in the Direct Currents (DC) module. In total, across the three experimental groups, students encountered questions linked to 17 learning objectives. These objectives, their order in the hierarchy, and abbreviations used throughout this report are listed in the following table.

IMITS Learning Objectives Addressed by Students Fall 2002

Objective Hierarchy (out of 36)	Learning Objective	Abbreviation
26	Time Constant	TC
19	Thevenin Equivalent	THE
17	Power	POW
16	Equivalent Circuits	EC
15	Voltage Divider	VD
14	Current Divider	CD
13	Sources	SOU
11	Ohm’s Law	OHL
10	Resistance	RES
9	Mesh Current Analysis	MCA
8	Node Voltage	NV
7	Kirchhoff’s Voltage Law	KVL
6	Kirchhoff’s Current Law	KCL
5	Basic Concepts	BC
4	Complex Algebra	CA
2	Algebra	AL
1	Graph Reading	GR

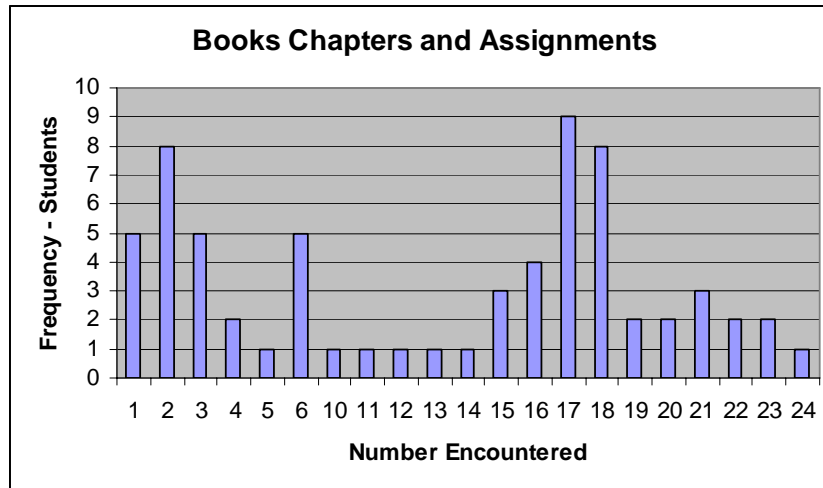
As reflected in Figure 1, the greatest number of objectives encountered by any one student was 15. Forty-three percent of students encountered 14 learning objectives. The median number of learning objectives encountered was 13; indicating students covered a substantial number of DC learning objectives.

F 1. Materials covered as indicated by number of objectives encountered.



Another indication of coverage was the number of different software events that students encountered. Figure 2 reflects the number of book chapters and assignments encountered by students (excluding encounters with tutors, supervisors, and hints). Over half of the students visited or re-visited 15 or more chapters and assignments.

F 2. Materials covered as indicated by book chapters and assignments.



Students in Cohort 3 did not use *IMITS* as extensively as did students in Cohorts 1 and 2. This difference is attributable to the degree to which the course instructor incorporated *IMITS* within the class. The instructor of Cohorts 1 and 2 incorporated all *IMITS* assignments related to the first semester course (i.e., DC #1 to DC #6 assignments) in some aspect of course activities (i.e., as homework, as lab assignment). The instructor’s overt and implied expectation that students use the software is reflected in the degree to which they did so (see table below). Note that variability on scores for Cohort 1 is less than for the other two cohorts.

Mean (and SD) of Learning Objectives and Books/Assignments by Cohort

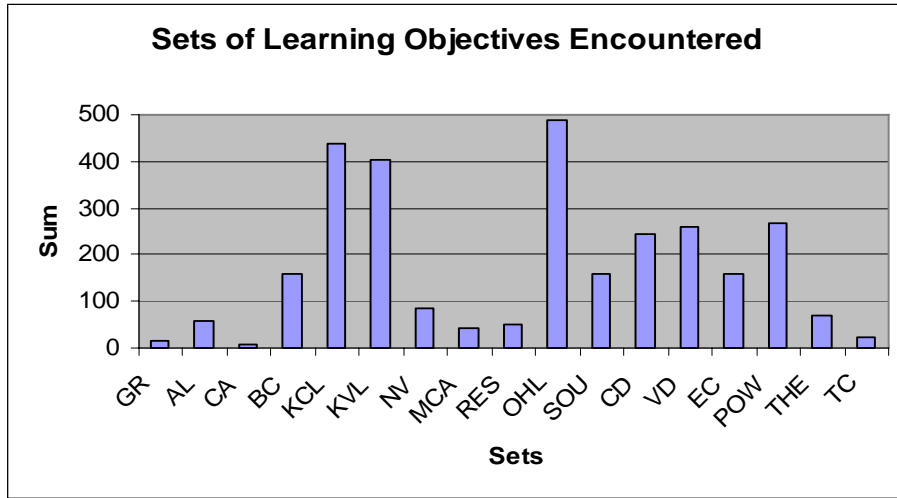
	Objectives	Books & Assignments
Cohort 1	13.25 (1.2)	17.54 (4.3)
Cohort 2	11.71 (4.9)	14.05 (7.1)
Cohort 3	5.81 (3.4)	3.11 (2.5)

B. Content: What content areas (e.g., topics) did students cover? Which topics were most frequently encountered by students? Did students in different cohorts address the same or different material?

As students used *IMITS*, they encountered embedded questions associated with a specific learning objective. At any given event they could encounter one or more questions for that objective. These occurrences are referred to as “sets of learning objective questions.” For example, a student reading the virtual chapter on Power could encounter a set of 2 questions on Ohm’s Law, a set of 3 questions relating to KVL, as well as a set of 4 questions relating to Power.

