RMLE Online
Research in Middle Level Education

Micki M. Caskey, Ph.D., Editor
Portland State University
Portland, Oregon

2008 • Volume 32 • Number 3 ISSN 1940-4476

# The Influence of Computer-Assisted Instruction on Eighth Grade Mathematics Achievement

Christopher H. Tienken

Seton Hall University
South Orange, NJ

James A. Maher

Maher Consulting
Holmdel, NJ

## Abstract

The issue of lower than expected mathematics achievement is a concern to education leaders and policymakers at all levels of the U.S. PK–12 education system. The purpose of this quantitative, quasi-experimental study was to determine if there was a measurable difference in achievement on the mathematics section of the state test for students ($n = 121$) from a middle school in New Jersey who received computer-assisted instruction (CAI) in drill and practice computation related to the eighth grade mathematics curriculum standards compared to students ($n = 163$) who did not receive the CAI. The results suggest that the CAI intervention did not improve student achievement significantly ($p > .05$). In two categories, students who received the CAI performed significantly lower than their peers in the comparison group. Students in the control group who scored in the 25th percentile on the seventh grade CTB/McGraw Hill TerraNova pretest outperformed their peers in the treatment group on the New Jersey Grade Eight Proficiency Assessment (GEPA) mathematics section. Likewise, Asian students in the control group outperformed all other students in treatment and control groups. The results fit within the existing knowledge on the subject of computer-assisted instruction and add support to the idea that practitioners should evaluate curriculum and instruction interventions for demonstrated success before they bring them into the learning environment.

## Introduction

The issue of lower than expected mathematics achievement is a persistent worry to some education leaders and policymakers at all levels of the U.S. PK–12 education system. The 1999 Third International Mathematics and Science Study Report (TIMSS-R) showed an example of the reported weaknesses of mathematics achievement of U.S. students compared to students in other industrialized countries. Grade 8 students in the United States ranked lower than 14 of the 38 participating nations (National Center for Education Statistics [NCES], 2000). In addition, 15-year-old students from the United States ranked between 16th and 23rd of 31 countries that participated in the mathematics portion of the 2000 Programme for International Student Assessment (PISA) administration (Organisation for Economic

Co-Operation and Development [OECD], 2004). On the national level, the 2005 (NCES, 2005) administration of the National Assessment of Education Progress (NAEP)[1] mathematics test indicated only 30% of grade 8 students scored "at or above proficient." While the validity of the NAEP achievement levels has not yet been demonstrated, the results influence policymakers. These achievement statistics raise concerns for some education leaders and policymakers about the mathematics achievement of U.S. middle school students.

Middle school students in New Jersey are not immune to this issue. New Jersey had a greater percentage of its students score proficient (30%) on the 2005 grade 8 NAEP mathematics test than the national average (24%). However, grade 8 NAEP New Jersey scale-score performance gaps exist between subgroups such as students eligible for free or reduced-price lunch and students not eligible for free or reduced-price lunch; 262 scale-score points and 292 scale-score points, respectively. This is a growing issue across the country. For example, the Southern Education Foundation (2007) reported that the percentage of economically disadvantaged students now outnumbers non-economically disadvantaged students in southern states. Childhood poverty rates range from a low of 20% in New Hampshire to a high of 84% in Louisiana. The expanding scourge of childhood poverty across the nation, and the corresponding negative influence on achievement, requires education leaders to use interventions with demonstrated records of success.

## Review of Related Literature

### Computer-Assisted Instruction and Student Achievement in Middle School Mathematics

We reviewed the results of experimental and quasi-experimental studies on the effect of computer-assisted instruction (CAI) on middle school student achievement in mathematics. An immediate issue with the middle school mathematics CAI knowledge dynamic was that few studies existed that met the federal definition of scientifically based research (SBR) and many of the studies that met the definition were conducted prior to the year 2000. In this section, we provide representative examples of the existing experimental and quasi-experimental studies on CAI drill and practice and achievement in middle level mathematics.

Roberts and Madhere (1990) found that CAI had a small positive effect on the overall mathematics achievement of 743 elementary and junior high school students. Students who participated gained 3.06 points on their Normal Curve Equivalent scores on a nationally normed standardized test of mathematics compared to students who did not have the CAI. Roberts and Madhere did not report effect sizes. Traynor (2003) found that CAI improved mathematics achievement of regular education, special education, and limited English proficient middle school students ($n = 161$) on a mathematics pretest-posttest when compared to traditional, teacher-directed practice techniques. The students comprised intact groups based on the way the middle school scheduled students into exploratory classes. Results were statistically significant ($p < .001$) with a moderate effect size ($d$) of 0.47 favoring the treatment group. Social scientists consider an effect size of 0.2 as small, an effect size in the range of $0.2 < d < 0.8$ as moderate, and an effect size greater than 0.8 as large (Cohen, 1988). Plano (2004) found that CAI activities for algebra had a non-significant predictive influence on student achievement overall but had a slightly significant influence on the algebra achievement of English language learners. Tienken and Wilson (2007) conducted a quasi-experimental, pretest-posttest control-group study and found a small, but statistically significant positive effect of CAI drill and practice computation exercises on the mathematics achievement of seventh grade students on the CTB/McGraw Hill TerraNova full battery mathematics test. They reported an effect size ($d$) of 0.12.

