An Experiment to Evaluate a Calculus CAI Package

Mark J. Kiemele, Lt Col, USAF
Department of Mathematical Sciences
United States Air Force Academy

I. INTRODUCTION

Advanced technology has provided an onslaught of new instructional materials and modes of operation for use in the classroom. The mathematics classroom is no exception, with a myriad of mathematical packages available that will do almost anything that modern technology will allow. While mathematical software packages abound, relatively little has been done to actually determine whether or not the new technology enhances the educational process, and if so, how. Like many mathematics departments across the nation, the Department of Mathematical Sciences at the United States Air Force Academy has been heavily involved in both the development and implementation of software packages that could be described as computer-aided-instruction or computer-enhanced-learning tools. A major part of our mathematics technology implementation process is our quality assurance program. In addition to software validity, an important goal of the quality assurance program is educational validity, that is, to determine if the implemented technology contributes to the educational environment in the manner in which it was intended.

In 1984 the USAF Academy initiated a program called the Microcomputer-in-the-Dorm(MID)/Local Area Network(LAN) project. The goal was for each cadet to have his or her own microcomputer in the dorm and to connect all faculty and cadet computing facilities via a LAN. At the same time, the Department of Mathematical Sciences (DFMS) undertook a project to develop a software package that could make as big an impact as quickly as possible. Since every incoming cadet (approximately 1400 each year) took three courses in calculus, it was decided to develop a Calculus Problem Generator. This package was not designed to replace the instructor in the classroom but rather to supplement the classroom instruction with a tool that could help cadets throughout their entire four-year stay--i.e., something they could use on their own without any interaction with the faculty. Thus, the mathematical package under scrutiny in this experiment is a calculus problem generator, which is an interactive program that can randomly generate more than 4 billion problems in 11 different functional areas of differential and integral calculus. Each problem is randomly generated at run time based on a user-chosen objective and one of three levels of difficulty. Problems are presented in multiple choice format, and a detailed step-by-step solution is presented only if the cadet asks for it or if the cadet answers the question incorrectly. At the end of each session, the student is provided a performance summary which includes the number of problems attempted and the number answered correctly, for each objective and for each level of difficulty within each objective.

II. PURPOSE AND DESCRIPTION OF THE EXPERIMENT

The general objective of the experiment was to evaluate the DFMS-developed Calculus Problem Generator when used as a review tool. The regular track for cadets in core mathematics comprised three calculus courses during their freshman year. After a summer of military training, cadets took Differential Equations, also a core course, in the fall semester of their sophomore year. Typically, the first several lessons in Differential Equations constituted a review of the calculus which, for many cadets, seemed to have become a bit rusty after a summer's layoff. Specifically, then, the purpose of this experiment was to answer the following questions:

(1) Can the Calculus Problem Generator be used as a viable alternative to the traditional in-class review?

(2) Is there any particular aptitude group whose performance is enhanced by application of this software package?

(3) Do cadets want to or like to use the CAI package and, if so, which cadets?

Additionally, we wanted to obtain data and feedback so that the Calculus Problem Generator could be improved.

The procedure used to accomplish these objectives was a controlled, completely randomized experiment in which each cadet was randomly selected and then randomly assigned to exactly one of the following three groups:

(1) Control Group - these cadets were not allowed to use the package and had to attend a traditional in-class review.

(2) Experimental Group - these cadets had to use the software package and were not allowed to attend in-class reviews.

(3) Swing Group - these cadets were allowed to choose what they wanted to do: use the software package or in-class review, but not both.

Due to limited computer resources, the Experimental and Swing groups had to be limited in number. The primary purpose of the Swing Group was to determine the interest level among cadets in the computer package. The experiment was conducted in the fall of 1985 and involved two courses: Math 211 (regular track for differential equations) and Math 200 (differential equations for the less technically inclined). Hence, the Math 200 group represented a population with a lower aptitude for the material than the Math 211 group. Additionally, out-of-class study times were obtained for each cadet and a questionnaire was administered to each cadet in order to help evaluate the software package. The response or criterion variable was the score on a common calculus review exam which consisted of 16 multiple choice problems (each worth 6 points) and 4 workout problems (each worth 10 points).