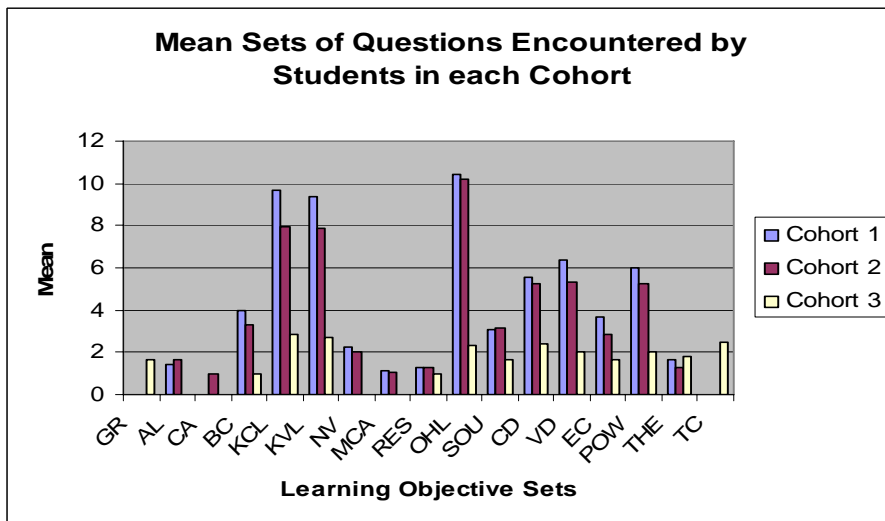
As illustrated in Figure 3, the most frequently encountered sets of questions corresponded to the following learning objectives: Ohm’s Law (OHL, 490), Kirchhoff’s Current Law (KCL, 438) and Kirchhoff’s Voltage Law (KVL, 405). The sets of learning objective questions least frequently encountered were Complex Algebra (CA, 6) and Graph Reading (GR, 15). The latter objective was only encountered by students in Cohort 3. These findings provide a sense of the topics covered by students and are consistent with content in a first semester circuits course.

Figure 3. Content covered as indicated by the total number of sets of questions encountered by students.



As previously noted, students in Cohort 3 covered less material than did students in Cohort 1 and Cohort 2. As observed in Figure 4, which illustrates the average number of sets of questions encountered by students in each cohort, students in Cohort 3 addressed similar but not identical objectives (generally in the same proportion) as did student in the other two cohorts. This type of variability, both in the amount of coverage and the type of material addressed, is expected as instructors differentiate integrate *IMITS* in their courses.

Figure 4. Content covered by each cohort as indicated by mean sets of questions.



Calculating the exact number of questions encountered for each learning objective (instead of encounters with sets of questions) produces a slightly different picture of the content covered by students using *IMITS*. This pattern is represented in Figure 5. (This definition of content coverage was used in the Effectiveness study.)

Although Ohm’s Law was the most frequently encountered set of questions, as seen previously in the Figure 3, 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.

Figure 5. Content covered as indicated by encounters with questions

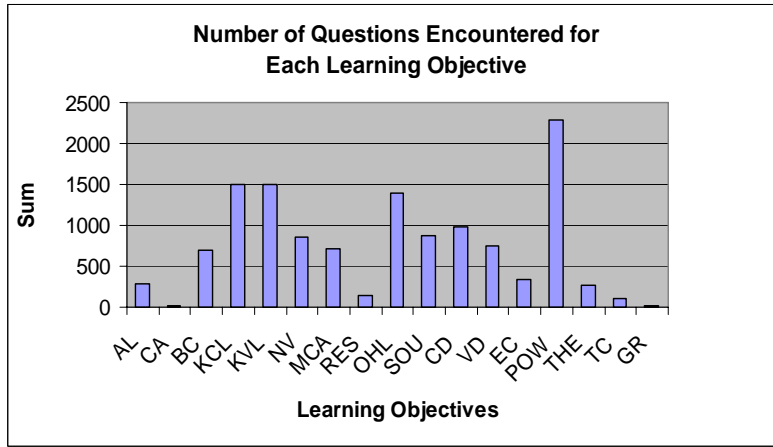
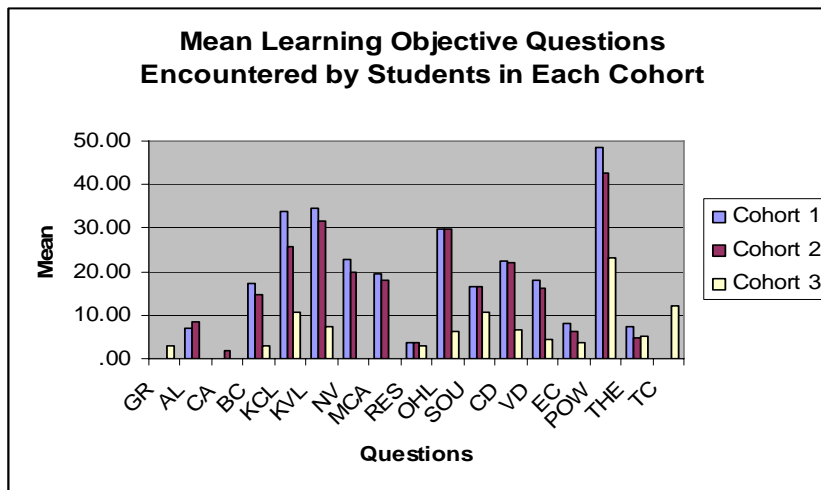


Figure 6 illustrates the mean questions for each learning objective encountered by students in each cohort. Students in Cohort 3 addressed similar but not identical objectives in the same proportion as did student in the other two cohorts.

Figure 6. Content covered by students each cohort as indicated by average number of questions encountered by students in each cohort.



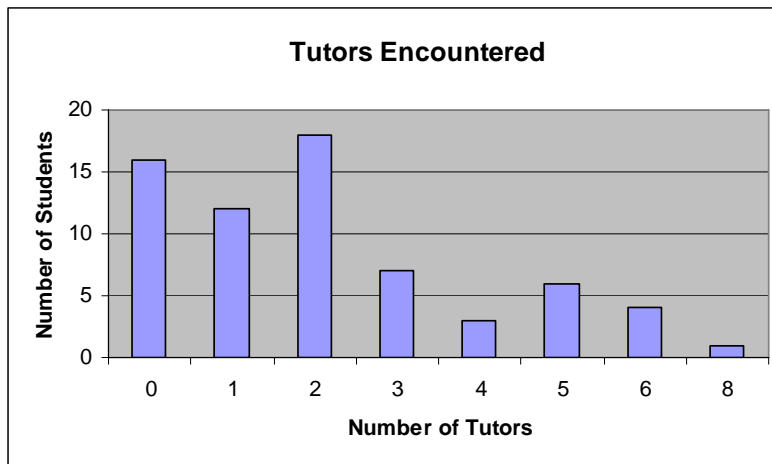
C. Student Performance: To what degree did students require guidance on solving questions? Which learning objectives were most difficult for students?

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 tutor guides the student through a problem-solving approach with the goal of improving understanding of the concept in question (i.e., Ohm’s Law).

