An Evaluation of the Merit Reading Software Program in the Calhoun County (WV) Middle/High School

Jerry D. Jones
William D. Staats
Noel Bowling
Robert D. Bickel
Michael L. Cunningham
Connie Cadle
Marshall University

Abstract

We were asked by Merit Software to conduct a quasi-experimental research study to evaluate the effects of its reading software on middle school students. Because the No Child Left Behind Act emphasizes the importance of evidence-based interventions and has set improving students reading comprehension as a goal, we agreed to take on this project. Scores from the Stanford Achievement Test, Ninth Edition (SAT-9) for both the treatment and control groups were compared for 2002 and 2003. Our findings showed achievement gains associated with each of the nine SAT-9 dependent variables we measured. Keywords: scientifically based reading, scientifically based research, scientific based research, reading research, research based evidence, research based reading, computer assisted instruction, computer based reading.

Merit Software, a publisher of educational software since the early 1980s, commissioned us to evaluate the effects of Merit reading software on students in grades six and eight during the winter of 2003.

We contacted administrators at the Calhoun County Middle/High School in Mount Zion, West Virginia to ask if they would participate in a control versus treatment group study. After reviewing the Merit programs, school officials agreed that this was a worthwhile endeavor. We coordinated the project with school administrators, trained the participating teachers in the use of the software, collected data, documented the project activities, and conducted data analysis.

PERSPECTIVES AND LITERATURE REVIEW

Although an argument can be made for including technology in schooling for its own sake (i.e., to prepare students for the technology-saturated environments they will face as adults), many policymakers and community members want evidence that technology supports student learning as they make decisions about technology investments. In the wake of the No Child Left Behind Act of 2002, with its emphasis on technology, computer software publishers have hastened to update and design skill intervention programs that have great appeal for district and school administrators and educators ("NCLB Dominates at FETC," 2003).

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Computer Assisted Instruction and Reading Improvement

In an experimental study on the effects of interactive software on children's reading comprehension, researchers reported that computer assisted instruction (CAI) designed to advance the reading skills of underachieving students effected gains, particularly with long and difficult passages. Reading comprehension increased more as students read longer and more difficult narratives from a CD-ROM displayed on a computer screen than when they read the same material in shorter and easier passages from the printed page (Greenlee-Moore & Smith, 1996). Greenlee-Moore and Smith used a three-way analysis of variance of comprehension scores on the Iowa Test of Basic Skills, revealing that reading from computers increased comprehension when subjects were reading longer and more difficult narratives.

A basic skills program in 18 West Virginia elementary schools produced powerful gains in the SAT-9 achievement test scores of fifth graders (Mann, Shakeshaft, Becker, & Kottkamp, 1999). This experimental project was one of the most comprehensive in terms of students covered (n=950) in West Virginia, which already had statewide educational programs at the time. Mann and colleagues tested the extent to which achievement gains were attributable to technology intervention. Testing showed an overall 11% gain on the SAT-9 from one year to the next (1996–97 to 1997–98) with an adjusted r_ of .094.

In its definitive 2000 report, the National Reading Panel noted a dearth of literature on the relationship between reading achievement and computer-based instruction (NIH, 2000). In particular, studies that separate students into control and treatment groups and then analyze the data at the student level are notably lacking. Four studies published since then corroborate the findings of this study, which contributes significantly to the research on this promising support tool for teachers' instruction.

A 2001 college pilot mentoring/tutoring project aimed at improving academic achievement in homeless, at-risk, urban elementary school students used computer software tutoring sessions and saw improvement in student reading achievement (Walport & Fitzpatrick, 2001). In this study, pre-assessment results for elementary school students at the beginning of the project were compared to those of their peers. Final assessment showed that the tutees made between three- and eight-month gains toward grade equivalency as indicated by the Slosson Oral Reading Test-R and the Star Assessments for Reading and Math.

Computer-based programs also appear to improve other academic areas that depend on reading skills. A study of effects of supplementing traditional instruction with a computer-based program on the Battle of Philippi for West Virginia History students yielded interesting results. Students were divided into three groups: one group received traditional instruction, one group was placed in a self-paced computer module, and one group received traditional instruction supplemented with the computer-based program. The students receiving the hybrid treatment showed the greatest increase in test performance, and the self-paced computer-instructed group showed the next-greatest increase (Hart, 2002).

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Ligas (2002) examined a five-year project focused on at-risk elementary and middle school students in Broward County, Florida. The study particularly focused on the project's effect on reading performance. Strong emphasis was placed on Direct Instruction, Accelerated Reader, and the use of CAI. The evaluation used a time series design to measure the project's impact on students' achievement over five years. The investigation compared a group of students who used software for 12 or more hours with students at the same school who did not use the software or used it for less than five hours. The first group outperformed the control group by 7.74 points on the SAT-8 Reading Comprehension average NCE scores.