Campbell, Peck, Horn, and Leigh (1987) found no significant difference in the mathematics achievement of third grade students who used CAI drill and practice activities compared to students who used only print drill and practice materials. Rosenberg (1991) found a negative influence of computers on instruction and achievement. He stated that the computer failed to deliver on the promises of increased efficiency (i.e., take less time for students to learn the concept) and effectiveness (i.e., higher student achievement than with traditional paper/pencil methods). Recent studies demonstrated similar results. Baker, Gersten, and Lee (2002) conducted a synthesis of studies on the influence and effect of CAI on mathematics achievement of low-achieving students. They found low achievers did not perform statistically significantly better. They observed an average effect size ($d$) of 0.01.

The empirical literature on CAI and middle school mathematics achievement is thin and the results are mixed. The findings related to middle school mathematics achievement and the use of CAI is congruent to those found in a recent report by the U.S. Department of Education, Institute of Education Sciences (IES). IES conducted a review of the effectiveness of CAI in mathematics on grade 6 student achievement and found no statistically significant effect, while an algebra CAI program had a positive statistically significant effect ($p < .05$) on student achievement in junior high school. The overall findings suggested mixed effects of CAI on student mathematics achievement (USDOE, 2007).

### Theoretical Perspective: Active Learning

Like CAI, active learning is designed to improve student achievement. Cooperstein and Kocevar-Weidinger (2004) noted that active learning occurs when (a) the learner can construct his or her own meaning, (b) current learning is developed on previous learning, (c) the learner is involved in meaningful social interaction, and (d) the learning is built using authentic involvement with the learning materials. Examples of active learning pedagogy include inquiry-based learning, discovery-based learning, hands-on learning, and problem-based learning. The roots of current active learning methodology reach back 200 years beginning with Pestalozzi's Object Teaching and Froebel's Kindergarten, and more recently by Dewey's ideas of experiential learning. Landmark projects during the 1930s and 1940s such as Wrightstone's study (1935), the New York City Experiment (Jersild, Thorndike, & Goldman, 1941), and the Eight Year Study (Aikin, 1942) demonstrated the power of active learning to have a positive effect on student achievement and attitudes toward learning compared to traditional approaches.

Some studies demonstrated that active learning was an effective method of enhancing students' learning. However, a glaring limitation of the recent literature in this area is that in many cases, quasi-experimental and experimental designs were not used, effect sizes were not reported, and overall methodology was suspect. Nonetheless, several studies reported positive outcomes. Hetland (2000) concluded that students' active involvement in music had an effect on the development of their spatial thinking. Wilson, Flanagan, Gurkewitz, and Skrip (2006) found that students' active involvement in origami resulted in increased problem-solving ability. Cerezo (2004) conducted a qualitative study and reported that active involvement in problem solving enhanced the learning of mathematics for at-risk female students. Huffaker and Calvert (2003) conducted a review of the literature related to active learning through online games and concluded that active learning was particularly useful when used in problem solving with computers. One complaint against active learning is that teachers sometimes mistakenly leave students on their own, and thus, the learning process becomes unguided and disconnected (Kirschner, Sweller, & Clark, 2006).

### Purpose

Middle level education leaders search for scientifically based interventions (U.S. Department of Education, 2002) to address issues related to improving mathematics achievement. The knowledge dynamic on the influence or effect of CAI on middle school mathematics achievement is not well developed and the results from previous studies are mixed. The results from this study add to the experimental/quasi-experimental CAI literature available to education leaders.

We present findings from an evaluation of a middle school mathematics intervention implemented during the 2004–2005 school year to improve students' mathematics performance on the New Jersey Grade Eight Proficiency Assessment (GEPA). The purpose of this quasi-experimental study was to determine if there was a measurable difference in achievement on the mathematics section of the GEPA for students from a middle school in New Jersey who received computer-assisted instruction (CAI) in drill and practice computation related to the eighth grade mathematics curriculum standards compared to students who did not receive the CAI.

### Problem

The central New Jersey school under study served 895 students in grades 7 and 8 during the 2004–2005 school year. Almost 34% of the students were eligible for the federal free or reduced-price lunch program and approximately 46% were non-white. The New Jersey Department of Education (NJDOE) rated the school "in need of improvement" *Level 4* during the 2003–2004 school year. Approximately 55% of the students in grade 8 scored *Partially Proficient* on the mathematics section of the GEPA. *Partially Proficient* is the lowest of three performance categories developed by the NJDOE. The need for improvement was urgent. Failure to improve could lead to sanctions such as restructuring the school or outsourcing the school to a private company.

Although some controversy exists about the effective use of CAI, particularly with respect to the drill and practice forms associated with simple knowledge development, the literature suggested a small, positive effect of active learning on mathematics achievement. The literature also suggested a positive influence occurred primarily when CAI integrated more complicated kinds of learning, such as open-ended, divergent problem solving. From the research reviewed, it was not clear, however, whether using active learning with simpler CAI processes such as those associated with computation-based drill and practice computer software and websites would have a positive influence on student achievement as measured by the GEPA.

### Questions

We examined how the use of a drill and practice CAI in combination with a less complex active learning follow-up exercise, direct instruction of how to use computer presentation software (Microsoft PowerPoint$^{TM}$) to communicate understanding of the drill and practice exercises, influenced student achievement of grade 8 mathematics skills and knowledge.

This study was guided by our desire to evaluate the influence of mathematics drill and practice CAI combined with the use of multimedia presentation software on mathematics achievement of the following groups of regular education grade 8 students: (a) total population of regular education students; (b) students who received basic skills instruction (BSI) in mathematics, language arts, or in both subjects; (c) various ethnic groups; and (d) socioeconomically disadvantaged (i.e., eligible for federal free or reduced-price lunch program).

