Effects of an Online Instructional Application on Reading And Mathematics Standardized Test Scores

Trey Martindale
University of Memphis

Carolyn Pearson
University of Arkansas at Little Rock

L. K. Curda
University of West Florida

Janet Pilcher
University of West Florida

Abstract

Standardized tests have become commonly used tools for accountability in public education in the United States. In Florida, the Florida Comprehensive Assessment Test (FCAT) is used to measure student achievement on grade-specific standards and benchmarks. Various agencies have developed computer-based and web-based software applications to improve student performance on these tests. The purpose of this study was to examine the impact of one such application, FCAT Explorer, on student FCAT scores. We used hierarchical analysis of variance and analysis of covariance to compare scores for schools that used FCAT Explorer, and schools that did not. We examined fourth, fifth, eighth, and tenth grade FCAT reading and mathematics scores for selected elementary schools and high schools. Student scores from elementary schools using FCAT Explorer were significantly higher than scores from elementary schools that did not use FCAT Explorer. At the high school level, we found no significant differences in scores between schools that used FCAT Explorer and schools that did not use the application. (Keywords: high-stakes standardized testing, accountability, computer-based instruction, online instruction, software evaluation.)

INTRODUCTION

One critical and controversial issue in American public education today is the use of high-stakes testing as a tool for accountability in schools. Although previous research has uncovered both positive and negative consequences of high-stakes testing (McColskey, 2000) and controversy over these tests continues (AERA, 2000), nearly all states are operating standardized testing programs. Many of these programs are attached to systems of rewards and punishment (Amrein & Berliner, 2002; O'Neil, 1992).

In the early 1970s, the Florida Commission on Education Reform and Accountability was formed. This commission recommended procedures for assessing student learning, with the goal of raising educational expectations. This goal was set in place to address the demand for a well-educated workforce in a state experiencing rapid growth in population and in commerce. The State Board of

Education adopted the commission's recommendations, and the Florida Legislature mandated statewide assessment of students in Grades 3, 5, 8, and 11 (see Florida Department of Education, 2005).

During the 1990s the recommendations became known as the Comprehensive Assessment Design and were expanded so that students would be tested in reading, writing, mathematics, and creative and critical thinking. The Commission also requested that educational content standards be developed and adopted. This prompted the Florida State Board of Education to develop the Sunshine State Standards, which codified what students should know and be able to do at each grade level. These standards were subdivided into what were called *benchmarks*. The Florida Comprehensive Assessment Test (FCAT) was designed to meet the requirements of the Comprehensive Assessment Design, and was aligned with the Sunshine State Standards. The FCAT has two basic components: a criterion-referenced test, which measures reading, writing, science, and mathematics; and a norm-referenced test, which measures students' performance against national norms.

In 1999, the Florida Legislature mandated that schools be assigned an annual performance grade, ranging from a high of A (making excellent progress) to a low of F (failing to make adequate progress). As of 2005, each school's performance grade is based on student FCAT scores and other factors including attendance, dropout rate, school discipline data, cohort graduation rate, and student readiness for college. Schools that receive either a grade of A or improve by at least two letter grade categories are rewarded with greater autonomy, including authority over the school's budget (Florida Statutes Ch. 99.398). Such schools are also eligible for the Florida School Recognition Program—a program of financial awards that are disbursed at the discretion of the school's staff and an advisory council.

Schools designated as performance grade category D or F are eligible to receive assistance and intervention toward improving performance (Florida Statutes Ch. 99.398). However, the state has been given authority to take action if a particular school does not improve. Students in any school that receives a grade of F for two consecutive years are eligible for a state voucher (opportunity grant), which allows the students to attend a higher-performing school in the same district, a higher-performing school in an adjoining district, or a private school. This program has been noted as one of the most aggressive test-based accountability measures in the nation.

COMPUTER-BASED AND ONLINE INSTRUCTION

One reaction to state high-stakes standardized testing has been to develop computer-based and Web-based software applications that prepare students for tests such as the FCAT. When used appropriately, computers, educational software, and Web resources can contribute in a variety of ways to effective learning environments (Herrington & Oliver, 1999; Martindale, Cates, & Qian, 2005; Snider, 1992). Within the classroom, instructional software use ranges from drill-and-practice for remediation to entire curricula and instructional processes. The development of multimedia and Web-based instruction may provide an opportunity to communicate with wide and diverse audiences, including students, teachers, administrators, and parents. A number of design and research efforts have been undertaken to develop computer- and Web-based instructional and practice materials focused on high-stakes testing (McDonald & Hannafin, 2003; Wright, Barron, & Kromrey, 1999). For an example, go to http://devbox.mediavue.net/fcat3/.