The expert system tracks a student’s performance on various assignments and activities and adapts to the learner based not on correct answers to assignments but on cumulative performance on questions targeted to each learning objective. This eliminates the problem of basing a program on a student’s guess about a right answer. Instead, the program adapts to a learner’s understanding of specific learning objectives (i.e., Ohm’s Law).

The performance of 24% of the students (n=16) was sufficiently high that not one tutor was evoked during the entire time they used *IMITS*. Eighteen percent (n = 12) evoked a tutor once during the use of *IMITS*; the performance of 27% (n = 18) activated the tutor twice. The performance of 31% (n = 21) dipped so that a tutor was activated between 3 and 8 times (see Figure 7).

Figure 7. Number of tutors encountered



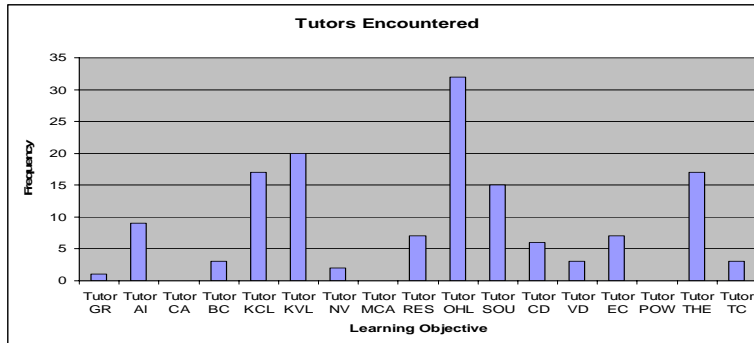
The number of tutors encountered by students in each cohort varied. Interpretation of the average number of tutors encountered needs to be made in the context of the amount of material covered by students in each cohort. Students in Cohort 1 covered more material, while students in Cohort 3 covered substantially less material.

Cohort	Number of Students in Experimental Group	Tutors Mean (SD)
1	24	3.17 (1.9)
2	21	1.63 (1.6)
3	30	1.50 (1.8)

Overall, the tutor for Ohm’s Law was most frequently activated. It was encountered 32 times (including for the same student). This learning objective was also the most frequently encountered set of questions (see Figure 3), 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 (20 and 17).

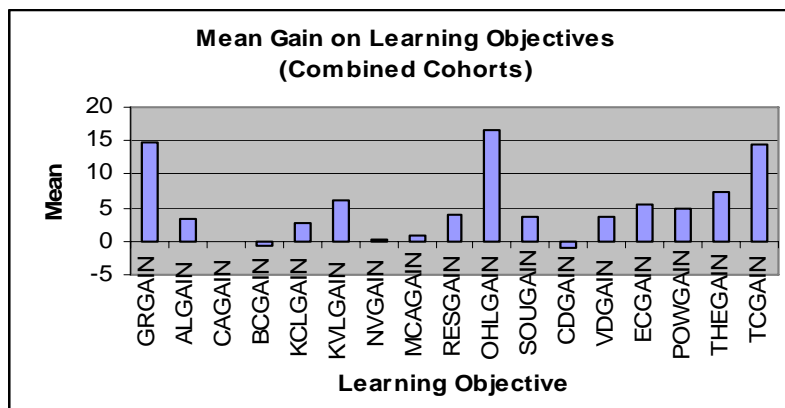
There were 17 activations of the tutor for Thevenin Equivalent (THE), although this learning objective was not frequently encountered. One explanation is that THE is a complex concept for students covering DC material (see Learning Objective Hierarchy Table). Given the complexity of this concept for novice learners, it is understandable that cumulative achievement on questions relating to this objective would remain low enough to activate the tutor. An anomaly, however, was the absence of activation of the tutor for the Power learning objective. This was the learning objective for which students encountered the most questions (see Figure 5) and it too is a relatively complex concept for novice learners.

Figure 8. Student performance on learning objectives as reflected in tutors activated.



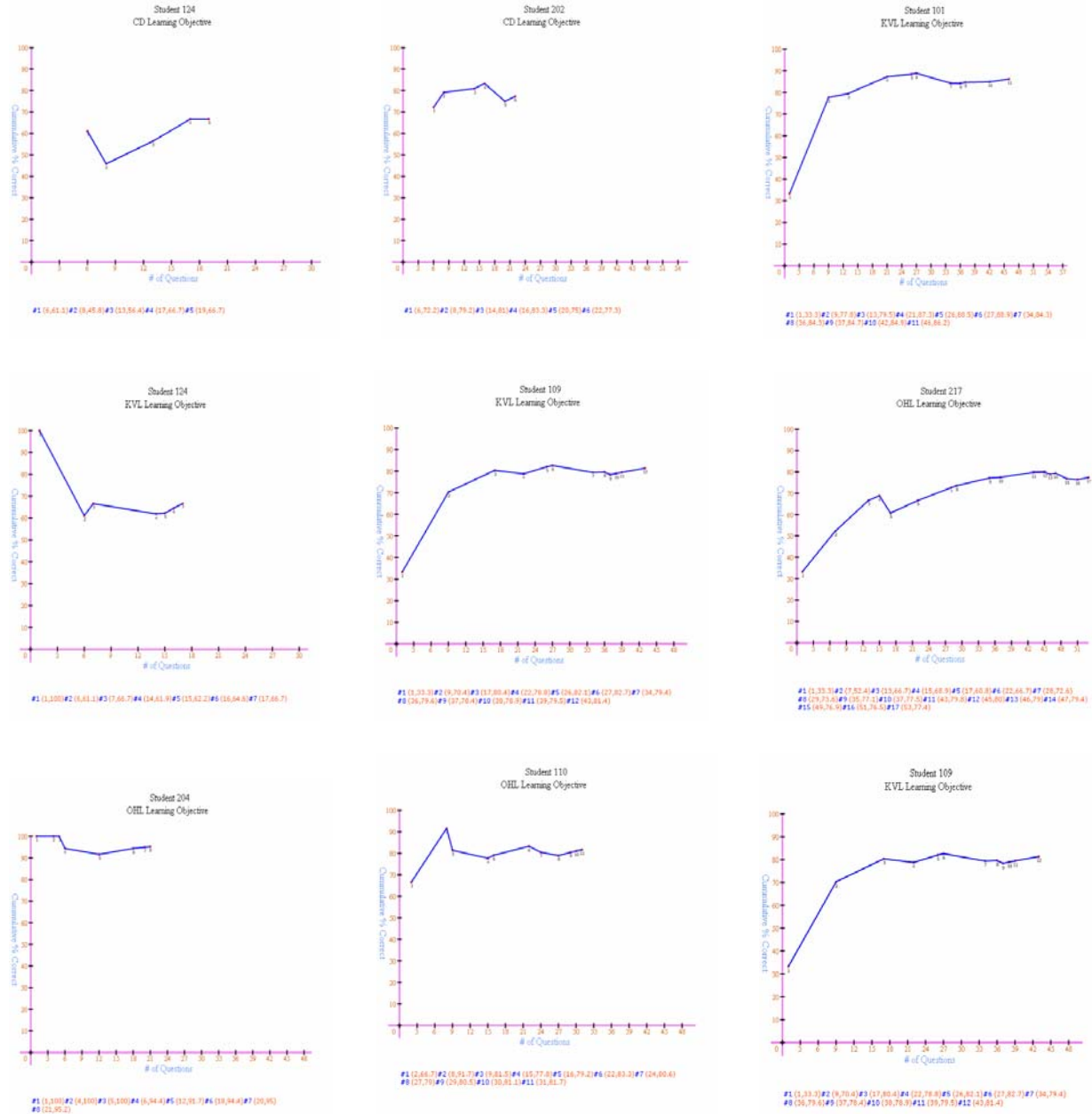
For each student, the cumulative learning gain was calculated for each learning objective. A strong indicator of the effectiveness of IMITS as a learning tool was that on all but 3 learning objectives (e.g., CA, BC, and CD) student learning improved. The average percent gain for each learning objective (combined cohorts) is represented in Figure 9.