III. EXPERIMENTAL RESULTS

Table 1 summarizes the results of each of the experimental groups with regard to the mean number of points scored on the common calculus exam. The total scores are broken down into the multiple choice and workout portions of the exam. Also shown are the results for each of the aptitude groups considered: Math 211, shown in brackets, and Math 200, shown in parentheses. A total of 808 cadets participated in this experiment, with the numeric breakdown as shown in the table. Using the sample mean data from the control and experimental groups, a simple F-Test for the equality of the two population means was conducted. The results indicate that the two means are not significantly different at the $\alpha = .05$ level (P-value = .63). Another F-Test for homogeneity of variances was conducted, verifying the assumption that the variability within the two populations was the same (data for this test is not shown).

Table 1. SUMMARY OF MEANS*

	CONTROL	EXPERIMENTAL	SWING
n = sample size =	487	158	163
	[373]	[117]	[127]
	(114)	(41)	(36)
Multiple Choice: 96	86.86	87.19	86.28
	[90.23]	[91.23]	[89.48]
	(75.84)	(75.66)	(75.00)
Workout: 40	28.51	27.29	27.47
	[31.42]	[29.64]	[29.98]
	(19.00)	(20.59)	(18.61)
Total: 136	115.37	114.48	113.75
	[121.64]	[120.87]	[119.46]
	(94.84)	(96.24)	(93.61)

*Note: Math 211 data is shown in brackets, i.e., [Math 211 data]
Math 200 data is shown in parentheses, i.e., (Math 200 data)
The combined totals for both aptitude groups are shown without brackets or parentheses.

Table 2 summarizes the study time data that was collected from the cadets. The study times represent averages for the total amount of time spent studying outside of class. The data indicates that the Experimental Group had a 38% decrease in average study time, compared to their Control Group counterparts (59 vs 95). This difference is significant at the $\alpha = .001$ level. Table 2 also shows that the lower aptitude group (Math 200) spent more time studying while still falling short of the performance measure of the higher aptitude group. This result was not surprising.

Table 2. STUDY TIME SUMMARY

By Group:	Control	Exp
Avg Study Time (min)	95	59
By Course:	Math 211	Math 200
Avg Study Time (min)	79	87
Avg Score (136)	121.05	94.91

Table 3. STUDENT RESPONSES TO QUESTIONS

Strongly Disagree	Disagree	No Opinion	Agree	Strongly Agree
1	2	3	4	5

1. "I would have preferred to use the software package on the micros instead of the in-class review, if I had been given the chance to do so." $\bar{X} = 3.09$, $s = 1.16$

2. "The use of the software package was beneficial to my review of calculus." $\bar{X} = 4.31$, $s = .74$

3. "I would have preferred to have had the traditional in-class review instead of using the software." $\bar{X} = 2.29$, $s = .86$

Finally, a portion of the questionnaire responses is given in Table 3. The first question was asked of those cadets who were not allowed to use the computer package, while questions #2 and #3 were asked of all cadets who used the software package, be they in the Experimental or Swing Groups. The average response, as well as the standard deviation, is shown following each of the statements.

IV. CONCLUSIONS

Statistically, no significant difference was detected between the performance of the Control and Experimental Groups on any of the portions of the common calculus exam. However, the Experimental Group was able to achieve this same level of performance with about a 38% reduction in study time (95 vs 59). This is a significant finding, both statistically (as shown in Section III) and educationally, considering the magnitude of the time management problem imposed upon cadets at the Academy. Cadets liked to use the computer package. 72% of the Swing Group opted to use the software, and the questionnaire results indicate similar findings. While the experiment was too limited in scope to make wholesale decisions regarding the range of use of the software package, the Calculus Problem Generator does appear to be a viable review tool.