Measuring the Effects of a Web-Based Practice Program on the FCAT Math Results of 5th Graders

Bahia Naime Diefenbach, Ph.D.

Christopher B. Sullivan, Ph.D.

Paper to be delivered at the Florida Educational Research Association Annual Conference
November 7 - 9, 2001, Marco Island, Florida

Measuring the Effects of a Web-Based Practice Program On the FCAT Math Results of 5th Graders

Learning in a Multimedia Environment

Several national and international research studies indicate that limited, traditional teaching approaches still dominate the math classrooms (Schifter, 1997; Russell, 1997). Moreover, teachers tend to teach as they were taught (Ball & Wilson, 1990). The benefit of using traditional paper and pencil study materials to prepare for paper and pencil standardized tests is intuitively and experientially apparent to all teachers (Scholz, 1995). Typically, student achievement in traditional instructional programs was measured by either student ability to recall factual information or by proficiency at a procedural or computational task. As a result, the items precisely matched the program's learning objectives and measured only the content included in the program (Dick and Carey, 1990). Often, higher order thinking skills or the ability to transfer the knowledge to different contexts was not among the teaching goals. The difficulty in expanding teachers' practices may result from a lack of personal experience with the new approaches.

Educational technologies have evolved from the printed programmed instruction to computers driven by hypermedia. Research on the effectiveness of computer technologies has consistently revealed that, when used appropriately, computers make excellent learning tools (Snider, 1992; Herrington & Oliver, 1999). In the classroom, computer applications have varied from the provision of drill and practice for remediation to structured curriculum and instruction. Recent developments in multimedia present opportunities and challenges for educators who want to develop effective instructional programs. New software applications bring the promise of creating superior learning environments relative to the traditional classroom as well as delivering these learning experiences to a greater number of students and more diverse audiences.

In recent years, some curriculum designers have advocated using computer technology for knowledge exploration and construction. Some have also stressed the importance of creating more learner focused learning environments in which the learner is provided with varying amounts of help and support (Hannafin & Scott, 1998). Such learning environments enhance the learning of mathematical skills by providing the learner with interactivity, immediate feedback, control of the pace of instruction, and individualized learning (Hawkins, 1993). Mathematics educators are intuitively attracted to dynamic mathematical programs, sensing that powerful learning outcomes are possible (Goldenberg & Cuoco, 1996). If the appropriate software is carefully chosen, it will have the flexibility to accommodate a variety of student learning styles (Hawkins, 1993; Schank, 1993). By employing innovative, multimedia technology to teach mathematics, educators have the opportunity to improve learning. These multimedia learning environments are ideal for a stimulating higher order of thinking (Schank, 1993; Paolucci, 1998). These views have been discussed and documented at length by such noted cognitive theorists as Bloom (1956) and Gagne (1987).

Innovative multimedia and online software provide opportunities for students to use the computer more often at school and at home. In school and out of school learning become complementary. Studies show that students who receive appropriate, carefully chosen computer tutorials as homework assignments achieve better academic results than those who receive traditional textbook exercises as homework assignments (Sasser, 1990-91). Moreover, with appropriate guidance about the meaningful use of technology, elementary teachers may rethink how children learn as well as develop working knowledge of mathematics concepts, and get feedback on student performance (Kim and Sharp, 2000).

Mathematics education promotes the view that mathematics should be taught and assessed in a variety of meaningful and authentic ways. When students are actively involved in an activity they are more likely to learn the mathematics content of the activity (Schoenfeld, 1992). Research shows that simulations of real-life situations, in which learners must rely on mathematical knowledge and skills to solve problems, help students incorporate mathematics reasoning as an important cognitive activity to arrive at a solution (Verzoni, 1997). By applying math skills and theories in real life situations, students establish connections between school learning and their interests outside school. This process enhances critical thinking and mathematics learning skills and improves the retention and transfer of learning. Therefore, students learn how to construct knowledge, conceptualize problems, and develop problem-solving skills (Goldenberg & Cuoco, 1996). Some software can support complex learning in math (Nicaise, 1997) and teach problem solving to students who struggle with learning difficulties (Babbitt and Miller, 1996).

Considerable research has involved cognitive technologies, but little has focused on students' learning traits (Shute, 1993). Identifying such learner traits as the learners' preference of teaching style and the amount of instruction (Freitag and Sullivan, 1995; Hannafin and Sullivan, 1996) could have critical developmental and implementation implications (Hannafin and Scott, 1998), and could have a considerable impact on students' success in such environments.

Multimedia materials show promising effects on students' acquisition of knowledge and can enhance teaching and learning for today's diverse students (Torrez, 2000). An important feature of multimedia software applications is their interactivity with their user and their ability to provide important feedback. These capabilities could significantly improve learning. While feedback seems to be important in the enhancement of learning, research indicated that this is true only under certain conditions (Cooper, 1998; Khine, 1996). In a learning situation, feedback may be broadly defined as information obtained by students regarding the accuracy of their performance in a learning task. Different types of feedback can be categorized according to their functions and characteristics (Dempsey and Sales, 1993). Knowledge of results is the simplest level of feedback, which provides responses such as “right” or “wrong”, “correct” or “incorrect” without giving the correct answer. Elaborative feedback is a higher order of post-response information, which not only contains the result of a learner’s response, but also provides reasons for why the response was wrong and provides the correct answer. A third situation is where no feedback is provided. This forces the learners to proceed through the instructional sequence without receiving any post-response information on what the learner has attempted.

Studies have examined immediacy of feedback, the amount of information in feedback, the type of task involved, the importance of error analysis, and response certitude. Researchers agree that informative feedback does benefit learning and enhances performance for several types of learning tasks. The research suggests that feedback is more effective when it relates to the correct answer (Kulhavy and Wager, 1993; Khine, 1996).

Multimedia teaching technologies still have one main disadvantage among other, cost. Cost is a major factor affecting ownership and use of computers and other new technologies (Fahy, 2000). Teachers have little time to determine which CD or web-based program really works to prepare students, and the cost of such programs is high enough that a wrong choice also exhausts a scarce budget.

The FCAT Explorer: 5th Grade Math Program

The FCAT Explorer is an educational program provided to Florida public schools by the Florida Department of Education at no charge. The FCAT Explorer is an educational web site that provides innovative practice programs and instructional support tools to strengthen the skills students need for success on the FCAT, in the classroom, and in life. As a learning tool, the FCAT Explorer can be used in the classroom, at home, or at the library, wherever there is a computer with Internet access.

The FCAT Explorer is based upon the curriculum outlined in Florida's Sunshine State Standards. It complements the curriculum by furnishing additional problem solving opportunities that the students can access through the computer.