A study by Traynor (2003) was designed to determine how CAI improved student performance among various types of students at Bloomington Middle School in California. Of 210 students in a particular language arts course, 161 were chosen to be the sample group, and completed a pretest and posttest for the capitalization subject area. This subject was chosen because it contained the largest number of pretest scores. The students were categorized into four groups for instructional purposes: special education, sheltered English immersion (non-English proficient), traditional English immersion (limited English proficient), and regular education. Data collected included pretest and posttest scores over approximately 70 days. A dependent t-test showed significant difference between pretest and posttest scores for the entire sample. Overall, treatment group students' posttest scores were significantly higher than those who did not use the computer-based program.

Recent research using the Peabody Literacy Lab suggests that integrated media (video, software, books-on-tape, low-level books) is an effective instructional tool for older struggling readers. The Orange County Public School District in Florida used the Lab as an intervention (presentation of new content/skills, guided student practice, feedback and corrections, independent student practice, and ongoing review) for 63 of their most disabled sixth through eighth grade readers from three schools. Using "anchored instruction," an animated tutor guides the student through exercises and delivers feedback through a digitized human voice. Using the Stanford Diagnostic Reading Test as pretest and posttest measures, significant interactions were found between the type of instruction and time of test in Auditory Vocabulary, Literal Comprehension, Inferential Comprehension, and Total Comprehension. Work at the Lab was the basis for Scholastic's Read 180 intervention program (Hasselbring & Goin, 2004).

BACKGROUND AND PURPOSE

Merit Software's comprehensive skills intervention program addresses public school grades 3–12. This program was designed to satisfy state requirements as measured on standardized instruments. The individual instructional modules cover skills necessary for academic achievement in reading (vocabulary and comprehension), grammar, spelling, math, problem solving, writing, and critical thinking.

Merit Software's programs are based on the theories of Skinner, Bloom, and the more contemporary constructivists (Anderson & Krathwohl, 2001). The

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

premise behind the Merit programs is that in addition to the basics, students should learn to connect concepts, solve unfamiliar problems, communicate ideas, and apply facts and skills that they have learned. This is accomplished through self-paced learning, encouragement to continue, understandable language and interesting content, opportunities to see things in different ways, clearly-delineated steps for dealing with material, straightforward instructions, and introductions to material.

Merit programs include pre-assessment, comprehension guides, comprehension practice, post-assessment, record keeping of individual student responses, and time-on-task indicators. The software was designed to give teachers quick and essential feedback to allow greater responsiveness to individual students' instructional needs and accomplishments.

Based on the capabilities of Merit reading software, the need for evidence-based research, and the concerns of the Calhoun County school district, these questions guided our study:

1. Does the software complement everyday classroom teaching?
2. Did the software affect students' achievement as measured by standardized tests?
3. What does use of this software intervention strategy imply for educational reform?

PROJECT SETTING

Calhoun County, WV was selected as the study site primarily because the school district has a strong history of adopting innovative ideas to improve instruction. It also has a stable, albeit impoverished, population. The 2000 US Census Report registers approximately 37% of children under 18 years of age in Calhoun County living below the poverty line and 28% of persons over age five years living with a disability. Calhoun County suffers the state's highest unemployment rate, 23.1%, according to a June 2003 report by *The Parkersburg News*. Despite this, the school district administrators demonstrated a strong interest and initiative in adopting innovative ideas, including technology, to improve instruction.

The 165,000-square-foot Calhoun Middle/High School is relatively new, built in 1998. It has a fully networked lab of Windows computers. Currently, it serves students and staff in grades 5–12 and has a total enrollment of 821. Students enjoy relatively small class size, which averages 18.8 students.

Participants

The summer before the start of 2002–2003 school year, the Calhoun County Middle/High School principal used the school board's computer-scheduling program to place sixth- and eighth-grade students into eight heterogeneous classrooms (four per grade) based on classroom assignment, grade level, and needed skills. The computer-scheduling program also assigned teachers to these classrooms based on grade and subject (reading/language arts) they taught. Before work with the software began, the sixth and eighth grade teachers chose

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

which classes would be the experimental group and receive the Merit reading software treatment, and which classes would be the control group, receiving regular instruction.

Curriculum

Calhoun County Middle/High School's language arts curriculum for grades six and eight is aligned to the West Virginia Instructional Goals and Objectives (WV IGOs) for grades six and eight. Teachers assign lessons using materials such as worksheets, textbooks, novels, and short stories. Teachers also employ other activities that do not use outside materials. These include journal writing, silent reading, and class discussions of readings.

Grade six students used *Elements of Literature* (Introductory Course) as their literature textbook and *The Writer's Craft* for grammar. Grade eight students used *Elements of Literature* (Second Course) for literature and *Elements of Language* (Second Course) for grammar. By matching the behaviors required by the Merit objectives and WV IGOs, we determined that the Merit objectives matched 75% of the WV IGOs for Grades Six and Eight Language Arts, against which the Calhoun County's current curriculum is aligned. The participating teachers selected and used the Merit units that were aligned with the WV IGOs in effect at the time of the study.