## Methodology

We used a quasi-experimental pretest/posttest control-group design because students comprised intact groups and random assignment of students was not possible. The design controlled effectively for most threats to internal validity (Campbell & Stanley, 1963). Internal validity is the extent that the experiment demonstrates a cause and effect relationship between the independent and dependent variables. The design overcomes the threat to internal validity posed by the interaction of selection of participants and maturation, the time between pretest and posttest, because of the large sample sizes of students and the short duration of the study. The pretest-posttest design mitigated further the threat posed by maturation because all participants experienced the pretest and posttest. Theoretically, any influences of maturation would be experienced by both groups, experimental and control, and thus, neutralize the maturation threat to internal validity.

We assigned teachers randomly to experimental (*n* = 2) and control (*n* = 2) groups and compared students based upon their pretest mathematics achievement. Because the pretest was part of an existing testing program, the potential threat to external validity posed by the interaction between the pretesting and treatment was reduced.

The study used a sample of eighth grade students and the total population of four eighth grade regular education mathematics teachers from one middle school in New Jersey. The NJDOE categorized the school as “needs improvement” based on lower than expected prior student achievement on the mathematics and language arts sections of the GEPA. The experimental group included 121 students and the control group included 163 students (total *n* = 284). We collected data from all students who met the following criteria: (a) received a valid score on the Grade 7 mathematics section of the TerraNova test (CTB-McGraw Hill, 2007), (b) received a valid score on the GEPA mathematics section, (c) enrolled in the school for the entire seventh and eighth grade years, and (d) enrolled in a regular education program in the school for the entire seventh and eighth grade years. We excluded students who received special education services from the analysis due to the individualized nature of those programs.

### Treatment

We assigned randomly the total population (*n* = 4) of eighth grade mathematics teachers to experimental and control groups prior to the start of the study. The teachers in the experimental group used mathematics drill and practice websites and slide presentation software with students. The teachers in the control group used neither the websites nor the presentation software. The purpose of the CAI treatment was to provide students practice with basic mathematics skills related to the Grade 8 New Jersey Core Curriculum Content Standards (NJCCCS). The mathematics websites provided students opportunities for drill and practice of computation in operations, fractions, geometry, data analysis, and algebra based on the NJCCCS and the school's mathematics curriculum. A site facilitator (i.e., district mathematics supervisor) observed the instruction of the teachers in the experimental group to monitor frequency of implementation, and when necessary, coached the teachers on how to access and use the mathematics websites.

After students became familiar with the CAI, the teachers taught them to use slide presentation software to create a digital "book report" to explain one aspect of mathematics they learned via the CAI. Each student used the slide presentation software to construct an explanation of the material he/she learned from using the drill and practice CAI. Upon completion of the CAI work, the students in the experimental groups presented the information to their classmates. The students used the CAI technology two sessions per week, 45 minutes per session, for 20 weeks. They used the CAI during their regularly scheduled mathematics period. There was no difference in the amount of time that the students in the experimental and control groups participated in mathematics instruction. The CAI was not an add-on and did not result in more mathematics time on task for the students in the experimental group.

The site facilitator ensured that the mathematics content was consistent for all teachers and that the teachers and students in the experimental group were the only ones using the mathematics websites and presentation software. The site facilitator conducted weekly classroom observations of the experimental and control teachers and reviewed lesson plans weekly. Teachers in the experimental group facilitated student creation of slide shows to demonstrate their understanding of mathematics concepts such as adding and subtracting fractions with unlike denominators.

### Hypotheses

We examined whether there is evidence to reject one or more of the following hypotheses:

> $H_0$: There is no difference in mean score achievement between the experimental and control group students on the mathematics section of the New Jersey GEPA for the following subsets of regular education students: (a) students who scored in the same quartile of the TerraNova grade 7 math assessment, (b) students who participated in similar basic skill instruction (BSI) math and/or reading remediation service programs, (c) students who did not participate in BSI math and/or reading remediation service programs, (d) students who were in the same ethnic group, and (e) students who participated in the same level of the school's free or reduced-price lunch program.

In addition, we examined if there was evidence that the odds of a student scoring at the proficient or above proficient level on the GEPA mathematics section was higher for the students in the experimental group compared to those in the control group.

## Analysis

The purpose of the statistical analysis is an examination of factors expected to explain success or failure on the New Jersey GEPA mathematics test. These factors include the experimental versus control curriculum (i.e., CAI enhanced vs. traditional), student achievement on the TerraNova mathematics pretest; student referral or not to basic skills instruction (BSI) sessions in math, language/reading, or both mathematics and language; ethnicity; and the student's socioeconomic status (via the level of participation in the school's free or reduce-priced lunch program).

Analysis of variance (ANOVA) methods were used to derive linear models of best fit for the raw data summarized in Tables 1 through 4. A factor was included in an ANOVA model only if the factor was statistically significant at the .05 level of significance or lower. The resulting model was used to estimate the residual variability not explained by the model and then to derive 95% confidence intervals for the predicted GEPA mathematics score for each group of students identified by the cell descriptors. The means of two groups of students are declared statistically significant when their corresponding 95% confidence intervals do not overlap.