The FCAT Explorer was developed in partnership with successful teachers, education specialists, instructional designers and testing professionals. The educational materials were designed using effective learning strategies, direct instruction procedures, principles of effective instructional design and cognitive learning theory (Gagne, 1987)**.** Moreover, the instructional materials in the FCAT Explorer reflect critical thinking attributes, strategies for learning mathematics and solving problems, and contextual variables influencing the incorporation of motivational components such as the Keller’s ARCS (Attention, Relevance, Confidence, Satisfaction) model. The FCAT Explorer is an effective instructional tool, capturing and holding students’ interest with a lively use of color graphics and a variety of subject matters (Attention). It measures and enhances the learning of the 5th grade math skills and benchmarks using real-life situations (Relevance). It also builds the confidence of the learner by providing hints, immediate guiding feedback, and the explanation of the correct answer (Confidence). Satisfaction is gained through acquiring tokens, and playing instructional games and the FCAT Explorer reports which show the student's progress (Satisfaction).

Cognitive learning in the social and physical sciences is used to provide a rich, exploratory environment and to teach the students how to construct knowledge, conceptualize problems, and develop problem-solving skills. The FCAT Explorer makes mathematics learning more authentic by including real-life applications of math concepts. Through the informational items, students investigate mathematical problems embedded in a real-world scenario, enhancing the students' quality of communication about mathematical concepts. The FCAT Explorer provides the user (administrator, teacher, parent, mentor or student) with a number of practice items, instructional games, reports, links to the math benchmarks and skills and to the Sunshine State Standards.

Approximately 2/3 of the problems in the FCAT Explorer: 5th Grade Math are formatted as multiple-choice problems. Approximately 1/3 of the problems are gridded response. One other type of problem is the charted response in which students practice FCAT-like graph construction with three sets of charted response problems. Each set consists of four problems that lead students to interactively construct bar graphs.

Each item in the FCAT Explorer consists of a stem, four distracters (for multiple-choice items) or a grid (for gridded items), guidance feedback for incorrect answers, a hint to direct the student's thinking to understand and reason through the problem, a glossary of mathematical terms, and an explanation of the correct answer to reinforce and improve their skills. When students incorrectly respond to a multiple choice problem or gridded response problem, they will be presented with answer-specific feedback or skill-specific feedback and hints. Students will then be given another opportunity to answer the problem. If they respond incorrectly a second time to a multiple choice or gridded response problem, the FCAT Explorer provides an explanation of the correct answer. An explanation of the correct answer is also provided if students correctly respond to a multiple choice or gridded response problem, to reinforce one or more correct solution techniques.

Students earn one game token each time they correctly answer a teacher-determined number of practice problems. When students have earned one or more game tokens, they may enter the Explorer's Arcade and select a game. The number of game tokens one has earned at any given time is displayed in the illustration of a game token that is shown on all of the navigation and problem pages. Each game reinforces a selection of the skills derived from the 34 Benchmarks. The games are timed to ensure that students don't spend too much time away from the practice problems. Game time limits range from five to eight minutes.

The various components of the FCAT Explorer generate an instructional program that is designed to accommodate individual learner traits to improve learning. It encourages students to use their thinking abilities to process learning at a higher level of complexity, and it teaches them how to organize content and knowledge to facilitate more complex processing. Moreover, it sharpens the students' skills to cultivate and develop their thinking. Improve learning efficiency and to address individual differences.

Background of the Study

The FCAT Explorer: 5th Grade Math program is offered to all Florida public school students by the Florida Department of Education. The program consists of an organized series of math practice items that the student answers online. In the final version of the FCAT Explorer**:** 5th Grade Math there are 148 items, including multiple choice, gridded response and extended response problems. The set of math items in the program are written to the 5th grade math benchmarks specified in the Sunshine State Standards, so that a student who works all the way through the program will cover items pertaining to all the benchmarks included in the FCAT. The student home page, or Explorations and Themes Generator, is shown in Figure 1.

Figure 1 Exploration and Themes Page of the FCAT Explorer: 5th Grade Math

The FCAT Explorer: 5th Grade Math was developed with the help of a group of elementary schools that were used as a pilot study group for the program. In the pilot study, there were 80 5th grade math items for the students to practice. The pilot study ran from January 8, 2001 through March 15, 2001, just after the completion of the 5th grade math FCAT. Schools that were approached for the pilot had already participated in the FCAT Explorer: 4th Grade Reading pilot study, and were recommended by educational professionals at a number of state level organizations. An equal number of schools was approached in grade levels A through D. Some schools did not accept the offer to be in the pilot study, and some schools that agreed to participate in the pilot study dropped out of the study midway through. The pilot study started with 22 schools and ended with 17 schools participating.

The fifth grade teachers in each pilot school were given a two hour training session in the FCAT Explorer: 5th Grade Math, and were helped to enroll their students in the program. Once enrolled, the students in each pilot school had access to the program, either at school or at home. Teachers were encouraged to contact the FCAT Explorer development team with any concerns, problems or observations that they might have; they could use a "Contact Us" link in the program, or they could call toll free. In addition, a liaison teacher in each of the pilot schools was contacted every two weeks to follow up on questions or concerns.

Student activity in the FCAT Explorer: 5th Grade Math program was tracked by recording the number of students actively using the site on a weekly basis and tracking the number of answers that students made to the items in the program. The outcome (correct or incorrect) of answers was also recorded, as well as the answers that students gave to each item. These data were used to conduct an item analysis after the pilot study was complete, and to calculate the P-values, or percent of students answering an item correctly, for each item, benchmark and strand. An overview of the schools participating in the pilot study, the number of students active in the program and the number of items they answered during the pilot study is given in Table 1.

Of the data from the pilot study, only the p-value calculations by strand are used in this study. These data are used to demonstrate the increment in a student's learning from working through the items in the program. The other datum used from the pilot study in this study is the fact that each of the 17 schools in the pilot study provided access to the program for their students. This creates a unique sub-population in Florida's schools, one that is compared to the rest of the schools in the state, none of which had access to the FCAT Explorer: 5th Grade Math.

Table 1. FCAT Explorer: 5th Grade Math Pilot Schools and Number of Items Answered

School Grade Level	Participating School	Students Participating	Total Items Answered	Average Items Answered
A	School 1	234	30,110	128.7
A	School 2	110	4,557	41.4
A	School 3	104	7,224	69.5
A	School 4	97	3,134	32.3
A	School 5	90	2,085	23.2
A	School 6	87	2,412	27.7
A	School 7	51	1,764	34.6
A	School 8	17	1,767	103.9
B	School 9	106	1,450	13.7
C	School 10	151	16,923	112.1
C	School 11	69	4,405	63.8
C	School 12	66	1,289	19.5
C	School 13	63	1,429	22.7
C	School 14	42	435	10.4
D	School 15	118	15,525	131.6
D	School 16	57	1,154	20.2
D	School 17	43	2,166	50.4
	Grand Total All Schools	1,505	97,829	53.3

Methodology for Analyzing the Effect of FCAT Explorer: 5th Grade Math

The foci of analysis for this paper are in two places, the learning of 5th grade math benchmarks and then demonstrating that knowledge on the State of Florida FCAT 5th grade math test. The FCAT Explorer pilot study ran from January to March, 2001, providing 5th grade students with a two month window in which to practice for the FCAT. The first focus compares the performance of the 17 FCAT Explorer: 5th Grade Math pilot schools, which used the program from January to March, 2001, on the year 2000 and year 2001 FCAT to the performance of the rest of Florida's elementary schools, that had no access to the FCAT Explorer. The second focus identifies the presence of a learning component in a computer-based learning program like the FCAT Explorer: 5th Grade Math. This step uses performance data from the 17 schools that participated in the pilot study of the program.