Well-designed instructional software can provide learning opportunities for students both at school and at home. Students who use appropriate tutorials, especially when completing homework assignments, have higher achievement than those who use traditional methods (Sasser, 1991). McDonald and Hannafin (2003) found that use of Web-based computer games designed for high-stakes test preparation promoted higher-order learning outcomes. These outcomes included increased meaningful dialogue among students and the identification of student misconceptions. Although these outcomes contributed to deeper understanding, no significant differences were found on test scores between those students who used the computer games and those who did not. Still, considerable research supports the hypothesis that online learning environments have a positive effect on learning outcomes (Goldenberg & Cuoco, 1996; Russell, 1997; Sanders, 2001; Schifter, 1997); can accommodate a variety of learning styles (Hawkins, 1993; Schank, 1993); can support higher-order learning, (Paolucci, 1998; Schank, 1993), especially in mathematics (Nicaise, 1997); and can teach problem-solving skills to those who struggle with learning difficulties (Babbitt & Miller, 1996).

Much research in cognition has concentrated on learner traits and learner control in online environments, and factors such as instructor style and amount of instruction available (Freitag & Sullivan, 1995; Hannafin & Scott, 1998; Hannafin & Sullivan, 1996). Perhaps one of the most important features of emerging technology is the capability for interactivity and opportunity for feedback. There is much research on the varied amounts of support (Hannafin & Scott, 1998), interactivity, feedback, pacing, and individualization (Hawkins, 1993), which may significantly improve achievement (Naime-Diefenbach & Sullivan, 2001). Feedback is particularly important for enhancing achievement, especially in terms of immediacy, amount of information provided, and the type of task involved (Khine, 1996; Kulhavy & Wager, 1993). Feedback and interactivity also influence learner motivation in online environments (Bolliger & Martindale, 2004; Hawkes & Dennis, 2003).

The potential benefits of learning technologies do not guarantee they will be used. Computer-based study materials to prepare students for tests like the FCAT can be prohibitively expensive, and difficult to manage (Fahy, 2000). Teachers may not have the time or expertise to determine which software or web-based application is appropriate and compatible, and wrong choices are lamentable in a climate of strained educational budgets (Kim & Sharp, 2000).

FCAT EXPLORER

FCAT Explorer (http://www.fcatexplorer.com), produced by Infinity Software, Inc. and provided by the Florida Department of Education to Florida public schools at no charge to the schools, was designed as an FCAT practice resource that is interactive, benchmark based, and accessible from any Internet-connected computer. Infinity won a competitive bid to produce this software to the Florida Department of Education's specifications. FCAT Explorer includes educator resources and a parent and family guide for third, fourth, sixth, eighth, and tenth grade reading and fifth, eighth, and tenth grade mathematics. FCAT Explorer was designed to align with the *Sunshine State Standards*, and provides practice items in a variety of formats that are linked to and reinforce each of the grade-level benchmarks. The application was developed in cooperation with a variety of expert teachers and testing specialists, with a foundation in the principles of instructional design and cognitive learning theory (e.g., Dick, Carey, & Carey, 2004). The design of FCAT Explorer reflects consideration of learner motivation by incorporating elements of Keller's ARCS model (Keller, 1987; Naime-Diefenbach, 1991).

PURPOSE OF THE STUDY

Prior examinations of FCAT Explorer and its impact on FCAT standardized test scores have been limited. One prior study examined the effects of FCAT Explorer on fifth, eighth, and tenth grade 2002 mathematics FCAT scores and found significant differences for all grade levels when comparing students who did and did not use the program (Sullivan & Naime-Diefenbach, 2002). Some design constraints were present, however: although data were aggregated at the school level, no control group was used to rule out maturation and other threats to internal validity, and only mathematics scores were examined. Because students' instruction through educational software should positively affect comprehension of standard curricula, this should be reflected in FCAT scores or other outcome measures (Naime-Diefenbach & Sullivan, 2001). The purpose of the current study was to determine if students who used FCAT Explorer scored higher on the FCAT reading and mathematics tests than those who did not use the software. We analyzed use of FCAT Explorer on 2001 and 2002 FCAT scores using a quasi-experimental design described in the following section.