Intervention

Students participating in this research project received intervention time on computers for two 45-minute sessions per week for four weeks, covering an average of 360 minutes. The tested modules addressed

1. Accu-Reading (Units 1 and 2)
2. Vocabulary Fitness (Units 1 and 2)
3. Grammar Fitness (Units 1 and 2)

Some eighth graders also received instruction from the Developing Critical Thinking module (Unit 1). Any extra sessions were devoted to skills the student and teacher agreed to work on. The program includes record keeping, which enables the teacher to account for student progress. The modules consist of all or a combination of four exercises:

1. Tryout, a pretest that can provide the teacher and student with information on relative skill strength.
2. Warm-up, which isolates a skill and provides several opportunities to practice it. The software provides feedback, and to finish, students need to respond correctly to set criteria.
3. Workout, a more rigorous exercise that combines these skills. Again, the soft ware offers feedback, and again the students need to respond correctly to set criteria to finish the exercise.
4. Finals, the posttest. The program includes record keeping that helps teachers monitor students' progress.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Procedures

The experimental group consisted of 116 students in three sixth grade and three eighth grade reading/language arts classes, and the control group consisted of 35 students in one sixth grade and one eighth grade class. All sixth and eighth graders participated in the study.

The same teacher taught each of the three sixth grade treatment classes. Another teacher taught all of the eighth grade treatment classes. Two Title I reading specialists also worked with many students in both the treatment and control groups. Separate instructors led the sixth and eighth grade control classes. The two treatment group teachers were trained to work with the software, whereas the two control group teachers and the two Title I reading specialists did not receive such training.

Merit Software trained a consultant to conduct a total of two days of training for the two teachers. The trainer guided the teachers through the software and familiarized them with the skills content areas. Training consisted of a review of the research background for the software's design, an explanation of the software design, a demonstration orientation, and useful tips for integrating the software content into traditional classroom content areas.

Calhoun Middle/High School students receive more than 90 minutes of language arts instruction each week. The Merit modules were used as part of the treatment group's curriculum. When the Merit software was not part of the class lesson, treatment group students received some of the same regular instruction as control students. Merit modules primarily replaced worksheets, textbooks, and some class discussions.

During the study, all sixth graders—control and treatment groups—read *Tuck Everlasting* by Natalie Babbitt. Sixth grade language arts teachers also assigned spelling activities from a word list and instructed students in writing a narrative work and journal writing. The eighth grade control group read *The Yearling* by Marjorie Kinnan Rawlings, and the eighth grade treatment classes read *The Giver* by Lois Lowry. All eighth graders were taught academic research techniques, and they worked on vocabulary, comprehension, and journal writing. They also wrote several narrative and expository pieces.

No other instructional software programs were used for language arts instruction during the study. Where the Merit programs displaced curricula, the treatment group students did not make up work that the control group completed.

STUDY DESIGN AND METHODOLOGY

Because practical considerations prevented random assignment of students to treatment and control groups, we used the repeated measures procedure available with SPSS (Statistical Package for the Social Sciences) 11.5 Mixed Models. We recognize that randomization is the most adequate all-purpose method of avoiding initial biases between groups (Shadish, Cook, & Campbell, 2002). This accommodates unbalanced designs in which the number of cases per group varies and the total number of cases is not constant across all observations. The pretest was the SAT-9 2002 and the posttest was the SAT-9 2003.

The study's evaluation design used what Campbell and Stanley (1963) call

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

"The Nonequivalent Control Group Design," one of the most widespread experimental designs in educational research. A pretest and posttest are given to experimental and control groups without pre-experimental sampling equivalence. The experimental and control groups constitute, and are composed of, naturally assembled collectives, such as classrooms (which are as similar as availability permits, yet not so similar that one can dispense with the pretest). From a pool of subjects, the experimenter randomly assigns individual subjects. The assignment of X to one group or the other is assumed to be random and under the experimenter's control. This design is not to be confused with the "Pretest–Posttest Control Group" design, in which the experimental subjects are assigned randomly from a common population to the experimental or control group.

We were primarily interested in regression coefficients corresponding to two variables: Merit Group and Merit Sessions. Merit Group refers to the students who were assigned to the treatment group and Merit Sessions refers to the number of 45-minute sessions in which each student participated.

The numerical magnitude of a statistically significant and positive regression coefficient corresponding to Merit Group tells how much, on the average, a student's achievement growth was increased simply by the fact that he or she was a participant in the treatment group. The numerical magnitude of a statistically significant and positive regression coefficient corresponding to Merit Sessions indicated how much additional achievement growth, on average, occurred with actual participation in each instructional session.

The rationale for using regression coefficients was to decrease the chances of confounding effects, or the influence of contributing variables. The use of regression coefficients allowed the study of the effects of each independent variable upon the other independent variables (e.g., achievement growth on specific areas of the SAT-9).