These two foci of study can be summarized in the following research questions:

1. Does use of the FCAT Explorer: 5th Grade Math pilot program lead to increased knowledge of the 5th grade math benchmarks?
2. Does the FCAT Explorer: 5th Grade Math pilot program provide an effective learning environment for 5th grade students?

For the first set of comparisons, the 2000 and 2001 mean school FCAT scores for 5th grade math were downloaded for all schools in Florida from the Florida Department of Education website (http://www.firn.edu/doe/sas/fcat/fcpress1.htm). The dataset available from this site lists the school-level FCAT mean scores for all of Florida's Elementary schools, for all grades. The FCAT mean scores for a school are created by averaging all the student FCAT scores for each grade level. In this study, the FCAT mean scores for 5th grade math were used. A pilot study sample of the 17 schools was formed, and the mean FCAT scores for these school are compared to all other Florida elementary schools, none of which had access to the FCAT Explorer.

There are several ways that the FCAT Scores can be compared. One is by direct comparison of the FCAT mean scores; in other words, the averaged 5th grade math pilot school FCAT mean scores are compared to the FCAT mean scores for all other schools. This is done for year 2000 and year 2001 FCAT. The second comparison looks at the amount of change in FCAT mean scores between 2000 and 2001, for the 5th grade pilot schools and all other schools. This is done by calculating the difference between each school's 2001 mean FCAT score and its 2000 mean FCAT score for 5th grade math; the figures are then averaged for each of the two sample groups.

This method is used to test for the positive effects of the FCAT Explorer: 5th Grade Math on FCAT scores. For both methods of comparison, difference of means tests are run to determine whether any differences between the pilot school sample and the rest of Florida's elementary schools are significant.

To demonstrate the presence of a learning component that is related to the use of the FCAT Explorer: 5th Grade Math program, pilot study data are used that log the number of correct answers made to the math items. The ratio of correct answers to all answers is the p-value of the item, or the average p-value of the benchmark or strand. This study presents the p-values by math strands for two types of answers in the FCAT Explorer: 5th Grade Math program, multiple choice and gridded response. The 5th Grade Math Strands include:

- Strand A: Number Sense, Concepts, and Operations
- Strand B: Measurement
- Strand C: Geometry and Spatial Sense
- Strand D: Algebraic Thinking
- Strand E: Data Analysis and Probability

An extra dimension to the p-value scores is provided by the interactivity of the FCAT Explorer computer based learning program. When the FCAT Explorer: 5th Grade Math program presents an item to the student in the form of a math word problem, the student enters the answer, either by selecting a letter from a multiple choice selection or typing in the answer on a gridded response entry grid. The program assesses the response and lets the student know if the answer was correct or incorrect. In either case, the program presents guidance feedback to the student.

If the student answers an item correctly, he or she is brought to a "correct answer explanation" web page. This page contains a correct solution for calculating the answer, sometimes offering more than one way to solve a problem. The guidance feedback explains and reinforces the correct answer. The student then moves on to the next item in the sequence.

If the student answers the item incorrectly, the program presents the student with "common error explanation" that is related to the common type of mistake given in the distracter answers in the multiple choice items, or calculated for the gridded response items. This guidance feedback explains why that specific answer is incorrect, and then provides a hint to direct the student's thinking to figure out the correct answer. The student is then brought back to the item and allowed to answer it again. Whether the student answers the question correctly or incorrectly the second time, he or she is brought to the "correct answer explanation" web page that explains and reinforces the correct answer. The student is then moved to the next item in the sequence.

From the cycle of having two attempts to answer an item comes the opportunity to compare the p-values for the student's first attempt and second attempt. Some of the students who answered incorrectly on their first attempt will answer correctly on the second try. Based on the school level data available from the pilot study, the percents of students answering correctly on their first and second attempts are compared for multiple choice and gridded response items. The results of these comparisons should indicate whether the FCAT Explorer: 5th Grade Math provides an effective learning resource for 5th grade math benchmark information. The pattern of employing common error feedback, hints and glossary terms for an incorrect answer, and correct answer explanation for all other answers, was used consistently throughout the FCAT Explorer: 5th Grade Math pilot program.

In addition to the guidance feedback, the pilot program had three educational games that were accessed through the use of tokens. Every time a student answered a question correctly, he or she received points toward a token – more points if the answer was correct the first time through, fewer if it was the second time through. When the student earned a token he or she could play one of three math games that reinforced calculation skills and math concepts. The games themselves focused on specific learning related to the benchmark.

The summary effect of the FCAT Explorer: 5th Grade Math on the 5th grade math skills of the pilot school students was expected to be positive, given all of the opportunities to learn math that were embedded in the program. The only reliable test data available for the pilot study was the March, 2001, FCAT 5th grade math test. Hence, the measures used for the second focus of the study come from the FCAT scores posted by the Department of Education, following the FCAT in year 2001.

FCAT Explorer: 5th Grade Math Pilot Study Results

Using data from the FCAT Explorer: 5th Grade Math pilot study, two analyses are conducted in this section. The first analysis compares the mean school FCAT scores for the 17 schools in the FCAT Explorer: 5th Grade Math pilot study with the FCAT scores of all other Florida schools. Difference of means tests is used to establish whether the difference in scores between the two groups is statistically significant. The second analysis examines the incremental proportion of students who make a correct answer to an item on the second try at it. This analysis compares two types of items, multiple choice and gridded response items, and interprets the increase in correct answers on second try as an indicator of software-related learning.