METHODS

Sample

Twenty-four schools were identified for participation in the study. The experimental group was composed of twelve schools—three schools each at four grade levels (Grade 4 for reading and 5, 8, and 10 for mathematics). Infinity Software, Inc. provided the researchers with usage-level data for all Florida schools who used FCAT Explorer. The 12 schools with the highest program use in the respective grades were then selected for this research and 12 schools in the same district or very nearby that did not use the program at all became the control group—three schools each at the same four grade levels. Each experimental school was selected based on its high percentage of student use of FCAT Explorer. A control school was matched to each experimental school by virtue of similar characteristics (e.g., same district, school size, and performance grade assigned by the state). Based on data from Infinity Software Inc., only students who had used FCAT Explorer were included in the experimental school data set. The total sample generated for the four grade levels was as follows: fourth (n = 586), fifth (n = 491), eighth (n = 1,379), and tenth (n = 1,505).

Procedure and Data Analysis

In both the experimental and the control group, three schools were nested under each condition (grade level). FCAT reading and mathematics scores were obtained from the Florida Department of Education database for the school years ending in 2001 and 2002. Because the focus of FCAT Explorer was on fourth grade reading and fifth, eighth, and tenth grade mathematics, these corresponding FCAT scores were used. Infinity Software provided data on student usage of FCAT Explorer (i.e., used versus not used) and number of items completed. These student data were matched to the FDOE database of student FCAT scores. This ensured that all students in the experimental group used FCAT Explorer, and all students in the control group did not use the application. Only students who had complete data for both school years were included in the analyses.

For each grade level, a hierarchical analysis of variance was conducted using FCAT scores as a repeated measure from 2001 to 2002 to control for maturation and initial group differences. Follow-up analyses (ANCOVA) ignored the grouping variables to more closely examine the strength of the treatment by examining the scores strictly by program usage or non-usage. In determining the strength of the treatment, all effect sizes were determined using Cohen's f.

RESULTS

For fourth grade reading, the main effect of the treatment (Explorer use versus Explorer non-use) was statistically significant ($F = 10.35$, $p < .01$, $f = .13$), and there was a statistically significant difference between the individual schools within the treatment ($F = 9.68$, $p < .01$, $f = .28$). Examination of the time effect (2001 to 2002 FCAT scores), the "Explorer use by time" interaction, and the "school within Explorer use by time" interaction revealed no statistically significant differences on reading scores. Table 1 reveals that regardless of school year, for fourth grade, the Explorer-use adjusted means were higher than the non-use means, although the effect sizes reported were moderate at best. Examination of the adjusted means over time, particularly within the Explorer-use, gives insight into the non-significant time effect, as the gain was minimal.

For fifth grade mathematics, the main effect of the treatment (use versus non-use) was statistically significant ($F = 4.46$, $p < .04$, $f = .09$), and there was a statistically significant difference between the individual schools within the treatment ($F = 20.60$, $p < .01$, $f = .46$). The time effect was also statistically significant ($F = 175.23$, $p < .01$, $f = .60$), as was the "school within Explorer use by time" interaction ($F = 13.11$, $p < .01$, $f = .37$), but not the "Explorer use by time" interaction. For fifth grade, again the Explorer-use adjusted means were higher than the non-use means with a moderate effect size at best. Both the Explorer-users and non-users made statistically significant gains in FCAT math scores, as indicated by the strong effect size for time.

For eighth grade mathematics, the main effect of the treatment (use versus non-use) was not statistically significant; however, there was a statistically significant difference between the individual schools within the treatment ($F = 35.88, p < .01, f = .36$). The examination of the time effect, the "Explorer use by time" interaction, and the "school within Explorer use by time" interaction revealed statistically significant differences across all "within" analyses ($F = 176.62, p < .01, f = .36$; $F = 4.31, p < .04, f = .06$; $F = 7.21, p < .01, f = .16$; respectively). For eighth grade, both Explorer-users and non-users made statistically significant gains in FCAT math scores, as indicated by the moderate effect size for time. But interestingly, after adjusting for initial group differences, the Explorer-users scored lower than the non-users in 2002.