Because practical considerations prevented random assignment to treatment and control groups, judiciously selected control variables, as outlined below, were used to address confounding. The study methodology uses the repeated measures procedure available with SPSS 11.5 Mixed Models to measure achievement growth over time. This accommodates unbalanced designs in which the number of cases per group varies, and the total number of cases is not constant across all observations. Growth curves are estimated using the information available, including instances in which test-takers skip one or more administrations of the treatment and then return.

In addition, the repeated measures procedure permits modeling of the error term to correct coefficient estimates for the effects of heteroscedasticity and autocorrelation. In the analyses reported below, autocorrelation was not present, but the error terms were modeled to accommodate departures from heteroscedasticity, thereby avoiding inflated standard errors and loss of statistical power.

Our growth model uses variables measured at three levels: Within students for repeated measures, between students, and between Merit and non-Merit groups. Use of multiple levels in our growth model reflects the fact that intercepts and slopes corresponding to relationships of interest may vary between students and within groups.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

For the analyses reported below, however, all intra-class correlations between the between-group and within-student levels were statistically non-significant. Moreover, although the intra-class correlations between the within-group and within-student levels were statistically significant, none of the slopes corresponded to relationships between the within-student independent variable. Time and our SAT-9 outcome measures had a statistically significant variance, although the intercept variance was statistically significant in each instance. As a result, the only parameter permitted to vary is the intercept corresponding to relationships between the within-student and within-group levels. The random intercept, thus, is specified as a function of the within-group independent variables:

$$?_{0J0} = ?_{000} + ?_{010}\text{GENDER} + ?_{020}\text{GPA} + ?_{030}\text{REPEAT} + ?_{040}\text{ETHNIC} + ?_{050}\text{LUNCH} + ?_{060}\text{EARLY} + ?_{070}\text{SPECIAL} + ?_{080}\text{PC} + u_{0J0}$$

The remaining intercept and slopes are fixed.

DATA ANALYSIS AND RESULTS

All variables are described in Table 1, and descriptive statistics computed before centering with respect to variable means appear in Table 2. Tables 3 through 11 report the regression analysis results for each dependent variable. All unstandardized regression coefficients are estimated using restricted maximum likelihood (REML) estimators, and the R^2_L goodness of fit measures are computed as shown in the tables.

Our primary interest is in coefficients corresponding to two variables, Merit Group and Merit Sessions. Merit Group is coded 1 for students who were assigned to the Merit group, and 0 for those not assigned to the Merit group. Merit Sessions refers to the number of 45-minute Merit instructional sessions in which each student participated.

Typically, we would expect variables created in this way to be closely correlated. In this instance, however, the Merit Group and Merit Session variables are orthogonal. This is due to the fact that Merit Sessions is best construed as an interaction effect computed by multiplying the Merit Group dummy variable by the number of Merit Sessions in which a student participated. Because the Merit Group variable is coded 0 for Merit non-participants, these products will be equal to zero. Because the Merit Group variable is coded 1 for Merit participants, these products will be equal to Merit Sessions. Given that Merit Groups is centered with respect to its grand mean and that Merit Sessions is centered with respect to group means, an interaction effect created using these variables will be orthogonal to them.

The coefficients corresponding to Merit Group and Merit Sessions, therefore, are interpreted such that the numerical magnitude of a statistically significant and positive regression coefficient corresponding to Merit Group tells how much, on average, a student's achievement growth is increased simply by virtue of the fact

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

that he or she is a Merit participant. The numerical magnitude of a statistically significant coefficient corresponding to Merit Sessions tells how much additional achievement growth, on average, comes with actual participation in each instructional session. The significance level for all group variables is .05.

Tables 3 through 11 show us that the Merit Group variable has a statistically significant and positive coefficient for seven of our nine SAT-9 dependent variables: Reading Vocabulary, Reading Comprehension, Math Problem Solving, Math Procedure, Language Expression, Science, and Social Science. Merit Sessions, moreover, has a statistically significant and positive coefficient for four of the nine dependent variables: Math Procedure, Language Mechanics, Language Expression, and Spelling.

In short, there are Merit achievement growth gains associated with each of our SAT-9 dependent variables.

Table 1: Variables

Independent	
Gender	Coded 1 if Male and 0 if Female
GPA	Grade Point Average at End of 2002–2003 School Year
Repeated Grades	Number of Grades Student Repeated
Ethnicity	Coded 1 if White and 0 Otherwise
Free Lunch	Coded 1 if Eligible for Free/Reduced-Cost Lunch and 0 Otherwise
Enrolled Late	Code 1 if Enrolled More than Six Weeks Late and 0 Otherwise
Special	Coded 1 if Designated Special Education and 0 Otherwise
Time	Coded 0 for Pretest and 1 for Posttest Over One Calendar Year
Personal Computer	Coded 1 if PC in the Home and 0 Otherwise
Grade	Either Sixth Grade or Eighth Grade
Size	Number of Students in Merit or Non-Merit Group
Merit Group	Coded 1 for Merit Groups and 0 Otherwise
Merit Sessions	Number of Merit Sessions in which Student Participated