The first analysis of the effects of the FCAT Explorer: 5th Grade Math uses the actual FCAT scores for year 2000 and year 2001 tests to examine whether use of the pilot program made a difference in the actual 5th grade math FCAT scores. For this analysis the FCAT Explorer: 5th Grade Math pilot study schools were grouped together and compared to all other Florida schools,

Figure 2. Comparison of 5th Grade Math FCAT Scores for FCAT Explorer: 5th Grade Math Pilot School and All Other Schools, 2000 and 2001

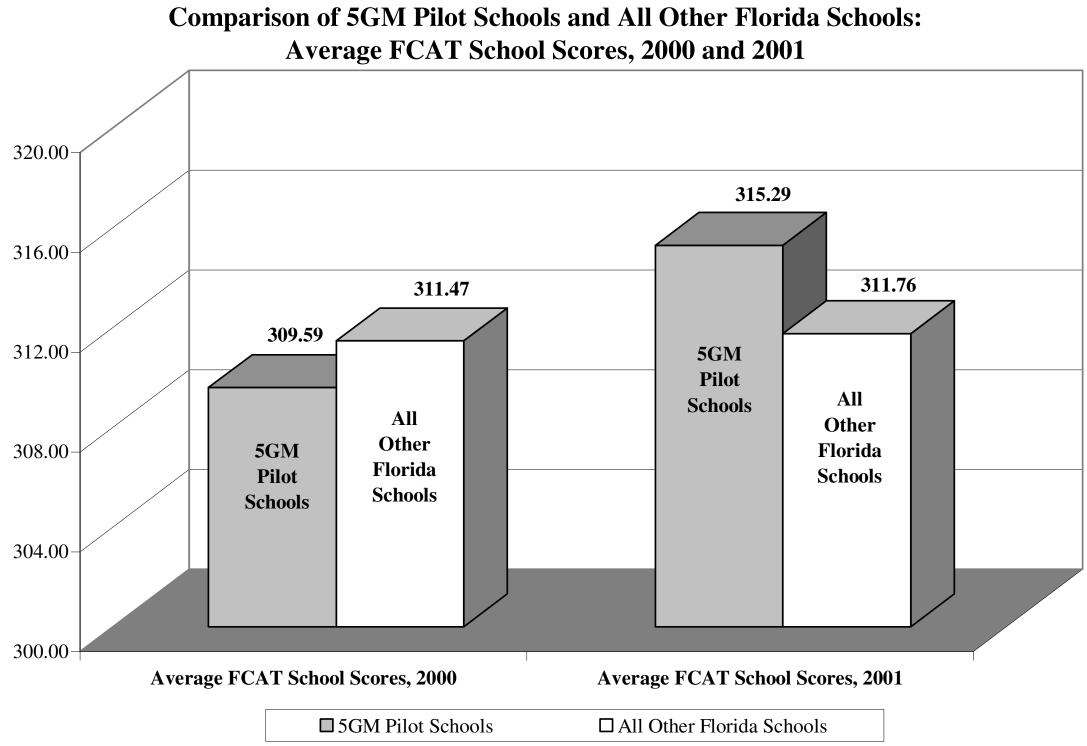


Table 2. Difference of Means Test Results Comparing 2001 5th Grade Math FCAT Scores for FCAT Explorer: 5th Grade Math Pilot School and All Other Schools

Independent Samples Test - FCAT Mean Scores, 2001

FCAT Mean Scores, 2001	N	Mean	Std. Deviation	Std. Error Mean
5GM Pilot Schools	17	315.29	30.76	7.46
All Other Florida Elementary Schools	1,542	311.76	25.40	0.65

	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	2.67	0.102	0.57	1557.00	0.569	3.54	6.21	-8.64	15.71
Equal variances not assumed			0.47	16.24	0.643	3.54	7.49	-12.32	19.39

none of which had access to the program. Figure 2 displays the results of this inquiry, showing the comparative 5th grade math FCAT scores for year 2000 and year 2001. The scores are followed by a difference of means test intended to determine if there is any statistically significant difference between the FCAT Explorer: 5th Grade Math pilot schools and all other schools.

From inspection of the column chart, it is clear that the FCAT Explorer: 5th Grade Math pilot schools did increase their FCAT scores between years 2000 and 2001, rising from an average of 309.6 in 2000 to 315.3 in 2001. Compared to the pilot schools, the other schools in Florida did not fare as well, increasing their FCAT scores by less than a point. Although pilot school FCAT scores indicate a better performance, the difference of means test for 2001 scores, shown in Table 2, indicates that there is no statistically significant difference between the two groups.

While the difference of means test did not indicate a significant difference, the additive evidence shows that the pilot school group did much better than the rest of Florida's schools, suggesting that the FCAT Explorer: 5th Grade Math program made some difference in the scores. A second approach to examining the effect of using the 5th grade math program is to look more closely at the difference in 5th grade FCAT scores in the pilot schools and in all of the other schools. To do this, a new field was created for each school that consisted of the school's means FCAT 2000 score subtracted from the school's mean FCAT 2001 score. The resulting number represented the positive or negative change in the FCAT score from one year to the next.

The average change in 5th grade FCAT school scores for the FCAT Explorer: 5th Grade Math pilot schools and all other schools is shown in Figure 3. In this chart it is clear that, on average, the pilot schools outperformed all of the other schools by more than five points. Such an improvement is quite impressive, considering that the group of other schools comprises most of

Figure 3. Comparison of the Change in 5th Grade Math FCAT Scores for FCAT Explorer: 5th Grade Math Pilot Schools and All Other Schools, from Year 2000 to Year 2001

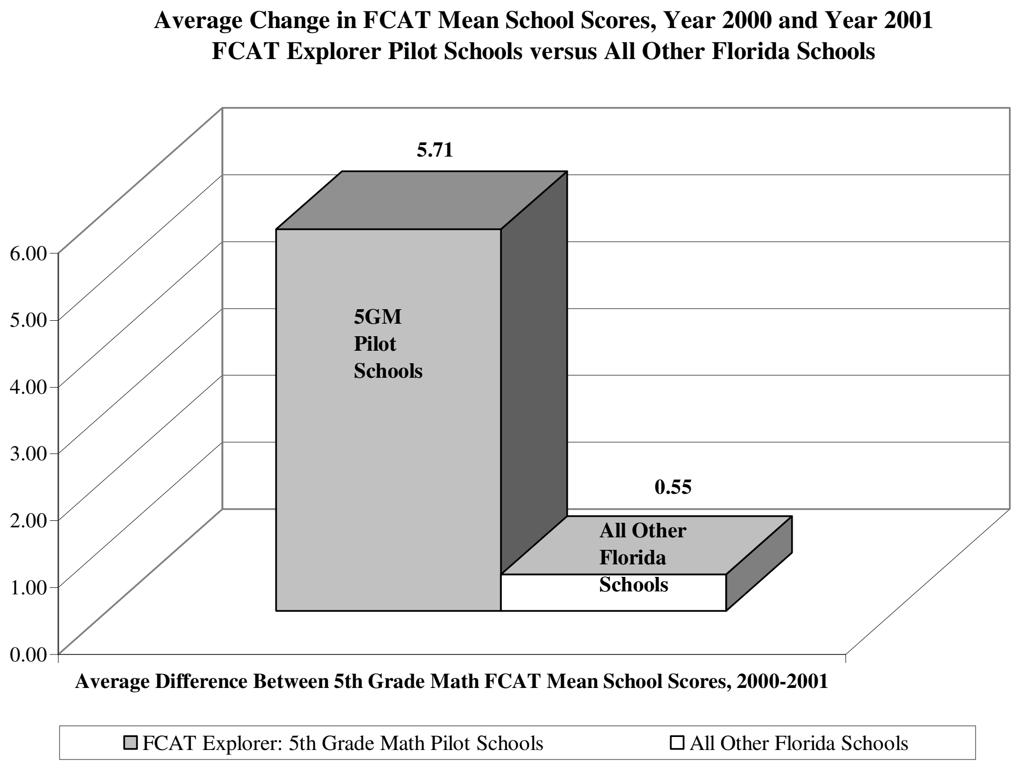


Table 3. Difference of Means Test Comparing the Difference in 5th Grade Math FCAT Scores for 5th Grade Math Pilot Schools and All Other Schools, from Year 2000 to Year 2001