Figure 9. Student learning as represented by mean gain percents on each learning objective.



Learning gains for each student for each learning objective were graphed separately. Space in this report limits replication of each graph. Examples for several learning objectives are presented to illustrate the variability in the learning curve for each student – and across students. In most cases, the learning curve progresses “upward” with “ups and down” (i.e., gains and losses) along the way. This is expected during acquisition of content knowledge. Note that the last two digits in the Title indicates the students rank within his or her cohort (i.e., Student 101 is ranked first in Cohort 1).

Figure 10. Illustrations of learning gains for select objectives (by student).

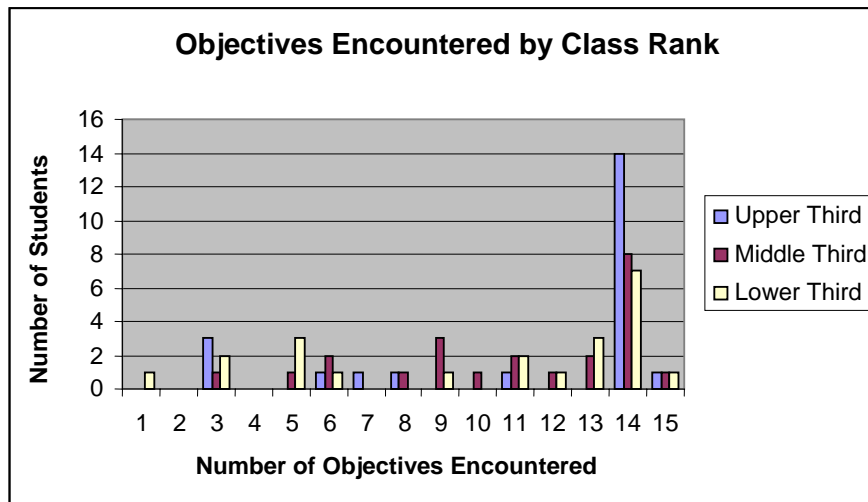


D. Adaptability and Usages based on Diverse Learners: Did students of differing ability level use *IMITS* in the same way? Were there differences among students in coverage, content, and performance patterns?

A goal of engineering education is the development of strategies and learning environments that will be used by and beneficial to students with a range of abilities. For the purpose of these analyses, ability is defined as a student’s class rank based on the final number of points earned in the course (as determined by the instructor). (Note: grade point average was significantly correlated with final points/class rank). Analyses of usable log files were made by combing the upper, middle, and lower thirds for each cohort (the n for each was 22, 23, and 22 respectively). Dramatic differences in coverage, content, or performance would indicate that upper and lower ability students interacted differently with *IMITS*.

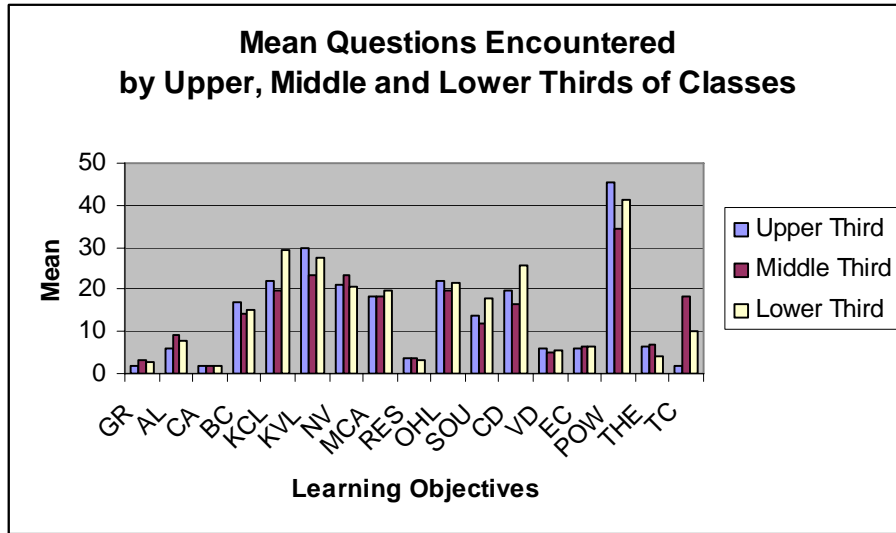
One indicator of coverage (amount of usage) was the number of objectives covered by students. The greatest number of objectives encountered by any one student was 15, with modal at 14, and the median at 13 (see Figure 1). Comparing Figure 11 to Figure 1, there is a similarity in the number of objectives encountered. The mean and median for the upper third was 11.45 and 14; for the middle third, 10.96 and 12; and for the lower third, 10.14 and 12.50. Students in the lower third encountered slightly fewer learning objectives (sum of 223 objectives compared to 252 in the middle and upper thirds). Although the modal number of objectives in each ranking group was 14, more students in the upper third of their class (n = 14) encountered 14 objectives. Only 8 students in the middle third and 7 students in the lower third engaged the same number of objectives. Correlational analysis revealed positive relationship ($r = .49, p = .000$) between the number of objectives addressed and the final number of points earned in class (a more precise indicator of class rank).