Dependent (Selected from SAT-9 Battery)	
PRETESTS (End of Fifth/Seventh Grade)	POSTTESTS (End of Sixth/Eighth)
Reading Vocabulary	Reading Vocabulary
Reading Comprehension	Reading Comprehension
Math Problem Solving	Math Problem Solving
Math Procedure	Math Procedure
Language Mechanics	Language Mechanics
Language Expression	Language Expression
Spelling	Spelling
Science	Science
Social Science	Social Science

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 2: Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Gender	0.53	0.50	0	1
GPA	2.74	0.76	0.00	4.00
Repeated Grades	0.37	0.68	0	4
Ethnicity	0.98	0.12	0	1
Free Lunch	0.58	0.49	0	1
Enrolled Late	0.69	0.25	0	1
Special	0.08	0.27	0	1
Time	0.51	0.50	0	1
Personal Computer	0.74	0.44	0	1
Grade	7.01	1.00	6	6
Size	13.05	2.93	8	17
Merit Group	0.82	0.38	0	1
Merit Sessions	5.96	3.62	0	12
Reading Vocabulary	20.72	5.03	5.00	30.00
Reading Comprehension	35.04	9.24	3.00	52.00
Math Problem Solving	32.41	7.94	10.00	49.00
Math Procedure	18.62	6.37	3.00	30.00
Language Mechanics	15.59	4.45	3.00	24.00
Language Expression	15.11	4.79	3.00	24.00
Spelling	19.08	5.61	4.00	32.00
Science	25.29	6.29	7.00	39.00
Social Science	20.59	6.58	5.00	35.00

Table 3: SAT-9 Reading Vocabulary

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	20.61	65.01	.000
Gender	2.16	3.07	.000
GPA	3.10	6.17	.000
Repeated Grades	- 0.43	-0.83	.297
Ethnicity	- 5.70	-2.09	.018
Free Lunch	- 1.15	-1.52	.065
Enrolled Late	1.21	0.90	.184
Special Education	-3.57	-2.95	.002
Time	1.42	3.98	.000
Personal Computer	0.03	0.04	.484
Grade	0.09	0.25	.401
Group Size	- 0.16	-1.38	.084
Merit Group	2.77	2.80	.003
Merit Sessions	0.07	0.68	.248

**Significance Levels for One-Tailed Tests*

$N = 141$; $R^2_L = 1693.4 – 1353.9/1693.4 = 20.0\%$

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Tables 3 and 4 show that the Merit Group variable has statistically significant and positive regression coefficients when the SAT-9 reading vocabulary and reading comprehension tests are used as dependent variables. Participation in the Merit Group increases achievement growth in reading vocabulary, on average, by 2.77 points, while participation in the Merit Group increases achievement growth in reading comprehension, on average, by 3.67 points.

Table 4: SAT-9 Reading Comprehension

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	34.50	62.21	.000
Gender	2.13	1.74	.041
GPA	7.39	8.56	.000
Repeated Grades	- 0.92	- 1.03	.152
Ethnicity	- 9.64	- 2.04	.021
Free Lunch	- 0.87	- 0.67	.251
Enrolled Late	2.55	1.05	.147
Special Education	- 2.75	-1.32	.093
Time	1.11	1.63	.052
Personal Computer	- 0.77	- 0.55	.291
Grade	0.84	1.30	.097
Group Size	- 0.14	- 0.67	.251
Merit Group	3.67	2.15	.040
Merit Sessions	0.13	0.73	.306

**Significance Levels for One-Tailed Tests*

N = 141; R^2_L = 2059.0 – 1660.6/2059.0 = 19.4%

Table 5: SAT-9 Math Problem Solving

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	32.12	62.66	.000
Gender	4.35	4.04	.000
GPA	6.52	8.53	.000
Repeated Grades	- 1.11	- 1.40	.081
Ethnicity	- 11.01	- 2.62	.013
Free Lunch	- 0.96	- 0.84	.201
Enrolled Late	0.59	- 0.03	.977
Special Education	- 4.46	- 2.40	.008
Time	1.03	2.27	.012
Personal Computer	- 0.21	- 0.17	.433
Grade	- 1.36	- 2.38	.009
Group Size	- 0.14	- 0.75	.227
Merit Group	2.77	1.84	.033
Merit Sessions	0.37	0.73	.232

**Significance Levels for One-Tailed Tests*

N = 141; R^2_L = 1968.2 – 1532.0/1968.2 = 22.2%

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Tables 5 and 6 report regression results for dependent variables for which Merit Software modules are produced but that were not used in math problem solving and math procedures. Both analyses yield statistically significant and positive regression coefficients corresponding to the Merit Group variable.