### Limitations

The small population of available teachers ($n = 4$) created external validity concerns and limited the ability to generalize results beyond the school in this study. Likewise, the demographic and socioeconomic makeup of the student population limited the ability to generalize student results beyond districts located in lower socioeconomic communities. Results may be different for students in schools located in higher socioeconomic communities. While the design was quasi-experimental and controlled for major threats to internal validity, the statistics used were to determine whether the CAI influenced achievement. Thus, the results do not demonstrate cause and effect, but merely the existence or lack of a relationship between CAI and achievement.

## Strengths

Potential internal validity issues posed by instrumentation were reduced because both groups took the same pretest and posttest assessments. The pretest was the mathematics section of a nationally normed, commercially prepared standardized test with reported full-test reliability estimates of .90 (CTB/McGraw-Hill, 1997). The posttest was the mathematics section of the New Jersey GEPA. The NJDOE reported full test reliability of .91 for the 2005 administration of the GEPA (NJDOE, 2005). Ecological validity issues were limited because the study took place in the school setting under existing constraints. We did not create artificial contexts and we worked within the existing confines (i.e., used only preexisting assessment tools and grading procedures, did not reassign students to alternative groupings, did not reassign staff to different grade levels). The potential external validity threat posed by the Hawthorne effect was mitigated because both groups used the same curriculum and textbook, spent the same amount of time in mathematics classes, and a site supervisor monitored the teachers in each group throughout the process to ensure continuity of instruction and program.

Threats due to maturation were accounted for as stated in the methods section. Issues due to temporal validity were accounted for by comparing achievement of the groups based on their quartile achievement from the grade 7 pretest. That is, achievement of students was not measured solely on a posttest, aggregate basis. We matched student achievement from the pretest quartiles and then compared the posttest achievement of the quartile groups. Thus, we were able to control for prior achievement of the students in each group.

CAI is a specific independent variable identified in the knowledge dynamic that can influence student achievement. Other variables that could potentially influence student achievement in mathematics include curriculum, the teacher, professional development, and special instructional programs such as special education, basic skills instruction, or gifted education.

Table 1

*Grade 8 Mathematics GEPA Score Mean/SD vs. Experimental/Control Group Placement & TerraNova Pretest Score Classification for Regular Education Students.*

| Classification | Terra Nova Pretest | Actual Mean/Standard Deviation, (Predicted Mean), (Sample Size), & 95% Confidence Intervals for Mean Predicted Grade 8 Math GEPA Score | |
|---|---|---|---|
| | | Experimental | Control |
| Regular Education | 25Q | 169.0/10.55<br>(169.0) (*n* = 14)<br>156.90 – 181.10 | 211.43/27.60<br>(211.43) (*n* = 21)<br>201.55 – 221.30 |
| | 50Q | 185.4/14.31<br>(185.41) (*n* = 37)<br>177.97 – 192.85 | 199.44/26.51<br>(199.44) (*n* = 25)<br>190.39 – 208.49 |
| | 75Q | 202.25/16.23<br>(202.25) (*n* = 40)<br>195.09 – 209.41 | 206.17/28.42<br>(206.17) (*n* = 53)<br>199.95 – 212.39 |
| | UQ | 218.87/18.86<br>(218.87) (*n* = 30)<br>210.60 – 227.13 | 206.20/32.47<br>(206.20) (*n* = 64)<br>200.55 – 211.86 |
| | Regular Class Statistics | 197.37/22.51<br>(197.37) (*n* = 121)<br>192.75 – 201.99 | 205.83/29.63<br>(205.83) (*n* = 163)<br>201.85 – 209.81 |

Table 2
*95% Confidence Intervals for Mean GEPA Score for BSI Math Referral and Experimental/Control Groups*

| BSI Math Referral Classification | Terra Nova Pretest | Mean/Standard Deviation, (Predicted Mean), (Sample Size), & 95% Confidence Intervals for Mean Predicted Grade 8 Math GEPA Score | |
|---|---|---|---|
| | | Experimental | Control |
| No | 25Q | — | 215.38/27.39<br>(216.81) (*n* = 18)<br>207.43 – 226.18 |
| | 50Q | 190.30/14.15<br>(188.91) (*n* = 23)<br>180.32 – 197.50 | 209.19/28.14<br>(212.99) (*n* = 16)<br>203.99 – 221.99 |
| | 75Q | 201.87/16.51<br>(202.72) (*n* = 38)<br>195.94 – 209.49 | 214.71/25.34<br>(213.99) (*n* = 42)<br>207.89 – 220.07 |
| | UQ | 218.87/18.86<br>(218.87) (*n* = 30)<br>211.08 – 226.65 | 216.16/29.72<br>(215.03) (*n* = 49)<br>209.38 – 220.68 |
| | No BSI Math Referral: Total | 204.55/19.97<br>(204.55) (*n* = 91)<br>199.90 – 209.20 | 214.67/27.53<br>(214.67) (*n* = 125)<br>210.70 – 218.64 |
| Yes | 25Q | 169.0/10.55<br>(169.0) (*n* = 14)<br>157.60 – 180.40 | 187.67/15.88<br>(179.16) (*n* = 3)<br>167.61 – 190.71 |
| | 50Q | 177.36/10.76<br>(179.65) (*n* = 14)<br>168.90 – 190.39 | 182.11/9.82<br>(175.35) (*n* = 9)<br>165.40 – 185.29 |
| | 75Q | 209.5/9.19<br>(193.45) (*n* = 2)<br>179.30 – 207.61 | 173.55/9.47<br>(176.34) (*n* = 11)<br>167.71 – 184.96 |
| | UQ | — | 173.67/15.31<br>(177.38) (*n* = 15)<br>169.27 – 185.49 |
| | No BSI Math Referral: Total | 175.6/14.37<br>(175.6) (*n* = 30)<br>167.50 – 183.70 | 176.74/13.08<br>(176.74) (*n* = 38)<br>169.54 – 183.93 |

As mentioned earlier, the curriculum, teachers, and professional development remained constant during the period under study. We accounted for special programs by excluding students in special programs from the analyses.