Figure 11. Number of Objectives encountered by students in the upper, middle, and lower thirds of their class (based on class ranking)



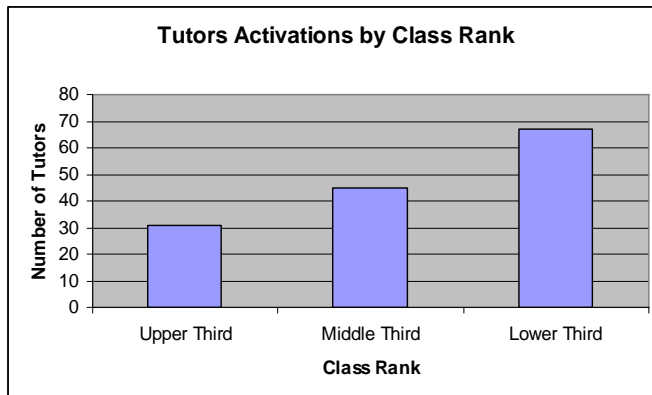
Irrespective of their class standing, students covered approximately the same material in *IMITS*. As noted previously, using the number of questions encountered for each learning objective as a measure of type of content addressed, POW, KVL, KCL and OHL were concepts (learning objectives) most frequently encountered. There were slight variations in learning objectives among the upper, middle, and lower thirds of each class, but, overall, this pattern held true when analyzed by class standing. Figure 12 illustrates this point.

Figure 12. Mean questions (on each learning objective) encountered by students in the upper, middle, and lower thirds of their class (based on class ranking)



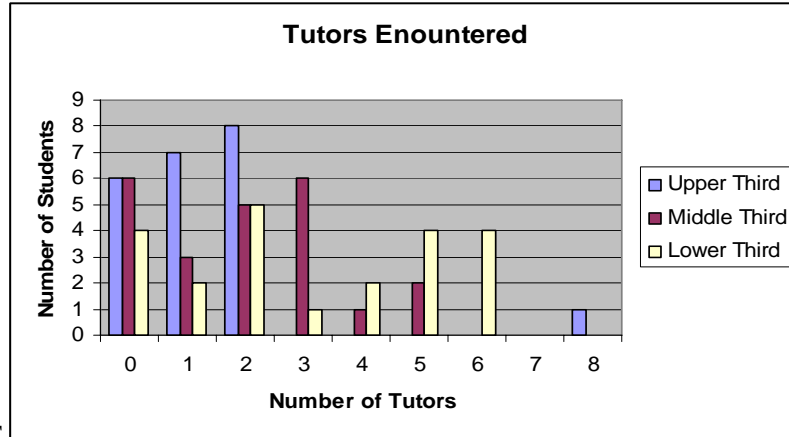
It would be expected that lower performing students would require additional guidance in the acquisition of concepts and in problem solving. This proved to be the case. Students in the upper third of their class encountered a total of 31 tutors, those in the middle third encountered 45 tutors, and those in the lower third activated 67 tutors. Although students in the lower third of their class covered slightly less material, when they did they activated twice as many tutors as those in the upper third. This usage pattern is consistent with low versus high performing students, and indicates that *IMITS* is meeting its design goal.

Figure 13. Allocation of number of tutors activated by class rank.



The following is a more detailed representation of tutors encountered by students in the upper, middle, and lower thirds of their class. For example, 6 students in the upper and middle thirds and 4 students in the lower third did not activate any tutors.

Figure 14. Number of tutors encountered by class rank.



Summary of User Log Files (for Usability)

Analysis of the *IMITS User Log Files* yielded the following key findings:

- A. Use of *IMITS* varied by cohort, with students in Cohorts 1 and 2 using *IMITS* substantially more than students in Cohort 3. Overall, as measured by the number of learning objectives (engineering concepts) and the number of software features (e.g., virtual books and assignments) encountered, students made sufficient use of *IMITS*.
- B. The material covered by students using *IMITS* was consistent with material covered in a first semester course on circuits. The most frequently encountered learning objectives (questions) were on Ohm’s Law, Power, Kirchhoff’s Current Law and Kirchhoff’s Voltage Law.
- C. As expected, student performance, as defined by the number of tutors activated, differed. The cumulative achievement on learning objectives of almost a quarter of the students was high enough not to receive a tutor; and about 30% received 3 to 8 tutors. The ratio of learning objectives to tutors differed by cohort. Areas in which students had the most difficulty were Ohm’s Law, Kirchhoff’s Current Law, Kirchhoff’s Voltage Law and Thevenin Equivalent.
- D. When grouped by performance based on class rank, students in the lower third covered slightly less material. That students in the lower third did not differ dramatically in coverage is heartening, as it is common for lower performing students to demonstrate less engagement with learning materials. Students in the lower third of the class covered the same content material as students in the middle and upper thirds. As expected, students in the lower third encounter almost twice as many tutors as students in the upper third, confirming the *IMITS* performance design.

FACULTY ENGAGEMENT AND INTERVIEWS

One of the most notable aspects of implementing *IMITS*, in both Usability and Effectiveness studies, was (a) variation in implementation and (b) instructor comfort and motivation.

The initial implementation of *IMITS* revealed several challenges:

- Variation in the use of *IMITS* was related to instructor's level of comfort with *IMITS*' pedagogy. *IMITS* uses authentic, project-based scenarios in which students are challenged to engage in higher-level problem solving. Response to the "guided constructivism" of *IMITS* varied. Some instructors saw the possibility of integrating *IMITS* as a way to support desired or planned changes to their curriculum. Other instructors were uncomfortable and found the pedagogical approach at odds with their direct-teaching philosophy.
- Instructors who were comfortable with the pedagogy of *IMITS* and who wanted to use the program required guidance and support, sought to integrate it into their teaching. As with any modification in teaching material, re-adjusting an established curriculum takes not only desire but time and effort.
- In some cases, the instructor's comfort level was related to the degree of their technological skills. Although installing (and using) *IMITS* does not require much skill, some instructors felt more comfortable with a "walk-through" by technical staff.
- An accepted educational maxim is that students value that what is valued by the instructor. This certainly seemed to hold true for *IMITS*. The instructor's value of *IMITS*, as operationalized by integration into the course, seemed to affect students' use of *IMITS*.