For tenth grade mathematics, the main effect of the treatment (use versus non-use) was not statistically significant; however, there was a statistically significant difference between the individual schools within the Explorer groups ($F = 4.85, p < .01, f = .13$). The examination of the time effect was statistically

Table 1. Means, Adjusted Means, and Standard Deviations of Fourth, Fifth, Eighth, and Tenth Grade FCAT Reading and Mathematics Score By Explorer and School within Explorer

		Time					
		FCAT 01			FCAT 02		
Grade	*n*	*Mean*	*SD*	*Adjusted Mean*	*Mean*	*SD*	*Adjusted Mean*
Fourth Grade							
Explorer Use							
School 1	3	329.33	9.23		306.00	13.08	
School 2	115	326.41	46.49		342.47	38.49	
School 3	164	297.60	48.79		308.80	45.67	
Total	282	309.69	49.62	317.79	322.50	45.69	319.09
Explorer Non-Use							
School 1	14	258.07	66.04		263.71	74.26	
School 2	142	293.62	63.10		309.82	50.78	
School 3	148	280.95	59.42		290.79	66.93	
Total	304	285.81	61.88	277.55	298.43	61.27	288.11
Fifth Grade							
Explorer Use							
School 1	124	258.65	65.15		314.02	48.44	
School 2	160	304.78	51.05		329.18	39.99	
School 3	153	290.01	57.22		330.14	51.74	
Total	437	286.52	60.31	284.48	325.21	47.20	324.44
Explorer Non-Use							
School 1	22	203.23	77.36		239.45	79.80	
School 2	26	305.19	55.75		350.96	42.97	
School 3	6	284.50	60.72		332.33	33.77	
Total	54	261.35	81.19	264.31	303.46	80.02	307.58

significant (F = 271.10, p < .01, f = .42), but not the "Explorer use by time" interaction and the "school within Explorer use by time" interaction. For tenth grade, both the Explorer-users and non-users made statistically significant gains in FCAT math scores, as indicated by the moderate effect size for time. After adjusting for initial group differences, the Explorer-users did not score significantly higher than the non-users.

We conducted additional follow-up analyses of the FCAT scores using analysis of covariance, collapsing across the schools and groups to more closely examine the strength of the treatment (use versus non-use of FCAT Explorer). Using the 2001 FCAT scores as the covariate, significant group differences were found in the fourth grade reading scores (F = 4.77, p < .03, f = .07), and the Explorer-users scored five points higher than the non-users, although the effect size was weak (Table 2). For the Explorer-users, the mean number of items attempted was 25 with a range of 1 to 201. For fifth grade, significant group differences were found in the mathematics scores (F = 12.76, p < .01, f = .11), with the Explorer-users scoring 12 points higher than the non-users. For the Explorer-users, the mean number of items attempted was 162 with a range of 1

Table 1, continued

		Time					
		FCAT 01			FCAT 02		
Grade	*n*	*Mean*	*SD*	*Adjusted Mean*	*Mean*	*SD*	*Adjusted Mean*
Eighth Grade							
Explorer Use							
School 1	195	307.47	60.06		317.34	50.09	
School 2	262	316.47	59.91		328.39	42.05	
School 3	144	310.86	49.26		319.15	42.89	
Total	601	312.21	57.64	311.60	322.59	45.22	321.63
Explorer Non-Use							
School 1	173	295.24	55.71		308.95	48.93	
School 2	278	296.50	56.77		317.06	49.36	
School 3	327	342.43	50.44		349.36	39.56	
Total	778	315.53	58.56	311.39	328.83	48.68	325.13
Tenth Grade							
Explorer Use							
School 1	22	324.73	30.68		348.23	25.31	
School 2	180	306.87	58.99		324.64	46.47	
School 3	329	311.94	51.20		333.14	36.84	
Total	531	310.75	53.38	314.51	330.89	40.29	335.34
Explorer Non-Use							
School 1	94	311.95	47.68		333.32	37.30	
School 2	442	306.20	51.33		324.71	39.92	
School 3	438	317.71	50.03		334.53	34.82	
Total	974	311.93	50.66	311.95	329.96	37.72	330.86

Table 2. Means and Standard Deviations of Fourth, Fifth, Eighth, and Tenth Grade FCAT Reading and Mathematics Scores by Usage

	Usage of FCAT Explorer							
	Used				Not Used			
Grade	*n*	*Mean*	*SD*	*Adjusted Mean*	*n*	*Mean*	*SD*	*Adjusted Mean*
Fourth Grade								
Reading 01	348	306.14	48.89		507	289.82	63.46	
Reading 02	381	314.96	48.74	309.72	542	296.26	63.55	304.17
Fifth Grade								
Math 01	786	284.99	57.46		90	259.50	76.20	
Math 02	861	323.13	47.16	323.33	109	291.61	77.40	311.05
Eighth Grade								
Math 01	675	312.21	56.72		1194	308.36	65.65	
Math 02	724	322.23	46.31	321.91	1370	318.98	55.76	322.51
Tenth Grade								
Math 01	641	306.23	53.20		1626	307.90	53.49	
Math 02	683	326.82	39.93	329.42	1816	324.42	43.07	327.92

to 987. For the eighth graders, significant group differences were not found in the mathematics scores, as evidenced by the negligible difference in means. For the Explorer-users, the mean number of items attempted was 59 with a range of 1 to 393. For tenth graders, significant group differences were not found in the mathematics scores and there was only a two-point difference in the means. For the Explorer-users, the mean number of items attempted was 27 with a range of 2 to 345.