With the SAT-9 math problem solving test as the dependent variable, membership in the Merit Group yields an increase in achievement growth equal, on

Table 6: SAT-9 Math Procedure

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	18.65	50.18	.000
Gender	2.64	3.20	.000
GPA	4.24	7.24	.000
Repeated Grades	- 1.56	- 2.56	.020
Ethnicity	- 1.41	- 0.44	.330
Free Lunch	0.37	0.41	.341
Enrolled Late	1.52	0.95	.171
Special Education	- 1.01	- 0.71	.239
Time	0.86	2.07	.019
Personal Computer	- 1.44	- 1.51	.067
Grade	- 2.95	- 6.74	.000
Group Size	0.12	0.89	.187
Merit Group	2.53	2.19	.014
Merit Sessions	0.38	3.17	.000

**Significance Levels for One-Tailed tests*

$N = 141$; $R^2_L = 1849.6 – 1449.1/1849.6 = 21.7\%$

Table 7: SAT-9 Language Mechanics

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	15.47	54.18	.000
Gender	0.41	0.65	.258
GPA	3.11	6.92	.000
Repeated Grades	- 0.80	- 1.70	.047
Ethnicity	- 0.42	- 0.17	.433
Free Lunch	- 1.08	- 1.60	.055
Enrolled Late	0.35	0.28	.390
Special Education	-2.76	- 2.54	.006
Time	0.67	2.18	.015
Personal Computer	- 0.77	- 1.05	.147
Grade	- 0.16	- 0.46	.323
Group Size	- 0.07	- 0.68	.248
Merit Group	0.93	1.05	.147
Merit Sessions	0.21	2.34	.010

**Significance Levels for One-Tailed Tests*

$N = 141$; $R^2_L = 1644.1 – 1310.6/1644.1 = 20.3\%$

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

average, to 2.77 points. Similarly, with the math procedures test as the dependent variable, membership in the Merit Group yields an increase in achievement growth equal, on average, to 2.53 points, while participation in each Merit Software session yields an additional 0.38 points.

Table 7 shows that, while the coefficient corresponding to Merit Group is not statistically significant, with the SAT-9 Language Mechanics score as the dependent variable, the coefficient corresponding to Merit Sessions is statistically significant and positive. For each instructional session in which a student actually participated, the achievement growth gain equaled, on average, 0.21 points.

Table 8: SAT-9 Language Expression

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	14.76	47.55	.000
Gender	0.94	1.37	.085
GPA	3.61	7.38	.000
Repeated Grades	- 0.49	- 0.97	.047
Ethnicity	- 4.02	- 1.50	.334
Free Lunch	0.70	- 0.95	.171
Enrolled Late	1.65	1.23	.391
Special Education	- 2.36	- 1.99	.023
Time	0.88	2.58	.005
Personal Computer	- 1.14	- 1.42	.078
Grade	0.04	0.11	.456
Group Size	0.57	0.49	.312
Merit Group	1.67	1.73	.042
Merit Sessions	0.19	1.86	.031

**Significance Levels for One-Tailed Tests*
N = 141; R^2_L = 1692.0 – 1359.9/1692.0 = 19.6%

Table 8 shows that, with the SAT-9 Language Expression test score as the dependent variable, both the Merit Group and Merit Software Sessions have coefficients that are statistically significant and positive. Membership in the Merit Group yields an achievement growth increase equal, on average, to 1.67 points. Furthermore, each Merit Software Session in which students actually participate yields an additional achievement growth increase equal, on average, to 0.19 points.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 9: SAT-9 Spelling

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	18.87	50.80	.000
Gender	0.09	0.11	.456
GPA	3.23	5.53	.000
Repeated Grades	- 0.31	- 0.51	.305
Ethnicity	- 2.46	- 0.77	.221
Free Lunch	-1.81	- 2.06	.020
Enrolled Late	1.36	0.86	.195
Special Education	- 5.91	- 4.17	.000
Time	0.16	0.43	.334
Personal Computer	- 1.15	- 1.20	.115
Grade	- 0.16	- 0.37	.356
Group Size	0.04	0.32	.375
Merit Group	1.58	1.38	.084
Merit Sessions	0.33	2.76	.003

**Significance Levels for One-Tailed Tests*
N = 141; R^2_L = 1762.6 – 1425.3/1762.6 = 19.1%

Table 9 shows that Merit Sessions has a statistically significant and positive coefficient for Spelling.

Table 10: SAT-9 Science

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	25.22	71.33	.000
Gender	3.99	5.10	.000
GPA	4.09	7.44	.000
Repeated Grades	- 1.69	- 2.92	.002
Ethnicity	- 810	- 2.66	.004
Free Lunch	0.50	0.59	.278
Enrolled Late	1.98	1.28	.100
Special Education	- 2.96	- 2.19	.014
Time	1.46	3.21	.001
Personal Computer	- 0.60	- 0.66	.256
Grade	- 1.28	- 3.08	.001
Group Size	- .010	- 0.76	.224
Merit Group	2.25	2.05	.020
Merit Sessions	0.14	1.21	.113

**Significance Levels for One-Tailed Tests*
N = 141; R^2_L = 1825.9 – 1459.1/1825.9 = 20.1%

Table 10 indicates that the Merit Group has a coefficient that is statistically significant and positive.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Table 11: SAT-9 Social Science