Group Statistics - Change in FCAT Mean Scores Between 2000 - 2001

Change in FCAT Mean Scores Between 2000 - 2001	N	Mean	Std. Deviation	Std. Error Mean
5GM Pilot Schools	17	5.71	7.94	1.93
All Other Florida Elementary Schools	1536	0.55	13.34	0.34

Independent Samples Test - Change in FCAT Mean Scores Between 2000 - 2001

	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	4.76	0.029	1.59	1551.00	0.112	5.16	3.24	-1.20	11.52
Equal variances not assumed			2.64	17.01	0.017	5.16	1.96	1.03	9.29

Florida's elementary schools.

To further examine the change data, a difference of means test was conducted to see if the difference between the pilot schools and all other schools was statistically significant. As it turns out, the pilot school change scores are statistically significant from all other school change scores (see Table 3). Because the variances of the two groups are not equal, a 2-tailed test is used to determine whether the t-test is significant; the reported significance in this test is .017, clearly less than the .05 confidence limit. If this finding is accepted, then clearly there is some factor that influenced the pilot schools to perform better on the FCAT. We would argue that it is the FCAT Explorer and the educational resources available in it.

As a check comparison in this analysis, two groups of schools were created to determine if the pilot school FCAT scores differed from what might be expected from a similar number of randomly selected schools. The first sample was based on a random sample of schools, matched in the same number of Grade A through D schools as the pilot school group. The second sample group was a 1% random sample of the dataset. In both cases, there was little difference between the mean scores of the comparison groups and there were no statistically significant findings.

The second analysis compares the performance of 5th grade students using the FCAT Explorer. The pilot program included multiple choice, gridded response and charted response items. Only the first two are reported on here. Multiple-choice items were the most numerous, comprising 73.6% of all items. The items were distributed unevenly across the mathematics strands, so that the students could answer more items from some strands than from others. The distribution of items across strands is given in Table 4. It is important to keep this distribution in mind when examining the results of the P-value analysis reported below. Because the number of items in each strand is relatively small, particularly for the gridded response items, one should be cautious in evaluating student responses to these items.

Table 4. Number of Items in Each Strand in the FCAT Explorer: 5th Grade Math Pilot Study

	Gridded Response Items		Multiple Choice Items		Total Multiple Choice and Gridded Response Items	
Strand	Number of Items	Percent of All Items	Number of Items	Percent of All Items	Number of Items	Percent of All Items
Strand A: Number Sense, Concepts, and Operations	6	8.3%	13	18.1%	19	26.4%
Strand B: Measurement	5	6.9%	7	9.7%	12	16.7%
Strand C: Geometry and Spatial Sense	3	4.2%	12	16.7%	15	20.8%
Strand D: Algebraic Thinking	2	2.8%	7	9.7%	9	12.5%
Strand E: Data Analysis and Probability	3	4.2%	14	19.4%	17	23.6%
Total	19	26.4%	53	73.6%	72	100.0%

The limitations of these data aside, some interesting observations can be made by examining the pattern of responses to the different math strands. Figure 4 presents the percentage of correct answers on first and second try for multiple choice items distributed proportionately for all of the students who answered the items. (See Appendix 1 for a more detailed table of correct answers by school grade for the pilot school sample.)

Looking at the pattern of correct first attempts at answering multiple-choice items, most strands hover around the average percentage of 52.0% correct answers on first attempt. On the second attempt, an average of 22.4% students answered the multiple-choice items correctly. We assume that this percentage represents the number of students who paid attention to the

Figure 4. Comparison of P-values for First and Second Attempts on Multiple Choice Items