DISCUSSION

Examination of the findings related to fourth grade reading and fifth grade mathematics revealed that students who used FCAT Explorer had significantly higher FCAT scores compared to students who did not use the program when controlling for previous scores and individual differences between schools. Additionally, when students were examined solely on usage or non-usage, FCAT Explorer users had significantly higher reading scores (fourth grade) than non-users, and significantly higher mathematics scores (fifth grade) than non-users. Although there were significant differences between those who used FCAT Explorer and those who did not, the treatment effect was still somewhat weak based on the effect sizes observed.

For eighth and tenth grade mathematics, the findings were different than that for fourth and fifth grade. We found no significant differences between the Explorer-users and non-users when controlling for previous scores and individual difference between schools. We did find a time effect, in that both users and non-users recorded increased FCAT scores from 2001 to 2002. The effect sizes (of the time effect) were fairly large (f = .36 and .42 for eighth and tenth, respectively). Counter-intuitively, the FCAT score increase was larger (but not of statistical or practical significance) for the eighth grade non-users than for Explorer users.

When examining the overall picture provided by these different analyses, evidence exists that the FCAT Explorer program is effective in the elementary grades and more effective for elementary rather than secondary students. It is important to note that effect sizes ranged from negligible to weak at best when examining usage of the program. The effect sizes were stronger when controlling for prior scores and individual differences between schools, particularly when examining time effects.

The low strength of the treatment is a limitation of this study, and we recommend further investigation of FCAT Explorer if it does become more fully integrated as a learning tool in the school districts across Florida. Mean usage levels of the application by students varied widely between elementary and secondary grade levels. There were also very wide ranges of usage levels within grades, especially for fifth grade, where the number of items attempted ranged from 1 to 987. The number of items attempted could explain part of the group differences that were found. A few students attempting a large number of FCAT Explorer items could explain a portion of the effects observed at the lower grade levels. What is not measured in this study is the quality of instructional time for students using FCAT Explorer, and all that is known about individuals is the range of frequency of "logging on" to the program. Concentrated time on task with encouragement from the teacher is undoubtedly important, along with the quantity of practice items a student completed within the application.

At the upper grade levels, one possible explanation for the lack of effects may be that high school teachers may not perceive the need for the program or have the time to implement it. It may be that elementary teachers perceive more pressure to prepare students to face the multiple rounds of FCAT testing in the years to come, and will use all available resources for test preparation. Conversely, high school teachers may perceive that program use at such late grades will have little effect on FCAT performance for students who already have a documented pattern of certain performances on standardized tests. Future research should examine teachers' use of the application, teacher perceived value of FCAT Explorer, and instructional time allotted for students to use the application during the regular school day.

CONCLUSION

The FCAT Explorer results presented here are promising, particularly for the elementary grades. The program appears to be usable by both teachers and students, and is aligned to the Sunshine State Standards. The cost of purchasing and supporting applications such as FCAT Explorer is a critical issue, and the state of Florida has much invested already in this particular program. The state should continue to critically evaluate such programs, and strongly encourage vendors to employ interoperability and open standards for the sharing of program services and data with other applications.

Contributors

Dr. Trey Martindale is an assistant professor of instructional design and technology at the University of Memphis. (Address: Dr. Trey Martindale, Department of Instructional Design & Technology, 406 Ball Hall, University of Memphis, Memphis TN 38152; treymartindale@gmail.com.)

Dr. Carolyn Pearson is a professor of educational foundations at the University of Arkansas at Little Rock. (Address: Dr Carolyn Pearson, Department of Educational Leadership, University of Arkansas at Little Rock, Little Rock, AR, 72204; lcpearson@ualr.edu.)

Dr. L. K. Curda is an associate professor of instructional technology at the University of West Florida. (Address: Dr. L. K. Curda, Division of Graduate Education, University of West Florida, Pensacola, FL, 32514; lcurda@uwf.edu.)

