Educational Research
Institute of America
www.eriaonline.com • email: info@eriaonline.com
An affiliate of Beck Evaluation & Testing Associates, Inc.

A Study of the Effectiveness of BIG IDEAS MATH MIDDLE SCHOOL

Big Ideas Learning

Report Number 513

July 2016

Contents

Educational Research Institute of America

Educational Research Institute of America (ERIA) is an educational research and development company that contracts with textbook companies, state departments of education, and school districts to conduct educational program tryout studies, develop instructional materials, analyze tests and test scores, and conduct technical studies for educational assessments. ERIA was founded in 1999 as an outgrowth of Indiana University Center for Research and Development, entitled the Center for Innovation in Assessment. Over the past decade, ERIA has conducted development and research projects for over 20 publishers, state departments of education, and other agencies.

Abstract

The focus of this study was the effectiveness of *Big Ideas Math®*, a mathematics program for middle school students published by Big Ideas Learning. The study included students from ten different schools in five different states. The average percentage of students in the eligible for free/reduced price lunch program is significantly lower than the national average percentage. This is also the case for the average percentage of non-Caucasian students.

The study was conducted with over 3,800 students enrolled in grades 6, 7, and 8. Only those students who took both a pretest and posttest were included in the data analysis. Teachers used the program for their math instruction five days per week and about 45 minutes per day.

Big Ideas Math was the only math program used by these teachers during the study. Pretests and posttests were written by math specialists based on the goals and objectives of the program at each grade level. In addition to analyzing the gain scores for the total group of students at each grade, analyses were conducted separately for higher- and lower- scoring mathematics students. Higher- and lower- scoring students were identified by the students' pretest scores. Those scoring highest on the pretests were designated as the higher-scoring mathematics students and those scoring lowest on the pretests were designated as the lower-scoring math students.

The average gain scores for the total group of students at each grade were statistically significant. In addition, the average gain scores for the higher- and lower- scoring groups at each grade level were also statistically significant. The effect sizes for all students at each grade as well as the higher- and lower-scoring groups were large. The effect sizes, in most cases, exceeded by a large margin the effect sizes needed to determine a substantively important level.

Overview of the Study

Houghton Mifflin Harcourt Publishers contracted with Educational Research Institute of America (ERIA) to conduct a complete academic year study of the effectiveness of the *Big Ideas Math* program for middle school students. The study compared assessments administered to students at the beginning of September 2015 to assessments administered around the middle of June 2016.

Research Questions

The following research questions guided the design of the study and the data analyses:

- Does the implementation of the *Big Ideas Math* Middle School program lead to improved student mathematics achievement?
- Does the *Big Ideas Math* Middle School program lead to differential effects on student achievement as a function of student ability level?

Design of the Study

The design was a quasi-experimental pretest/posttest study of the implementation of the *Big Ideas Math* program for middle school students. The schools used the program during the 2015–2016 academic year. A total of 62 teachers in five different states participated in the study. Teachers reported using the program 5 days a week with an average usage time of about 40 to 45 minutes per day.

Instructional Program

The *Big Ideas Math* Middle School program is described by Big Ideas Learning as follows:

> *The* Big Ideas Math *Middle School program was developed with the Common Core State Standards for Mathematics and uses the Standards for Mathematical Practice as its foundation.* Big Ideas Math *is the only program to offer a Regular Pathway, Compacted Pathway, and Advanced Pathway through middle school mathematics. The program was written by renowned authors Dr. Ron Larson and Dr. Laurie Boswell and features cutting-edge technology to support the curriculum.*
>
> Big Ideas Math *has been systematically developed using learning and instructional theory to ensure the quality of instruction. Students gain a deeper understanding of math concepts by narrowing their focus to fewer topics at each grade level. Students master content through inductive reasoning opportunities, engaging activities that provide deeper understanding, concise stepped-out examples, rich, thought-provoking exercises, and a continual building on what has been previously taught.*

Description of the Assessments

The pretest and posttest used in the study were developed by ERIA mathematics curriculum experts. Tests were developed to match the content of the *Big Ideas Math* program and also to emphasize the National Council of Teachers of Mathematics (NCTM) Standards.

The pretest and posttest used at grades 6, 7 and 8 included a total of 30 multiple-choice items for each test and 10 short-constructed responses. Only the multiple-choice test items were used to analyze the test reliabilities.

Table 1 provides the statistical results for the administration of the posttest for the three posttests used. The KR 20 reliability and the standard error of measurement for the posttest indicates the posttest score results were reliable for arriving at decisions regarding the achievement of the students to whom the tests were administered.