Parameter	Parameter Estimate	t Value	Sig. Level*
Intercept	20.31	51.06	.000
Gender	2.82	3.20	.001
GPA	4.84	7.74	.000
Repeated Grades	- 0.97	- 1.49	.068
Ethnicity	- 6.94	- 2.03	.021
Free Lunch	0.37	0.39	.348
Enrolled Late	- 3.16	- 1.85	.032
Special Education	- 1.33	- 0.87	.192
Time	0.62	1.30	.097
Personal Computer	- 0.89	- 0.66	.256
Grade	- 1.23	- 2.62	.004
Group Size	- 0.11	- 0.71	.239
Merit Group	2.35	1.91	.028
Merit Sessions	0.20	1.60	.055

**Significance Levels for One-Tailed Tests*
N = 141; R^2_L = 1855.5 – 1499.1/1855.5 = 19.2%

Table 11 shows that the Merit Group has a coefficient that is statistically significant and positive.

CONCLUSIONS

This study confirms the results of the research that precedes it: Computer based instruction, when used to complement everyday classroom instruction by teachers, can significantly improve the performance of middle school students' reading skills as measured by standardized tests.

The Merit Group variable had a statistically significant and positive coefficient for seven of the nine SAT-9 dependent variables: Reading Vocabulary, Reading Comprehension, Math Problem Solving, Math Procedure, Language Expression, Science, and Social Science. Merit Sessions, moreover, had a statistically significant and positive coefficient for four of the nine dependent variables: Math Procedure, Language Mechanics, Language Expression, and Spelling. In summary, Merit achievement growth gains were associated with each of the SAT-9 dependent variables.

Membership in the Merit Group increased achievement growth in SAT-9 Reading Vocabulary, on average, by 2.77 points, and achievement growth in Reading Comprehension, on average, by 3.67 points. Membership also yielded an average gain of 1.67 points for SAT-9 Language Expression; furthermore, each session in which the experimental members participated yielded a gain of 0.19 points.

With a suitable complement of controls in place, membership in the Merit Group increased growth in SAT-9 Reading Vocabulary by 13.1% of the total sample mean score, which was 20.72. Membership also increased achievement in Reading Comprehension by 10.5% of the sample mean score of 35.04. (Table 12).

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Furthermore, the progress in critical thinking and reading comprehension skills that the Merit intervention cultivates would likely transfer to other academic areas, including math and science. Although the company does not produce learning modules for Science and Merit modules for Math were not used, analyses showed statistically significant and positive regression coefficients corresponding to the Merit Group. Membership yielded an increase in achievement in Math Problem Solving, on average, of 2.77 points. Similarly, in Math Procedure, membership yielded an average improvement of 2.53 points with an additional 0.38 points for each session. Finally, in Science, membership yielded an achievement growth, on average, of 2.25 points.

Table 12: Merit Group Percentage Achievement

SAT-9 Groups	Sample Mean	Point Gain	Percentage Gain
Reading Vocabulary	20.72	2.77	13.1%
Reading Comprehension	35.04	3.67	10.5%
Language Expression	15.11	1.67	11.1%
Math Problem Solving	32.41	2.77	8.5%
Math Procedure	18.62	2.53	13.6%
Science	25.29	2.25	8.9%
Social Science	20.59	2.35	11.4%

Consistent with our quasi-experimental approach, causal inferences are premised on four criteria: association, causal priority, nonspuriousness, and rationale. Statistically significant regression coefficients satisfy the association criterion. Because the Merit intervention was invoked before the second administration of the SAT-9, the causal priority criterion is satisfied. The control variables are intended to satisfy the nonspuriousness criterion, meaning that they enable us to tentatively judge that the relationships between Merit Group, Merit Sessions, and our achievement growth dependent variables are not effects of one or more common cause.

The case for a compelling rationale rests on the answer to the question: Does participation in Merit interventions have implications for educational reform? Given this new era of accountability, many education officials at the school and district levels might understandably give priority to improved test-taking ability. We can say with a good deal of confidence that participation in Merit interventions yielded higher average scores on a widely used, high-stakes standardized test.

The evaluation team for this study recommends that Merit Software conduct another study of the use of Merit reading software using an experimental research design. The project could focus on the three main groups emphasized by the NCLB: minorities, low SES, and special education students. Such a study could employ a stratified random sampling for these three groups.

The sampling design should meet four broad criteria:

Goal orientation. The sampling design should be tailored to the evaluation design and should be based on the study's goals and objectives.

Measurability. The sampling design provides the data for the necessary analysis.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Practicality. The conceptual design should conform to the actual situation.

Economy. Economy requires that the evaluation objectives be met with available time, financial, personnel, and any other necessary resources.

The study could use a proportional allocation, whereby each of the three strata contributes to the sample number that is proportional to its size in the population (e.g., in the grades/schools selected for the study). Stratified sampling guards against wild samples, ensures that no population will be omitted from the sample, and avoids overloading in certain sub-populations. The proposed study could address the time variable (i.e., the amount of time that teachers of the treatment group spent on each Merit software program and how much time the school district allocated to language mechanics and vocabulary instruction as set forth in their curriculum guides).