IMITS was initially developed as an intelligent tutor to supplement traditional classroom learning. However, several instructors saw the potential for *IMITS* to play a more integral role in the classroom. They provided several suggestions for the future. One suggestion was to modify *IMITS* so that it could be used in the classroom as part of collaborative assignments for teams of students. In fact, one instructor did use *IMITS* in this fashion during a few lab activities. By adding more components (i.e., project planning), one instructor thought that it could be used to aid development of senior design projects. Another instructor suggested that *IMITS* be modified to link the software to in-class learning by including feedback to the instructor on each student's performance. This would provide diagnostic information on each student as well as formative evaluation on class performance, thus permitting the instructor to better tailor classroom teaching. This instructor also added that if *IMITS* was network-based, the flow of information, and its use on campus (in and of class) could be "seamless." Another idea was to provide a communication element to connect students to the instructor or real-world professional.

EFFECTIVENESS STUDY

The research study addressed Goal II: Design an adaptable learning environment tailored to the needs of all students; Goal III: Improve student learning of fundamental concepts and problem solving in introductory electrical engineering circuits courses through use of a problem-based intelligent tutoring program; and Goal IV: Implement integration of an intelligent tutoring system in traditional electrical engineering classes.

Research Design

Same as described on page 18 of this report.

Sample and Sampling Procedure

Same as described on page 18 of this report.

Procedures

Same as described on page 19 of this report.

Treatment Implementation

This study was designed so that integration of the software was under the control of the instructor in collaboration with researchers. The manner in which the software was used was unique to each site. This variability of implementation was deliberate; the goal was to reflect the manner in which the software might be used by an instructor in any classroom. Upon final distribution of the software, it is expected that instructors will use the software based on their unique epistemological and pedagogical views as well as on their program's curriculum. Instructors in this study varied in how they used the software. One instructor incorporated the software (in the experimental group) as part of weekly lab activities. The other instructor provided the software to students to use as a tutor and only minimally integrated the software in regular teaching activities. Approaching this effectiveness study as a field rather than a controlled experiment yielded a design with high external validity and increased threats to internal validity.

Faculty Training

Prior to the start of Fall 2002 classes, instructors from each site (plus the instructor at the Fall 2002 usability site) participated in a one-day workshop led by researchers at Temple University. Instructors were provided (a) an overview of the software design, (b) a description of the research design and protocols, (c) sample syllabi and lesson plans, and (d) a demonstration of these sample instructional strategies. Instructors were provided an opportunity to use the software and to discuss ways in which they might integrate the expert system within their course. One of the benefits of the workshop was the opportunity instructors had to discuss their teaching

approaches and to brainstorm with each other (and with the researchers) uses of the software that were consistent with their teaching strategies and course objectives. This discussion yielded information that will be of benefit to the expert systems designers regarding future modifications to the software.

Instrumentation

Classroom Achievement Measures

For students in the experimental study, outcome data were obtained on select assignments found in the software (e.g., DC assignments #1 through #6). Instructors were asked to assign at least two *IMITS* problems. Students in Cohorts 1 and 2 completed 6 *IMITS* problems (each as part of a larger assignment, resulted in imprecise capture of the desired dependent variable). Students in Cohort 3 completed 2 *IMITS* problems. Students in the experimental group completed these assignments as part of their software usage. Students in the control group were provided written versions of the problems. At each site, the instructor developed usual achievement measures (i.e., homework, quizzes, and examinations).

In order to manage analyses, only major classroom achievement measures (e.g., examinations) and relevant achievement measures (e.g., *IMITS* assignments) were used as dependent variables. For Cohorts 1 and 2, the dependent variables were four lab assignments (which incorporated three *IMITS* problems) and three examinations. For Cohort 3, the dependent variables were three *IMITS* problems and two examinations. Analyses were conducted separately for each cohort.

User Log Files

Data were obtained from log files automatically generated by the software. A description of the generation of user log files was previously provided. For purposes of the effectiveness study, the dependent variables derived from the log file were gain scores (for each learning objective). A gain score for a given learning objective 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. For example, a given student's gain score on the learning objective for Ohm's Law might be: Last encounter (100% correct) - First encounter (50% correct) = Gain score (50%).

EFFECTIVENESS RESULTS

Two major questions guided the evaluation of the software's effectiveness on learning:

- A. Did students who used *IMITS* (experimental condition) learn more (e.g., score higher on achievement measures) than their counterparts in the control condition?
- B. Did performance on the *IMITS* embedded learning objective questions improve the more the software was used?

A between-subjects analysis using analysis of variance (with grade point average as a covariant for Cohort 2) was used to answer the first question. The dependent variables were examinations as well as performance on lab and *IMITS* problems. Because classroom achievement measures were instructor-made, analyses were conducted separately for each cohort.

To answer the second question, did students' learning of important engineering concepts (learning objectives) improve the more they used the software, 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 on each learning objectives.

CLASSROOM ACHIEVEMENT: BETWEEN GROUPS ANALYSES

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, $sd = 2.9$).

LEARNING OBJECTIVES GAIN SCORES: REGRESSION ANALYSES

To determine effectiveness of software usage on achievement of learning objectives, data from user log files from the three cohorts were combined. Seventeen learning objectives were covered by students. Analyses of student performance on Graph Reading, Complex Algebra and Time Constant were not conducted as too few students encountered questions on these learning objectives. Regression analyses were conducted on the remaining objectives.

Fourteen 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 (Temple or not). School was included in the model only for 9 of the 14 analyses; as either were no or too few Temple students covered that the remaining five objectives.

For ten learning objectives, the number of questions encountered by a student is directly related to the learning gain score. These objectives were Algebra, Basic Concepts, Node Voltage, Mesh Current Analysis, Resistance, Ohm’s Law, Sources, Voltage Divider, Equivalent Circuits, and Thevenin Equivalent. For Power, enrollment at a particular school predicted learning gain. None of these variables proved significant for KVL, KCL or CD.

The regression analyses are displayed in the table below. The unstandardized coefficient (regression coefficient) is the statistic in the middle column. This coefficient quantifies the effect of the independent variable on the dependent variables. Using as an example the gain score for algebra as the dependent variable, the results reveal that, *ceteris paribus*, for every 1 question encountered on the algebra objective, the gain score increased 2.43 percentiles.