Table 1
Posttest Test Statistics

Test	Reliability*	SEM**
Grade 6 Post-test	.84	2.37
Grade 7 Post-test	.84	2.37
Grade 8 Post-test	.83	2.31

**Reliability computed using the Kuder-Richardson 20 formula.*
*** SEM is the Standard Error of Measurement.*

Test Item Discrimination

In addition to determining the reliability and standard error of measurement of a test, the quality of a test can be evaluated by computing the discrimination of each test item. Test item discrimination is a measure of test validity. Item discrimination focuses on whether test items validly separate students on the basis of their knowledge of the test content.

Test item discrimination can range from -1.0 to +1.0. If the discrimination of a test is above 0, it means that the students who scored higher on the test answered the item correctly more often than students who scored lower on the test. If the discrimination is below 0, it would have a negative discrimination meaning that the students who scored lower on the test answered the question correctly more often than students who scored higher on the test.

All tests will have a range of item discriminations. It would be best, however, if a test had no negative discriminating items and all positive discriminating items were above +.10.[1] However, that is very seldom the case with any test. We can, however, examine a test to see how many "psychometrically" good items there are on a test. The average discrimination of all the items on a test should be above +.15. The highest discriminations are seldom above +.50.

A scale that can be used to evaluate the discrimination of test items and the number of items for each of the two tests used in this study is provided in Table 2. The table shows that the percentages of acceptable, good, or excellent items range from 87% to 90%. The average test item discriminations for all of the tests are excellent.

[1] *Item discrimination is determined by the quality of the test item but also by the effects of instruction and the performance level of students to whom the test is being administered.*

Table 2
Item Discrimination for Posttests

Item Discrimination	Discrimination Values	Grade 6	Grade 7	Grade 8
Below 0	Poor items	0	0	1
+.01 to +.10	Weak items	1	0	0
+.11 to +.20	Acceptable	1	1	1
+.21 to +.30	Good	1	3	2
+.30	Excellent	27	26	26
% of Items Acceptable, Good or Excellent		*90%*	*87%*	*87%*
Average Test Item Discriminations		*.41*	*.42*	*.41*

Description of the Study Sample

Table 3 provides the demographic characteristics of the schools included in the study. It is important to note that the school data does not provide a description of the make-up of the classes that participated in the study. However, the data does provide a general description of the schools and, thereby, an estimate of the make-up of the classes included in the study.

The percentages of students classified as non-Caucasian ranged from 6% to 38% with an average of 14%. By comparison, 49.8% of the students enrolled in U.S. public schools were classified as minority.[2] The sample population shows that in terms of average percentages of non-Caucasian, the study sample was quite a bit lower than the U.S. population percentage in terms of non-Caucasian students.

The percentages of students enrolled in free/reduced lunch programs ranged from 1% to 48% and averaged 28% across the sample of schools. This average percentage was also much lower than the reported national average of 48% for students enrolled in free/reduced lunch programs in public schools.

[2] *The National Center for Educational Statistics* (NCES) reported that for the 2011–2012 school year, 48.1% of public school students were enrolled in free/reduced lunch programs. No free/reduced lunch data were available for the 2012–2013 school year. Also, the NCES reported that for the 2012–2013 school year, 49.8% of public school students were classified as minority (non-Caucasian) students.

Table 3
Demographic Characteristics of the Schools Included in this Study

State	Location	Grades	Enrollment	Non-Caucasian	Free/Reduced Lunch
MA	Suburban	6 to 8	1,091	12%	14%
MI	Suburban	6 to 8	766	6%	24%
MO	Rural	7 to 12	501	9%	43%
WA	Suburban	7 to 8	827	18%	32%
WI	Suburban	8 to 12	812	38%	47%
WI	Suburban	5 to 8	687	8%	16%
WI	Suburban	5 to 8	739	9%	15%
WI	Urban	5 to 8	622	12%	48%
WI	Suburban	6 to 8	644	22%	1%
WI	Urban	6 to 8	686	6%	38%
Average			738	14%	28%

Data Analyses and Results

Standard scores were used for all data analyses. Raw scores were converted to standard scores with a mean of 300 and a standard deviation of 50. Data analyses and descriptive statistics were computed for the students' standard scores.

For most of the comparisons, paired comparison *t*-tests were used to determine if differences in pretest and posttest scores were significantly different. The comparisons were conducted for differences between the *Big Ideas Math* September 2015 pretest and the *Big Ideas Math* June 2016 post-test. The ≤.05 level of significance was used as the level at which differences would be considered statistically significant.

In addition, effect size (Cohen's *d*) was computed for each of the comparisons. This statistic provides an indication of the strength of the effect of the treatment regardless of the statistical significance. The interpretation of Cohen's *d* statistic as guided by the American Institute for Research (AIR) states that, "According to guidelines from the *What Works Clearinghouse*, an effect size of .25 or greater is considered to be 'substantively important'." [3] Beyond the level considered to be substantively important, interpretations of effect sizes in this report include the following guidelines:

.20 to .49 = small

.50 to .79 = medium

.80+ = large

[3] *What Works Clearinghouse Procedures and Standards Handbook* (Version 2.1), page 23, Sept. 2011.