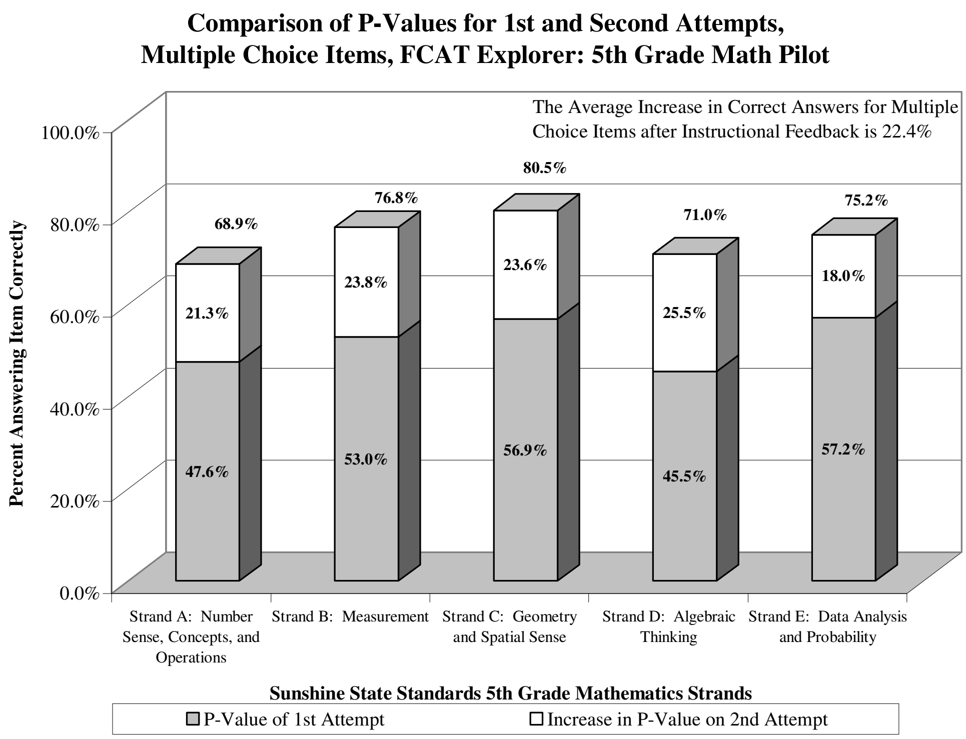
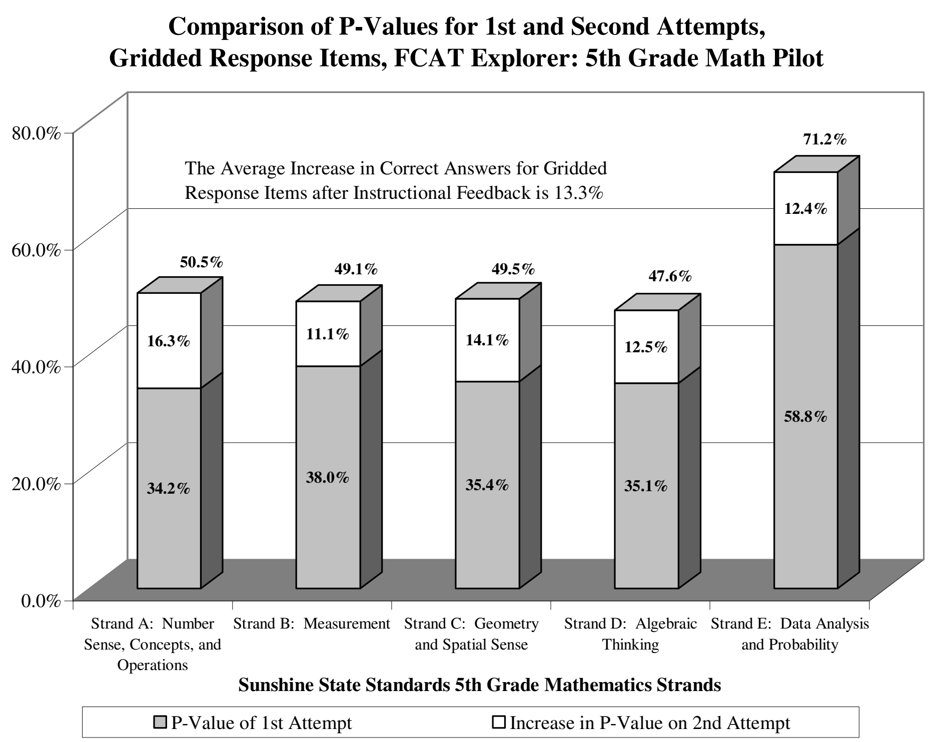


Figure 5. Comparison of P-values for First and Second Attempts on Gridded Response Items



guidance feedback after the first incorrect answer, and were able to select the correct answer a second time around. If this is the case, then this percentage provides an additive measure of the effect of using a software program such as the FCAT Explorer: 5th Grade Math. In other words, 22.4% more students can answer the item correctly than could the first time they tried to answer the math problem. This additional percentage must in some way be related to using the software program between the first and second attempts at the problem. To the extent that the increase is due to the effectiveness of the FCAT Explorer: 5th Grade Math, the p-values of the first and second correct answers track in some way the effect of using the learning software.

The second set of p-values for the gridded response items displays a similar story. Figure 5 shows the percentage of correct answers on first and second attempts for the gridded response items, distributed proportionately for all of the students who answered the items. The p-values of these items are much lower than those of the multiple-choice items. (See Appendix 2 for a more detailed table of correct answers by school grade for the pilot school sample.)

The percentage of students who answered correctly on the first attempt fluctuates more dramatically than with the multiple choice questions; however, the small number of items in each strand also affects the totals, so any conclusions should be made with caution. Another factor in the gridded response items is that the answer must be exact, rather than selected from a list, which could drive down the number of correct first answers. The average percentage of students answering correctly the first time around is 40.3% and the number of students answering correctly the second time is only 13.3%. However, this 13% is still an additive indication that the FCAT Explorer: 5th Grade Math provided enough learning guidance for this percentage of students to calculate the correct answer when given a second opportunity.

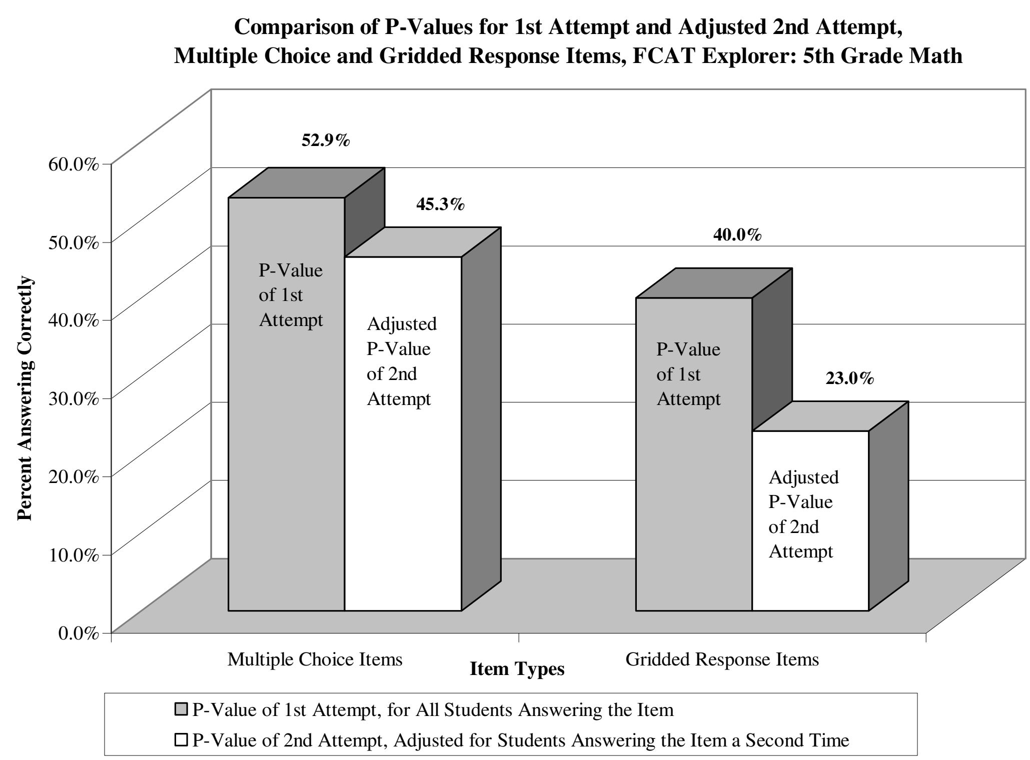
One finding from these two charts is that there is a fairly consistent percentage of students who will answer an item correctly, though this differs somewhat between the multiple choice and gridded response items. There is also a solid core of students who can answer an item correctly on the second try, the percentage again differing by item type. In the FCAT Explorer: 5th Grade Math, the period between answering an item the first time and answering it a second time is filled with math guidance feedback, intended to help the student understand how to figure out the math problem. It is up to the student to attend to this educational resource material, but it appears that about 13% to 22% of all the students could correctly solve the math item after reading the math hints.