R square notes the percent of variation in the dependent variable, in this case, gain score that is explained by the independent variables taken together. In the Algebra example, 58.9% of variation in the gain score is explained by the variables in the model taken together.

In educational research, effect size is being used as an indicator of the practical significance of findings. In regression analysis, the proportion of variance accounted for, R square, is a measure of effect size of all entered variables on the dependent variable (Cohen, 1988).

Variables	<u>B</u>	Significance
Gain Score on Algebra		
Number of AL Questions *	2.43	.000
Grade Point Average	1.10	ns
Race	1.49	ns
R ² = .589		
Adj. R ² = .549		
Gain Score on Basic Concepts		
Number of BC Questions *	.395	.050
Grade Point Average	.812	ns
Race	-.699	ns
R ² = .120		
Adj. R ² = .051		
Gain Score On Node Voltage		
Number of NV Questions *	3.53	.000
Grade Point Average	-.558	ns

Race	2.90	ns
$R^2 = .442$		
Adj. $R^2 = .410$		
Gain Score on Mesh Current Analysis		
Number of MCA Questions *	.446	.000
Grade Point Average	6.7E-02	ns
Race	-1.2E-02	ns
$R^2 = .643$		
Adj. $R^2 = .608$		
Gain Score on Resistance		
Number of RES Questions *	5.80	.000
Grade Point Average	-3.7E-02	ns
Race	1.83	ns
$R^2 = .849$		
Adj. $R^2 = .836$		
Gain Score on Ohm's Law		
Number of OHL Questions *	1.7	.001
Grade Point Average	6.80	ns
Race	-9.94	ns
School	27.0	ns
$R^2 =$		
Adj. $R^2 =$		
Gain Score on Voltage Divider		
Number of VD Questions *	1.47	.001
Grade Point Average	-2.42	ns
Race	5.61	ns
School	11.35	ns
$R^2 = .306$		
Adj. $R^2 = .235$		
Gain Score on Equivalent Circuits		
Number of EC Questions *	2.47	.000
Grade Point Average	4.15	ns
Race	6.95	ns
School	2.12	ns
$R^2 = .422$		
Adj. $R^2 = .371$		
Gain Score on Power		
Number of POW Questions	.282	.054
Grade Point Average	-2.02	ns
Race	1.53	ns

School *	25.16	.004
R ² = .268		
Adj. R ² = .206		
Gain Score on Sources		
Number of SOU Questions *	.373	.009
Grade Point Average	-1.87	ns
Race	-8.74E-02	ns
School	3.15	ns
R ² = .183		
Adj. R ² = .119		
Gain Score on Thevenin Equivalent *		
Number of THE Questions	1.15	.004
Grade Point Average	.632	ns
Race	4.45	ns
School	1.86	ns
R ² = 2.12		
Adj. R ² = .124		

* denotes statistical significance

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 (see Figure 5), the better they learned that concept.

The following Figures illustrate this relationship. For illustrative purposes, graphs are presented for Algebra, Ohm's Law, Power, Thevenin Equivalent learning objectives. For viewing ease, graphs are depicted by cohort.

Figure 15. Algebra learning objective: Relationship between number of questions encountered and cumulative gain scores.

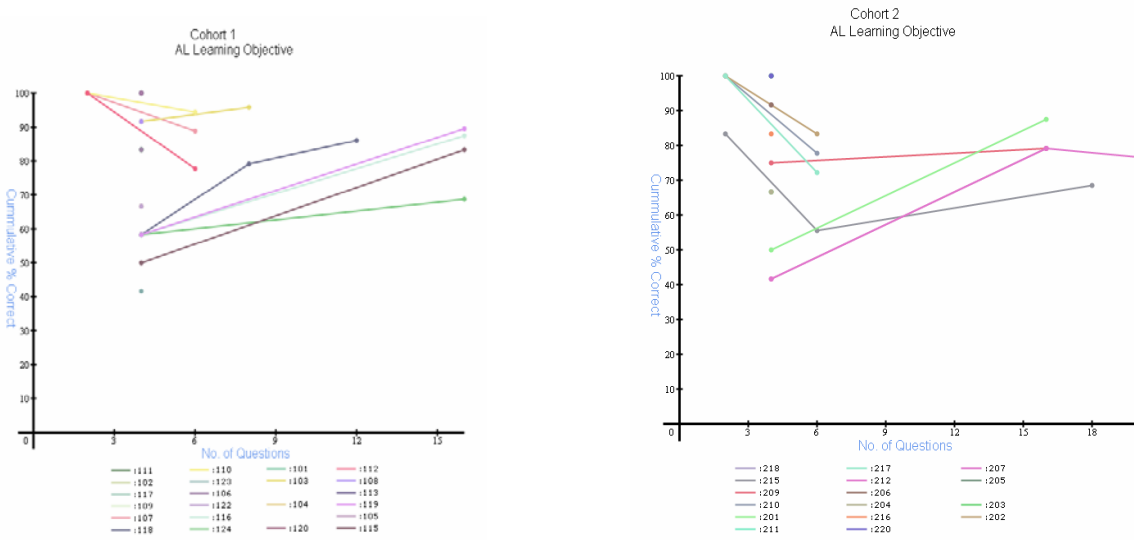


Figure 16. Ohm's Law learning objective: Relationship between number of questions encountered and cumulative gain scores.

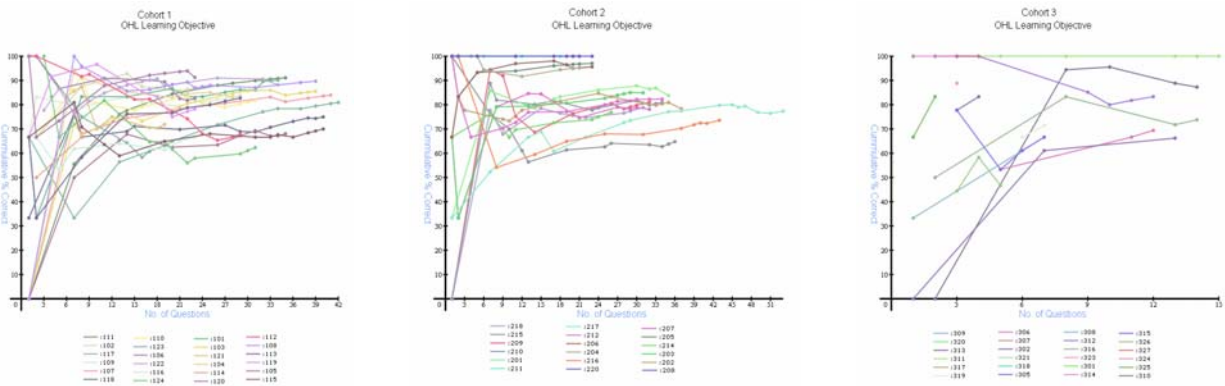


Figure 17. Power learning objective: Relationship between number of questions encountered and cumulative gain scores.