Dr. Janet Pilcher is a professor and Dean of the College of Professional Studies at the University of West Florida. (Address: Dr. Janet Pilcher, Dean, College of Professional Studies, University of West Florida, Pensacola, FL, 32514; jpilcher@uwf.edu.)

References

AERA position statement concerning high-stakes testing in pre-K–12 education. (2000, July). Washington, DC: American Educational Research Association. Retrieved February 28, 2005, from http://www.aera.net/policyandprograms/?id=378.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, *10*(18). Retrieved February 24, 2005, from http://epaa.asu.edu/epaa/v10n18/.

Babbitt, B. C., & Miller, S. P. (1996). Using hypermedia to improve the mathematics problem-solving skills of students with learning disabilities. *Journal of Learning Disabilities*, *29*, 391–401.

Bolliger, D. U., & Martindale, T. (2004). Key factors in determining student satisfaction in online courses. *International Journal on E-Learning, 31*(1), 61–67.

Dick, W., Carey, L., & Carey, J. O. (2004). *The systematic design of instruction* (s[ixth] ed.) New York: Longman.

Fahy, P. J. (2000). Achieving quality with online teaching technologies. Paper presented at Quality Learning 2000 Inaugural International Symposium, Calgary, AB. (ERIC Document Reproduction Service No. ED445197)

Florida Department of Education. (2005). Ensuring highest student performance, and leading the nation in raising student achievement. Retrieved March 9, 2005, from http://www.fldoe.org/k12/.

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.

Goldenberg, 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* (pp. 381–420). Hillsdale, NJ: Erlbaum.

Hannafin, R. D., & Scott, B. N. (1998). Identifying critical learner traits in a dynamic computer-based geometry program. *The Journal of Educational Research*, *92*(1), 3–12.

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

Hawkes, M., & Dennis, T. (2003). Supporting and assessing online interactions in higher education. *Educational Technology, 43*(4), 52–56.

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*, *10*(1), 3–24.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development, 10*(3), 2–10.

Khine, M. S. (1996). The interaction of cognitive styles with varying levels of feedback in multimedia presentation. *International Journal of Instructional Media, 23*(3), 229–237.

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*, *19*(4), 317–338.

Kulhavy, R. W., & Wager, W. (1993). Feedback in programmed instruction: Historical context and implications for practice. In J. Dempsey & G. Sales (Eds.). *Interactive Instruction and Feedback* (pp. 3–20). Englewood Cliffs, NJ: Educational Technology Publications.

Martindale, T., Cates, W. M., & Qian, Y. (2005). Analysis of recognized web-based educational resources. *Computers in the Schools, 21*(3/4), 110–117.

McColskey, W. H., & McMunn, N. (2000, October). Strategies for dealing with high-stakes state tests. *Phi Delta Kappan*, *82*(2), 115–120.

McDonald, K. K., & Hannafin, R. D. (2003). Using web-based computer games to meet the demands of today's high-stakes testing: A mixed method inquiry. *Journal of Research on Technology in Education, 35*(4), 459–472.

Naime-Diefenbach, B. (1991). *Validation of attention and confidence as independent components of the ARCS motivational model*. Unpublished doctoral dissertation. Florida State University, Tallahassee.

Naime-Diefenbach, B., & Sullivan, C. B. (2001). *Measuring the effects of a web-based practice program on the FCAT math results of fifth graders*. Paper presented at the Florida Educational Research Association Annual Conference, Marco Island, FL.

Nicaise, M., (1997). Computer-supported apprenticeships in math and science. *The Journal of Computers in Mathematics and Science Teaching*, *16*(4), 443–465.

O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, *49*(8), 14–19.

Paolucci, R. (1998). The effects of cognitive style and knowledge structure on performance using hypermedia-learning system. *Journal of Educational Multimedia and Hypermedia*, *7*(2/3), 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* (pp. 32–47). London: Falmer.

Sanders, W. B. (2001). *Creating learning-centered courses for the world wide web*. Boston & London: Allyn and Bacon.

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. ED412122)

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

Sullivan, C. B., & Naime-Diefenbach, B. (2002, November). *Measuring the effects of a web-based practice program on the FCAT math results*. Paper presented at the annual meeting of the Florida Educational Research Association, Gainesville, FL.

Wright, V., Barron, A. E., & Kromrey, J. D. (1999). Preparing students for high-stakes testing in Florida: An interactive website. *Florida Journal of Educational Research, 39*(1), 79–94.