There is one downside to the analysis of Figures 4 and 5. They represent percentages that are reflective of all of the students who answered items, but in the second attempt items they do not reflect the students who actually attempted the items a second time. If we compare the percentages of only those students who attempted the items, a different pattern emerges. Table 5 and Figure 6 present the results of this analysis. The p-value of the first attempt is unchanged, because that is derived from all of the students answering the item. However, the p-values of the second attempt show a very different picture. For the multiple-choice items, the p-value of the second attempt doubles, to 45.3%. This percentage is a closer estimate of the proportion of students able to answer the item correctly after guidance feedback, and it beats the one in three chance that the student would answer correctly through a lucky guess. The increase in the p-value for gridded response questions is even more heartening. While only ten percentage points higher than in Figure 5, there is little probability that the student could have answered correctly by chance, so all of the gains are due to the guidance feedback in the FCAT Explorer. These findings point to a clear effect on students using the FCAT Explorer during the pilot study.

Table 5. Comparison of P-values for Only Those Student Who Attempted to Answer Items

Strand	Number Attempting Item on First Attempt	Number Answering Correctly on 1st Attempt	P-value of 1st Attempt	Number Attempting Item on Second Attempt	Correct 2nd Attempt	P-value of 2nd Attempt
Multiple Choice Items						
Strand A	6,551	3,118	47.6%	3,433	1,395	40.6%
Strand B	2,709	1,436	53.0%	1,273	656	51.5%
Strand C	4,051	2,294	56.6%	1,757	978	55.7%
Strand D	2,915	1,327	45.5%	1,588	743	46.8%
Strand E	9,170	5,249	57.2%	3,921	1,646	42.0%
Total	25,396	13,424	52.9%	11,972	5,418	45.3%
Gridded Response Items						
Strand A	3,659	1,253	34.2%	2,406	594	24.7%
Strand B	2,073	787	38.0%	1,286	231	18.0%
Strand C	1,328	470	35.4%	858	188	21.9%
Strand D	809	284	35.1%	525	101	19.2%
Strand E	1,858	1,097	59.0%	761	227	29.8%
Total	9,727	3,891	40.0%	5,836	1,341	23.0%

Figure 6. Comparison of Adjusted P-values for Students Who Answered Items



Conclusions

The data collected in this study were used to determine the efficacy of the FCAT Explorer as a web-based educational resource. There were several findings reported that point to its effectiveness. The first comparison of FCAT mean school scores between year 2000 and year 2001 showed the schools that had used the FCAT Explorer: 5th Grade Math had a better average score than all other schools in 2001. Next, the 5th grade math pilot schools were compared to all other elementary schools in Florida in terms of the change in FCAT mean school scores between year 2000 and year 2001. As suggested by the first comparison, the change in FCAT mean school scores was greater by over five points for the 5th grade math pilot schools, and the difference was statistically significant. Clearly, something is at play in the difference in FCAT scores, which we argue is use of the FCAT Explorer.

The second analysis used score data to generate primary and secondary sets of p-values for each math item. The first p-value measured the success with which students answered each math item on first encounter. The second p-value measured the increase in the total number of students who answered correctly on the second try. On the first try, 52% of the students answered the multiple choice items correctly, and 40% answered the gridded response questions correctly, These total percentages increased by 22% and 13% respectively after students were exposed to the guidance feedback. This increase was interpreted as an additive effect due to using the FCAT Explorer.

One drawback of the p-value analysis above is that the second attempt percentage is taken from the total number of students attempting the items rather than only from those students who actually attempted an answer on second try. An analysis was made of the percentage of success on the second try, compared to chance alone. Figure 6 shows that of those students who had a second try, an average of 45% answered correctly. This compares favorably with the 33% that would be expected by chance alone. Similarly, for the gridded items, on second try the percentage of correct answers was 23%. There is little probability that the students could have answered this well by chance. This indicates that the FCAT Explorer's constructive feedback to incorrect answers has a positive effect on the students' ability to solve the problem.

The conclusions of this study underline the positive effect of using a web-based program on learning math. Using the FCAT Explorer clearly enhances students' math skills as reflected on their performance in the year 2001 FCAT. Moreover, the various components of the FCAT Explorer practice items such as feedback, hint, glossary term definitions, graphics, and the explanation of the correct answer have helped students direct their thinking, learn math strategies and reason in different ways to solve a specific problem. The FCAT Explorer has reached its goal of helping students master the 5th grade benchmarks, by motivating them to tackle math problems presented in real-life situations.

Appendix A.
P-values of FCAT Explorer: 5th Grade Math Pilot Study Answers to Multiple Choice Items

School Grade	Correct 1st Attempt	P-value of 1st Attempt	Correct 2nd Attempt	P-value of 2nd Attempt	P-value of 1st and 2nd Attempt Combined	Total
Strand A: Number Sense, Concepts, and Operations						
A	1,874	48.7%	844	21.9%	70.6%	3,851
B	36	35.0%	21	20.4%	55.3%	103
C	681	44.5%	337	22.0%	66.5%	1,530
D	527	49.4%	193	18.1%	67.5%	1,067
Total	3,118	47.6%	1,395	21.3%	68.9%	6,551
Strand B: Measurement						
A	940	56.5%	404	24.3%	80.7%	1,665
B	18	36.0%	12	24.0%	60.0%	50
C	332	48.9%	167	24.6%	73.5%	679
D	146	46.3%	73	23.2%	69.5%	315
Total	1,436	53.0%	656	24.2%	77.2%	2,709
Strand C: Geometry and Spatial Sense						
A	1,741	57.1%	747	24.5%	81.6%	3,048
B	28	71.8%	3	7.7%	79.5%	39
C	525	54.5%	228	23.7%	78.1%	964
D	0	0.0%	0	0.0%	0.0%	0
Total	2,294	56.6%	978	24.1%	80.8%	4,051
Strand D: Algebraic Thinking						
A	806	46.2%	467	26.8%	73.0%	1,743
B	9	32.1%	9	32.1%	64.3%	28
C	308	45.6%	175	25.9%	71.4%	676
D	204	43.6%	92	19.7%	63.2%	468
Total	1,327	45.5%	743	25.5%	71.0%	2,915
Strand E: Data Analysis and Probability						
A	3,367	59.1%	1,061	18.6%	77.7%	5,697
B	79	46.7%	23	13.6%	60.4%	169
C	1,052	53.4%	355	18.0%	71.4%	1,971
D	751	56.3%	207	15.5%	71.9%	1,333
Total	5,249	57.2%	1,646	17.9%	75.2%	9,170

Appendix B
P-values of FCAT Explorer: 5th Grade Math Pilot Study Answers to Gridded Response Items