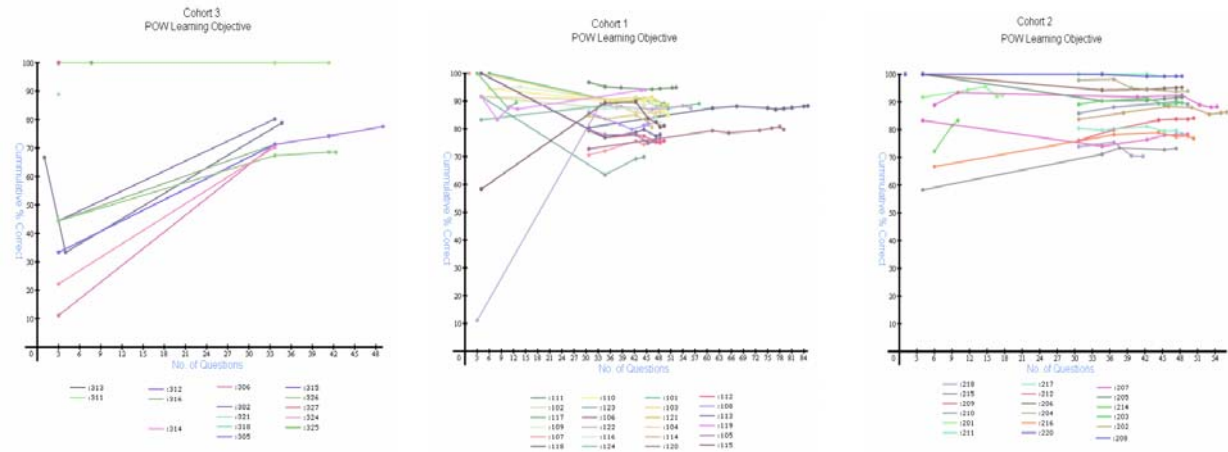
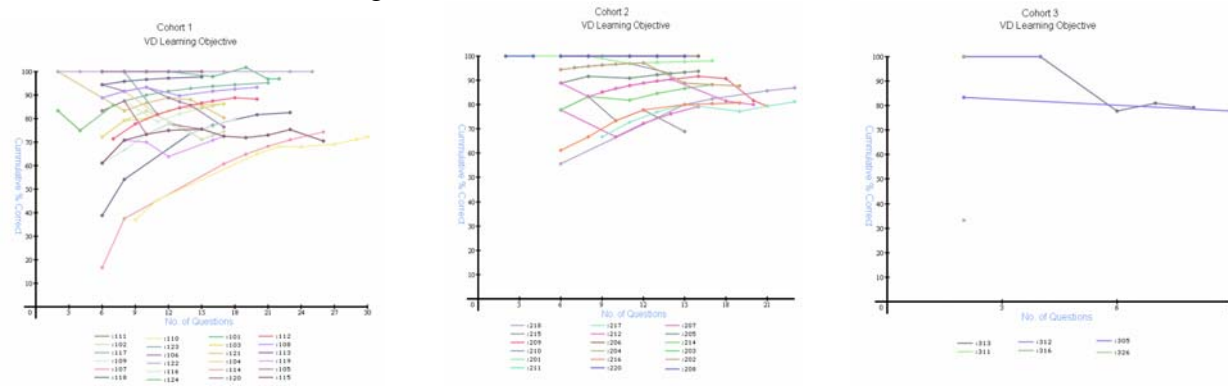


Figure 18. Thevenin Equivalent learning objective: Relationship between number of questions encountered and cumulative gain scores.



Limitations

Because a quasi-experimental design lacks random assignment, interpretation of findings must be made accordingly. The recent Institute of Education Sciences report (2003) noted that findings from quasi-experimental design can provide *possible evidence of effectiveness* if they meet guidelines such as (a) equivalent groups, (b) non-optional participation, (c) prior assignment to intact group, and (d) other standards such as low attrition, valid outcome measures and reporting of tests for statistical significance (p. 12). To a substantial extent, this study met these criteria. Groups were tested for equivalence on important variables; Cohort 2 differed on grade point average, which was controlled for as a co-variant in the analysis of variance and the advantage of the regression analyses is that it takes into account relevant variables. The study meet other standards as well: the intervention was administered as part of the course curriculum, no special criterion was used to assignment treatment to groups, attrition was low, and statistical tests were included in reporting. A limitation was that (a) no standardized achievement tests were used (however, tests that were used were those normally constructed by class instructors as well as problems created for *IMITS*) and (b) variability in implementation was not documented.

There was a possible diffusion of treatment as the instructor of one class mentioned the software program to students in the control group, but it is not known what effect that comment had on the students.

Because of this variability of testing and implementation, data from cohorts had to be analyzed separately. This approach necessitates multiple replications in order to establish effectiveness. Future effectiveness studies on IMITS can taken into account and correct for some of the limitations of these initial investigations.

Summary of Research Findings

Analysis of classroom achievement measures and gain scores on learning objectives derived from user log files yielded the following key findings on the effective of *IMITS*:

- A. There was support for the effectiveness of *IMITS* using classroom measures. For each cohort, the software group outperformed the control group (at a statistically significant level) on two classroom achievement measures: Labs #4 and #5 (DC #5 problem) (Cohort 1); Lab #4 and Exam #3 (Cohort 2), and DC assignments #3 and #4 (Cohort 3).
- B. There was very 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 10 learning objectives, the number of questions encountered predicted learning gain on that objective.

References

- Butz, B. P. (2001). The learning mechanism of the Interactive Multimedia Intelligent Tutoring System (IMITS). *Journal of Engineering Education*, 90, 543-548.
- 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.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Institute of Education Sciences (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington D.C.: Us. Department of Education, National Center for Education Evaluation and Regional Assistance, December
- 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.