Interpretive validity was strengthened through the quasi-experimental design and the way in which we monitored the implementation of the treatment. Organizational, structural, and instructional conditions other than CAI for the experimental group were remarkably stable during the 20-week period.

## Results

Table 1 relates GEPA mathematics test performance for the experimental and control groups of students to the student's performance on the grade 7 TerraNova pretest and provides the mean and standard deviation GEPA math summary statistics for each quartile of student scores on the TerraNova pretest. The Analysis of Variance (ANOVA) model with a full set of significant interaction terms was used to derive predicted 95% confidence intervals for the mean cells.

Performance on the GEPA mathematics test was correlated with the student's performance on the TerraNova mathematics pretest for regular class students. It is useful to contrast the GEPA test scores of experimental and control groups by comparing students who scored in similar quartiles of the TerraNova mathematics pretest. In Table 1, the 95% mean confidence interval estimates overlap for all comparisons except one. Namely, regular education students in the control group who scored within the 25th percentile of the TerraNova mathematics test, performed higher, statistically significant ($p < .05$), on the GEPA mathematics test than did students in the experimental group. An effect size was calculated using the formula developed by Glass (1976) where the difference of mean of the experimental and control groups is divided by the standard deviation of the control. An effect size of 1.53 favoring the control group students in the 25th quartile was observed.

The first hypothesis stated there is no difference in achievement on the mathematics section of the New Jersey GEPA between regular education students in the experimental and control groups who scored in the same quartile on the grade 7 TerraNova pretest. The results suggest a difference favoring control group students who scored in the 25th percentile on the TerraNova pretest in grade 7. Overall, there is not evidence that the CAI program influenced the average achievement of students in the experimental group positively compared to the students in the control group.

Table 2 relates GEPA mathematics performance for the experimental and control groups to whether the student participated in basic skills instruction (BSI) mathematics remediation as well as the student's quartile performance on the grade 7 TerraNova pretest. An ANOVA model with two interaction terms (TerraNova pretest score—experimental/control group interaction and a BSI mathematics referral—experimental/control group interaction) was used to derive predicted 95% confidence intervals for the mean cells in Table 2.

Performance on the GEPA mathematics test correlated highly with the student's performance on the TerraNova mathematics test. The data provide evidence that students in the control group not referred for mathematics BSI services scored statistically significantly ($p < .05$) higher on the GEPA mathematics test than did the corresponding experimental group (See the non-overlapping 95% confidence intervals in Table 2 for the No BSI referral group totals of the experimental and control groups). An effect size of 0.36 favoring the control group students who did not participate in mathematics basic skills was observed.

Table 3
*95% Confidence Intervals for Mean GEPA Score for Ethnicity and Experimental/Control Groups*

| Ethnicity | Actual Mean/SD, (Predicted Mean), (Sample Size), & 95% Confidence Intervals for Mean Predicted Grade 8 Math GEPA Score | |
|---|---|---|
| | Experimental Classes | Control Classes |
| Asian/Pacific Islanders | 207.2/15.87<br>(207.20) (*n* = 5)<br>183.21 – 231.19 | 237.67/34.40<br>(237.67) (*n* = 6)<br>215.76 – 259.57 |
| Black/African American | 190.47/17.99<br>(190.47) (*n* = 57)<br>183.37 – 197.58 | 185.41/27.88<br>(185.41) (*n* = 74)<br>179.17 – 191.64 |
| Hispanic/Latino | 199.18/27.77<br>(199.18) (*n* = 11)<br>183.00 – 215.36 | 197.13/31.41<br>(197.12) (*n* = 24)<br>186.17 – 208.08 |
| White | 202.36/22.82<br>(202.36) (*n* = 74)<br>196.13 – 208.60 | 199.44/31.30<br>(199.44) (*n* = 135)<br>194.82 – 204.05 |

The data suggest that BSI eligibility is a strong predictor of student achievement on the GEPA mathematics test.

The third hypothesis states that there is no difference in achievement on the mathematics section of the New Jersey GEPA between students in the experimental and control groups who are classified in the same ethnic group. Table 3 relates GEPA mathematics test performance for the experimental and control groups to the student's ethnicity. An ANOVA model with an ethnicity-experimental/control group interaction was used to derive predicted 95% confidence intervals for the mean cells in Table 3. The data provide evidence that Asian/Pacific Islanders in the control group, on average, outperformed the other ethnic groups on the GEPA mathematics test and whites, on the average, outperformed the blacks. However, the data do not provide evidence that there was a difference between the performance of the black and the Hispanic/Latino groups. In the experimental group, there was not a statistically significant difference ($p < .05$) in the means of the four ethnic groups in the study. Therefore, we conclude that the data do not provide evidence that any one ethnic group in the experimental population outperformed any other on the GEPA mathematics test. Overall, the data do not provide evidence that the CAI program benefited any ethnic group in the study other than the Asian/Pacific Islander students in the experimental group. Those students scored statistically significantly higher ($p < .05$) than the Asian/Pacific Islander students in the experimental group. An effect size of 0.88 favoring the Asian/Pacific Islander students in the control group was observed.