School Grade	Correct 1st Attempt	P-value of 1st Attempt	Correct 2nd Attempt	P-value of 2nd Attempt	P-value of 1st and 2nd Attempt Combined	Total
Strand A: Number Sense, Concepts, and Operations						
A	791	35.5%	411	18.4%	53.9%	2,228
B	11	23.9%	10	21.7%	45.7%	46
C	258	30.2%	112	13.1%	43.3%	854
D	193	36.3%	61	11.5%	47.8%	531
Total	1,253	34.2%	594	16.2%	50.5%	3,659
Strand B: Measurement						
A	509	41.0%	161	13.0%	53.9%	1,242
B	10	43.5%	2	8.7%	52.2%	23
C	138	29.0%	51	10.7%	39.7%	476
D	130	39.2%	17	5.1%	44.3%	332
Total	787	38.0%	231	11.1%	49.1%	2,073
Strand C: Geometry and Spatial Sense						
A	321	36.9%	128	14.7%	51.7%	869
B	4	36.4%	1	9.1%	45.5%	11
C	83	31.4%	36	13.6%	45.1%	264
D	62	33.7%	23	12.5%	46.2%	184
Total	470	35.4%	188	14.2%	49.5%	1,328
Strand D: Algebraic Thinking						
A	190	37.7%	74	14.7%	52.4%	504
B	3	30.0%	0	0.0%	30.0%	10
C	59	33.0%	23	12.8%	45.8%	179
D	32	27.6%	4	3.4%	31.0%	116
Total	284	35.1%	101	12.5%	47.6%	809
Strand E: Data Analysis and Probability						
A	777	62.2%	161	12.9%	75.1%	1249
B	23	45.1%	3	5.9%	51.0%	51
C	274	54.0%	60	11.8%	65.9%	507
D	23	45.1%	3	5.9%	51.0%	51
Total	1,097	59.0%	227	12.2%	71.3%	1,858

References

Babbitt, B. C. and Miller, S. P., (1996). Using hypermedia to improve the mathematics problem-solving skills of students with learning disabilities. Journal of Learning Disabilities v 29 (July '96) p. 391-401.

Ball, D. L., & Wilson, S.M. (1990). Knowing the subject and learning to teach it: Examining assumptions about becoming a mathematics teacher (Research report 90-7). East Lansing, MI: National Center for Research on Teacher Learning, Michigan State University.

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals, Handbook I: Cognitive domain. New York: Longman & Green.

Cooper, S. B. (1998). Instructor-created computer tutorials for students in an elementary mathematics education course. Journal of Computing in Childhood Education v 9 no1 p. 93-101.

Dempsey, J. and Sales, G. (1993) (Edited). Interactive Instruction and Feedback. Englewood Cliffs: Educational Technology Publications.

Dick, W., & Carey, L. (1990). The systematic design of instruction. Glenview, IL: Scott, Foresman.

Fahy, P. J. (2000). Achieving quality with online teaching technologies. Paper presented at Quality learning 2000 Inaugural International Symposium, Calgary, Alberta, March 2000. (ERIC, ED 445 197.

Freitag, E. T., & Sullivan, H. J. (1995). Matching learner preference to amount of instruction: An alternative form of learner control. Educational Technology Research and Development, 43(2), 5-14.

Gagne, R. M. (1987). The conditions of learning (4th ed.). New York: Holt, Rinehart, and Winston.

Goldengerg, E. P., & Cuoco, A. (1996). What is dynamic geometry? In R. Lehler & D. Chazan (Eds.), Designing learning environments for developing understanding of geometry and space. Hillsdale, NJ: Erlbaum.

Hannafin, R. D., & Sullivan, H. J. (1996). Learner preferences and learner control over amount of instruction. Journal of Educational Psychology, 88, 162- 173.

Hannafin, R. D.; Scott, B. N. (Sept./Oct. 1998). Identifying critical learner traits in a dynamic computer-based geometry program. The Journal of Educational Research (Washington, D.C.) v. 92 no1 p. 3-12.

Hawkins, J. (1993). Technology and the Organization of Schooling. Communications of the ACM, 36 (5), 30 – 34.

Herrington, J. & Oliver, R. (1999). Using Situated Learning and Multimedia to Investigate higher-Order Thinking. Journal of Interactive Learning Research; v 10 no1 p3-24.

Khine, M. S. (1996), The Interaction of cognitive Styles with varying levels of feedback in multimedia presentation. International Journal of Instructional Media v. 23 no3 p. 229-37.

Kim M. K.; Sharp, J. M. (2000). Investigating and measuring preservice elementary mathematics teachers' decision about lesson planning after experiencing technologically-enhanced methods instruction. The Journal of Computers in Mathematics and Science Teaching v.19 no4 p.317-38.

Kulhavy, R.W. and Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In Dempsey, J. and Sales, G. (Edited). Interactive Instruction and Feedback. Englewood Cliffs: Educational Technology Publications.

Nicaise, M., (1997). Computer-supported apprenticeships in math and science. The Journal of Computers in Mathematics and Science teaching v.16 no4 p.443-65.

Paolucci, R. (1998). The effects of cognitive style and knowledge structure on performance using hypermedia-learning system. Journal of Educational Multimedia and Hypermedia v. 7 no2-3 p. 123-50.

Russell, T. (1997). Teaching teachers: How I teach IS the message. In J. Loughran & T. Russell (Eds.). Teaching about teaching: Purpose, passion and pedagogy in teacher education. London: Falmer.

Sasser, J. E. (1990-1991). The effect of using computer tutorials as homework assignments on the mathematics achievement of elementary education majors. Journal of Computers in Mathematics and Science Teaching, 10(2), 95-102.

Schank, R.C. (1993). Learning via multimedia computers. Communications of the ACM, 36(1), 54-56.

Schifter, D. (1997). Learning mathematics for teaching: Lessons in/from the domain of fractions. Newton, MA: Education Development Center, Inc. (ERIC Document Reproduction Service No. ED 412 122).

Scholz, J. M. (1995, April). Professional development for mid-level mathematics. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 395 820).

Shcoenfeld, A. (1992) Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D.A. Growa (Ed.), Handbook of Research on mathematics teaching and learning (pp. 334-370). New York: Macmillan.

Shute, V. J. (1993). A comparison of learning environments. In S. P. Lajoie & J. Derry (Eds.), Computers as Cognitive tools (pp. 46-64). Hillsdale, NJ: Erlbaum.

Snider, R. C. (1992). The machine in the classroom. Phi Delta Kappan, 74 (4), 316 – 323.

Torrez, N. (Feb. 2000). Developing Culturally Consonant Curriculum Using the Technology of the New Millennium. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education (52nd, Chicago, IL, February 26 – 29, 2000).

Verzoni, K. A. (Oct. 1997) Turning students into problem solvers. Mathematics Teaching in the Middle School v.3 p.102-7.