In summary, the proposed study should focus on identified needs to help schools and school districts meet or exceed the new student achievement standards required by the NCLB Act. In West Virginia, 37% of all high, middle, and elementary schools are below the required standards for low SES students. All 55 West Virginia school districts have one or more schools that do not meet the new standards. These districts include urban, suburban, and rural schools that would provide a variety of research settings where certain variables (such as class size, length of school day/year, state adopted curriculum and textbooks, attendance and graduation requirements as well as state funding) could be controlled.

The state-adopted testing program would enable school districts to disaggregate test data by quartile for students who are economically disadvantaged, from racial and ethnic minority groups, and from those who have disabilities and have limited English proficiency. The evaluators could assist school principals with random assignment of students to classrooms that could become treatment and comparison groups. The proposed study could include tests of Merit software in reading, mathematics, and other subjects.

Conducting new, creatively designed studies focused on the NCLB standards would enhance student achievement and help establish the same credibility that the public sees for the uses of technology in banking, medicine, manufacturing, commerce, and many other fields.

Contributors

Jerry D. Jones is an associate professor of Leadership Studies at the Marshall University Graduate College in South Charleston, West Virginia. He has been a superintendent of two school districts in West Virginia. Most recently he has focused on helping school districts manage their resources effectively. (Address: Jerry D. Jones, Marshall University Graduate College, 100 Angus E Peyton Drive, South Charleston, WV 25303; jonesjd@marshall.edu.)

William D. Staats has also held managerial positions in school districts. He was a Superintendent at the Wood County Schools in West Virginia and an Assistant Superintendent for Program and Pupil Evaluation in the Duval County Schools in Florida.

Noel Bowling is a professor of Reading Education at the Marshall University Graduate College. He has published numerous articles on reading education.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

Robert D. Bickel is a professor of Advanced Educational Studies at the Marshall University, Huntington, West Virginia campus. He has published several papers in education and public policy relating to research and evaluation.

Michael L. Cunningham is an associate professor of Leadership Studies at the Marshall University Graduate College and a former elementary school principal. Connie Cadle is a doctoral student in Marshall University's Leadership Studies department.

References

Anderson, L., & Krathwohl, D. (Eds.) (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives.* White Plains, NY: Longman Publishing Group.

Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally College Publishing Company.

Greenlee-Moore, M. E., & Smith, L. L. (1996). Interactive computer software: The effects on young children's reading achievement. *Reading Psychology*, *17,* 43–64.

Hart, T. A. (2002). *The effectiveness of hypermedia in West Virginia history: A study of the application of the hypercard program, the battle of Philippi.* Unpublished doctoral dissertation. West Virginia University, Morgantown, WV.

Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading & Writing Quarterly*, *20*, 123–144.

Ligas, M. R. (2002), Evaluation of Broward County Alliance of Quality Schools Project. *Journal of Education for Students Placed at Risk, 7*(2), 117–139.

Mann, D., Shakeshaft, C., Becker, J., & Kottkamp, R. (1999). *West Virginia story: Achievement gains from a statewide comprehensive instructional technology program. Research report.* (Eric Document Reproduction Service No. ED 429 575).

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved June 9, 2004 from http://www.nichd.nih.gov/publications/nrp/smallbook.htm.

NCLB Dominates at FETC. (2003). *Technology and Learning*, *23*(9), 6.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin Co.

Skinner, B. F. (1958). Teaching machines. *Science, 128*(130), 969–977.

Traynor, P. L. (2003). Effects of computer-assisted instruction on different learners. *Journal of Instructional Psychology, 30*(2), 137–143.

Wolpert, G., & Fitzpatrick, C. (2001). *Collegians, community, kids and computers.* Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, March, Dallas, TX.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.

APPENDIX

Student Coursework Materials Cited in the Study

Babbitt, N. (1975). *Tuck Everlasting*. New York: Farrar, Straus and Giroux.

Blau, S. (1998). *The Writer's Craft*. Evanston, IL: McDougal Littell.

Lowry, L. (1993). *The Giver*. Boston: Houghton Mifflin/Walter Lorraine Books.

O'Dell, L., Vacca R., Hobbs R., Irvin, J., & Warriner J. (2001). *Elements of Language, Second Course*. Austin, TX: Holt, Rinehart and Winston.

Probst, R. (2000). *Elements of Literature, Introductory Course*. Austin, TX: Holt, Rinehart and Winston.

Probst, R. (2000). *Elements of Literature, Second Course*. Austin, TX: Holt, Rinehart and Winston.

Rawlings, M. K. (1985). *The Yearling*. New York: Atheneum Books for Young Readers.

Copyright © 2004, ISTE (International Society for Technology in Education), 800.336.5191 (U.S. & Canada) or 541.302.3777 (Int'l), iste@iste.org, www.iste.org. All rights reserved.