Table 4 relates GEPA mathematics performance for the experimental and control groups to the student's level of participation in the school's free or reduced-price lunch program. An ANOVA main effects model (no interaction term) was used to derive predicted 95% confidence intervals for the mean cells in Table 4. The fifth hypothesis states that there is no difference in achievement between students in the experimental and control groups based on the level of eligibility for the federal free or reduced-price lunch program.

The data in Table 4 provide evidence that the students in the experimental and control non-subsidized lunch group performed better, on the average, on the GEPA mathematics test than did the students in the free or reduced-price lunch group (see the non-overlapping 95% confidence limits for these groups in Table 4). For example, we observed a statistically significant difference ($p < .05$) in the mean achievement score of students in the experimental group not eligible for free or reduced-price lunch compared to those eligible for free or reduced-price lunch. We observed an effect size of 0.35 favoring students in the experimental group not eligible for free or reduced-price lunch. Likewise, we observed an effect size of 0.56 favoring the students in the control group not eligible for free or reduced-price lunch compared to their group members who were eligible. Overall, the data do not provide evidence that, on average, the CAI program benefited students in any one of the school lunch programs.

Table 5 examines the odds of students passing the GEPA math test as a function of the student's BSI

Table 4
*95% Confidence Intervals for Mean GEPA Score for Free Lunch and Experimental/Control Groups*

| Student's Free Lunch Classification | Actual Mean/SD, (Predicted Mean), (Sample Size), & 95% Confidence Intervals for Mean Predicted Grade 8 Math GEPA Score | |
|---|---|---|
| | Experimental Classes | Control Classes |
| Free Lunch | 191.61/20.67<br>(187.45) (*n* = 26)<br>180.39 – 194.50 | 185.21/27.00<br>(186.85) (*n* = 66)<br>180.92 – 192.79 |
| Reduced-Price Lunch | 199.64/21.22<br>(198.59) (*n* = 14)<br>189.07 – 208.10 | 197.4/28.04<br>(197.99) (*n* = 25)<br>188.98 – 207.00 |
| Non-Subsidized Lunch | 198.90/22.19<br>(200.05) (*n* = 107)<br>195.25 – 204.84 | 200.28/33.02<br>(199.45) (*n* = 148)<br>195.25 – 203.65 |

language/reading service profile and the student's BSI math service profile. A logistic main effects model (no significant experiment/control group effect and no interaction terms) was used to derive predicted probabilities and odds of passing the GEPA math test for each cell in Table 5. The model was also used to derive 95% confidence intervals for the relative odds of passing the GEPA math test. At the .05 significance level, the logistic model found no statistically significant difference between the experimental and control groups regarding the percentage/odds of a student passing the GEPA math test.

More than half, 56.94%, of the students who were not referred to language and/or reading remediation passed the GEPA math test, compared to 26.47% of those who were referred to language/reading remediation. On the average, of those referred neither to language/reading nor math remediation, an estimated 68.19%, passed the GEPA math test. Of those students referred to both language/reading and math remediation only an estimated 3.89% passed GEPA math test.

## Conclusions

In summary, the data suggest that the school under study was successful in identifying a large number of students (110 out of 283 regular students) who required language, reading, and/or math basic skills instruction; however, the remediation program in general, with or without CAI, demonstrated limited success in bringing students up to the level required to pass the GEPA math test.

The drill and practice CAI and student multimedia slide show demonstrations did not have a statistically significant positive influence on student achievement on the GEPA mathematics test. The data suggest that CAI may have had a negative influence on student achievement, as only an estimated 68.19% of those students referred neither to language arts

Table 5
*Actual and Logistic Model Predicted Percent and Odds of Students Passing the GEPA Math Test as a Function of the Student's BSI Language/Reading BSI Math Service Profiles*

| Language or Reading Referral | Math Referral | Actual % (Model Predicted) of Students Passing Math GEPA Test | | Actual (Model Predicted) Odds of a Student Passing the Math GEPA Test | |
|---|---|---|---|---|---|
| | | Experimental | Control | Experimental | Control |
| No | No | 67.57%<br>(68.19%)<br>(*n* = 74) | 68.69%<br>(68.19%)<br>(*n* = 99) | 2.08<br>(2.14) | 2.19<br>(2.14) |
| | Yes | 9.52%<br>(11.69%)<br>(*n* = 21) | 13.64%<br>(11.69%)<br>(*n* = 22) | 0.11<br>(0.13) | 0.16<br>(0.13) |
| | No Language or Reading Referral: | 54.74%<br>(56.94%)<br>(*n* = 95) | 58.68%<br>(56.94%)<br>(*n* = 121) | 1.21<br>(1.32) | 1.42<br>(1.32) |
| Yes | No | 35.29%<br>(39.60%)<br>(*n* = 17) | 42.31%<br>(39.60%)<br>(*n* = 26) | 0.55<br>(0.66) | 0.73<br>(0.66) |
| | Yes | 11.11%<br>(3.89%)<br>(*n* = 9) | 0.0%<br>(3.89%)<br>(*n* = 16) | 0.12<br>(0.04) | 0.0<br>(0.04) |
| | Yes Language and/or Reading Referral: | 26.92%<br>(26.47%)<br>(*n* = 26) | 26.19%<br>(26.47%)<br>(*n* = 42) | 0.39<br>(0.36) | 0.35<br>(0.36) |

nor mathematics BSI passed the GEPA mathematics test. Of students referred to both language arts and mathematics BSI, only an estimated 3.89% passed the GEPA mathematics test.

The CAI drill and practice program was not an effective intervention for increasing achievement on the GEPA. It did not improve the experimental group students' proficiency on the GEPA mathematics test. In two categories, students who received the CAI performed statistically significantly lower than did their peers in the control group. The academically weakest students, those students in the control group who scored in the 25th percentile on the grade 7 TerraNova pretest, outperformed their peers in the experimental group on the GEPA mathematics section. Students in the control group not referred to mathematics BSI remedial instruction outperformed the corresponding group of students in the experimental group.

These findings trouble us for three reasons. First, the teachers used CAI instruction two mathematics periods per week for 20 weeks leading up to the GEPA test. The 90 minutes a week spent on drill and practice CAI may have been better spent on problem solving and critical thinking. Half the points on the GEPA mathematics test come from open-ended problem-solving questions (NJDOE, 2005).

Second, more than 35% of the students in the district participated in BSI mathematics programs. CAI did not influence positively the achievement of the regular education students who struggled academically. In fact, the students in the control group who scored in the lowest quartile of the TerraNova pretest significantly outscored their peers in the experimental group. This suggests that the CAI program may have had a negative influence on some of the district's academically weakest students. The drill and practice CAI used during this study did not have a positive influence on the test scores of low-achieving students compared to similar students in the control group, nor did it influence positively the performance of non-Caucasian students.

Third, the CAI program did not improve the performance of the district's neediest students, those eligible for free or reduced-price lunch. Leaders looking for an intervention to increase the achievement of economically disadvantaged students should take note of the findings presented. In this case, drill and practice CAI was not an effective intervention to overcome the debilitating influence of poverty on student learning.

An ancillary finding included that students enrolled in the BSI programs had the lowest odds of passing the GEPA mathematics section and they demonstrated the lowest scale scores as a group on the test. A universal goal of BSI programs in New Jersey, and in fact, the main focus of the federal Title I program, is to improve student achievement for students eligible for free or reduced-price lunch. Furthermore, section 101 of the NCLB Act (No Child Left Behind [NCLB PL 107-110], 2002) calls for closing the achievement gap between subgroups of students. The basic skills program did not help students in the Title I subgroup achieve proficiency (Note: Only 3.89% of the students requiring language/reading and math BSI services passed the math section of the GEPA test.).

Middle school leaders might be well served to revisit the history of their profession to inform future actions related to restructuring traditional basic skills programs. For example, the recommendations from the Cardinal Principles of Secondary Education (Commission on the Reorganization of Secondary Education, 1918) and the results of the Eight-Year Study (Aikin, 1942) suggested the positive influence of problem-based curriculum and instruction over traditional methods such as drill and practice. Middle level leaders should consider retooling ineffective drill and practice basic skills programs and begin to incorporate problem-based instruction or other types of active learning into their programs and in future uses of CAI.

The school in this study was successful in identifying a large number of students (110 of 284 regular students) who required language arts and/or mathematics BSI; however, the schoolwide BSI program demonstrated limited success in bringing the students in the experimental or control groups up to the level required to attain proficiency on the GEPA mathematics test. While both the students' BSI language arts service profile and the students' BSI mathematics service profile were significant predictors of the odds of the student passing the GEPA mathematics test, the students' mathematics service profile was the more discriminating predictor. The CAI drill and practice was unable to influence positively student performance for those students.

While readers should not generalize the results of this study to general forms of CAI used in other middle schools, the results may prompt middle school leaders to evaluate carefully interventions used to improve student achievement against criteria for success before bringing them into the school environment. Interventions should first and foremost do no harm. Ultimately, they should improve student achievement by using effective and appropriate means to achieve an agreed upon, productive, and ethical end. In education, one desired end is to help develop students who can think critically and solve authentic problems. This study provides further evidence that CAI drill and practice activities void of problem solving will not help students achieve that end.

## References

Aikin, W. M. (1942). *The story of the Eight-Year Study: With conclusions and recommendations*. New York: Harper and Brothers.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Elementary School Journal*, *10*3(1), 51–73.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago: Rand McNally College Publishing.

Campbell, D. L., Peck, D. L., Horn, C. J., & Leigh, R. K. (1987). Comparison of computer-assisted instruction and print drill performance: A research note. *Educational Communication and Technology Journal*, *35*(2), 95–103.

Cerezo, N. (2004). Problem-based learning in the middle school: A research case study of the perceptions of at-risk females. *Research in Middle Level Education Online, 27*(1). Retrieved on June 1, 2007, from http://www.nmsa.org/portals/0/pdf/publications/RMLE/rmle_vol27_no1_article4.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Commission on the Reorganization of Secondary Education. (1918). *Cardinal principles of secondary education*. Washington, DC: U.S. Bureau of Education, Bulletin No. 35.

Cooperstein, S. E., & Kocevar-Weidinger, E. (2004). Beyond active learning: A constructivist approach to learning. *Reference Services Review, 32*(2), 141–148.

CTB/McGraw-Hill. (1997). *TerraNova Technical Bulletin 1*. Monterey, CA: Author.

CTB-McGraw Hill. (2007). Glossary of Assessment Terms. Retrieved on June 1, 2007, from http://www.ctb.com/articles/article_information.jsp?CONTENT%3C%3Ecnt_id=10134198673250329&FOLDER%3C%3Efolder_id=1408474395222381&bmUID=1210949935335

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher 5,* 3–8.

Hetland, L. (2000). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education*, *34*(3/4), 179–238.

Huffaker, D. A., & Calvert, S. L. (2003). The new science of learning: Active learning, metacognition, and transfer of knowledge in e-learning applications. *Journal of Educational Computing Research*, *29*(3), 325–334.

Jersild, A. T., Thorndike, R. L., & Goldman, B. (1941). A further comparison of pupils in “activity” and “non-activity” schools. *Journal of Experimental Education, 9,* 307–309.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86.

National Center for Education Statistics (NCES). (2000). *Highlights from the Third International Mathematics and Science Study—Repeat*. Office of Educational Research and Improvement. U.S. Department of Education: Office of Educational Research and Improvement. Retrieved May 3, 2007, from http://nces.ed.gov/pubs2001/2001027.pdf

National Center for Education Statistics (NCES). (2005). *The Nation’s Reportcard*. U.S. Department of Education: Institute of Education Sciences. Retrieved May 3, 2007, from http://nces.ed.gov/nationsreportcard/pdf/main2005/2006453.pdf

New Jersey Department of Education. (2005). *Grade eight proficiency assessment. Technical report*. #1505.69. Author.

No Child Left Behind (NCLB). (2002). Act of 2001, P.L. No. 107-110, § 115, Stat. 1425.

Organisation for Economic Co-Operation and Development. (2004). *Learning for tomorrow’s world: First results from PISA 2003*. Retrieved May 3, 2007, from http://www.oecd.org/document/55/0,2340,en_32252351_32236173_33917303_1_1_1_1,00.html

Plano, G. (2004). *The effects of the Cognitive Tutor Algebra on student attitudes and achievement in a 9th grade algebra course*. Unpublished doctoral dissertation. Seton Hall University, South Orange, NJ.

Roberts, V. A., & Madhere, S. (1990). *Chapter I resource laboratory program for computer-assisted instruction (CAI) 1989–1990*. Evaluation report. Evaluation report for the District of Columbia Public Schools, Washington, DC.

Rosenberg, R. (1991). Debunking computer literacy. *Technology Review, 94*(1), 58–64.

Southern Education Foundation. (2007). *A new majority: Low income students in the South's public schools.* Retrieved on Nov. 11, 2007, from http://www.sefatl.org/pdf/A%20New%20Majority%20Report-Final.pdf

Tienken, C. H., & Wilson, M. (2007, fall/winter). The impact of computer-assisted instruction on seventh-grade students' mathematics achievement. *Planning and Changing, 38*(3/4), 181–190.

Traynor, P. (2003). The effects of computer-assisted instruction on different learners. *Journal of Instructional Psychology, 30*(2), 137–151.

U.S. Department of Education. (2002). *Proven methods: Scientifically based research*. Retrieved April 14, 2008, from http://www.ed.gov/nclb/methods/whatworks/research/index.html

U.S. Department of Education. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Retrieved May 2, 2008, from http://ies.ed.gov/ncee/pdf/20074005.pdf

Wilson, M., Flanagan, R., Gurkewitz, R., & Skrip, L. (2006, September). *Understanding the effect of origami practice, cognition and language on spatial reasoning.* Paper presented at the Fourth Annual Science, Origami, Mathematics, and Education Conference, at California Institute of Technology, Pasadena, CA.

Wrightstone, J. W. (1935). *Appraisal of newer practices in selected public schools.* New York: Teachers College Press.

## Endnote

[1] The following quotes regarding the documented flaws in the NAEP achievement levels are from the 2002 Executive Summary NAEP Reading Report Card (USDOE, 2003):

> "As provided by law, NCES, upon review of a congressionally mandated evaluation of NAEP, determined that achievement levels are to be used on a trial basis and should be interpreted with caution" (USDOE, p. xi).

> "In 1993, the first of several congressionally mandated evaluations of the achievement level setting process concluded that the procedures used to set the achievement levels were flawed… In response to the evaluation and critiques, NAGB conducted an additional study of the 1992 reading achievement levels before deciding to use them for reporting the 1994 NAEP results. When reviewing the findings of this study, the National Academy of Education (NAE) panel expressed concern about what it saw as a confirmatory bias in the study and about the inability of the study to address the panel's perception that the levels had been set too high" (USDOE, p. 14).

> "First, the potential instability of the levels may interfere with the accurate portrayal of trends… it is noteworthy that when American students performed very well on an international reading assessment, these results were discounted because these results were contradicted by poor performance against the possibly flawed NAEP reading achievement levels in the following year" (USDOE, p. 14).

> "The most recent congressional mandated evaluation conducted by the National Academy of Sciences (NAS) relied on prior studies of achievement levels…The panel (NAS) concluded NAEP's current achievement-level-setting-procedures remain fundamentally flawed. The judgment tasks are difficult and confusing; raters' judgments of different item types are internally inconsistent; appropriate validity evidence for cut scores is lacking, and the process has produced unreasonable results" (USDOE, p. 15).

**Reference**

U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics. (2003). *The nation's reportcard: Reading 2002*. NCES 2003-521, by W. S. Grigg, M. C. Daane, and J. R. Campbell. Washington, DC.