Grade 6 Results

Table 4 shows that the average scores of the 1540 grade 6 students participating in the study increased their average test scores at a statistically-significant level. The effect size was substantively important and was classified as large.

Table 4
Grade 6 Total Group Paired Comparison t-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Pretests	1540	276	35.81	45.721	≤.0001	1.13
Posttests	1540	325	50.17			

The total group of 1540 grade 6 students was divided into two equal-sized groups based on their pretest scores. The 770 students scoring lowest on the pretest were considered to be lower-achieving mathematics students while the 770 scoring highest on the pretest scores were considered to be higher-achieving mathematics students.

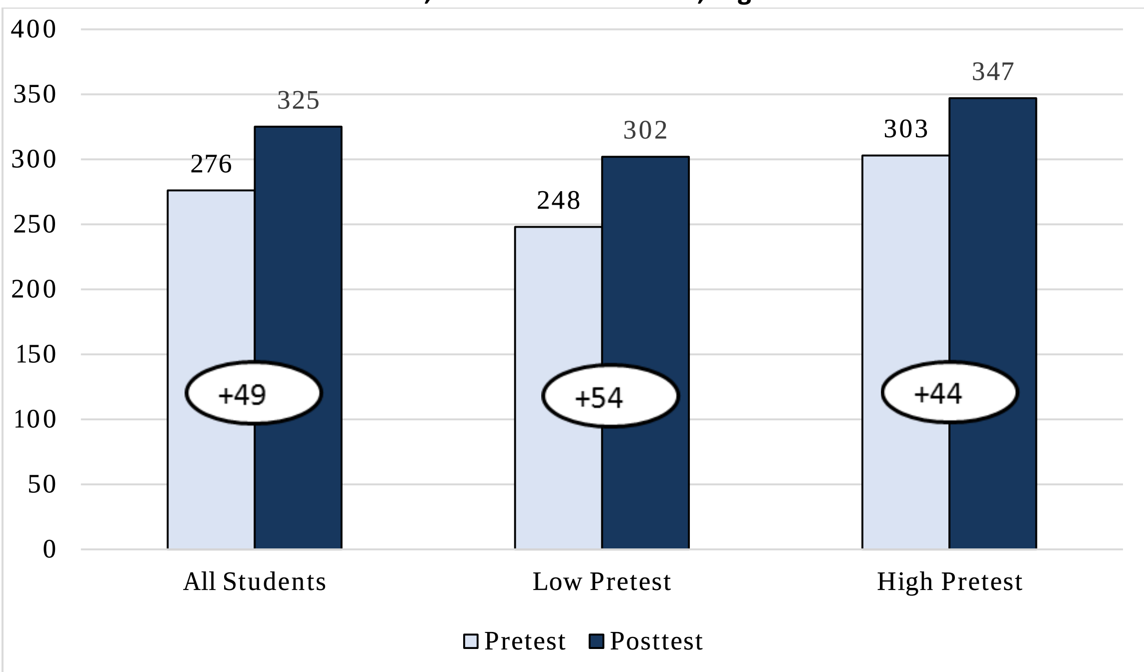
Table 5 shows that both groups made statistically-significant gains. The effect sizes for both groups were substantively important and were classified as large.

Table 5
Grade 6 Paired Comparison t-test Results
Higher- and Lower-Scoring Pretest Groups

Test	*Number of Students*	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Higher-Scoring Group						
Pretest	770	303	27.29	29.309	≤.0001	1.11
Posttest	770	347	48.40			
Lower-Scoring Group						
Pretest	770	248	17.67	35.800	≤.0001	1.71
Posttest	770	302	41.43			

Figure 1 provides a graphic representation of the gains achieved by the grade 6 students. In this full-year study, the grade 6 students increased their average standard scores by 49 points. The lower-achieving mathematics students increased their scores by 54 points while the higher-achieving mathematics students increased their scores by 44 points.

Figure 1
Grade 6 Pretest/Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Grade 7 Results

Table 6 shows that the average scores of the 1282 grade 7 students participating in the study increased their average test scores at a statistically-significant level. The effect size was substantively important and is classified was large.

Table 6
Grade 7 Total Group Paired Comparison t-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	***Mean Standard Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Pretests	1282	275	36.41	44.190	≤.0001	1.16
Posttests	1282	325	49.13			

The total group of 1282 grade 7 students was divided into two equal-sized groups based on their pretest scores. The 641 students scoring lowest on the pretest were considered to be lower-achieving mathematics students while the 641 scoring highest on the pretest scores were considered to be higher-achieving mathematics students.

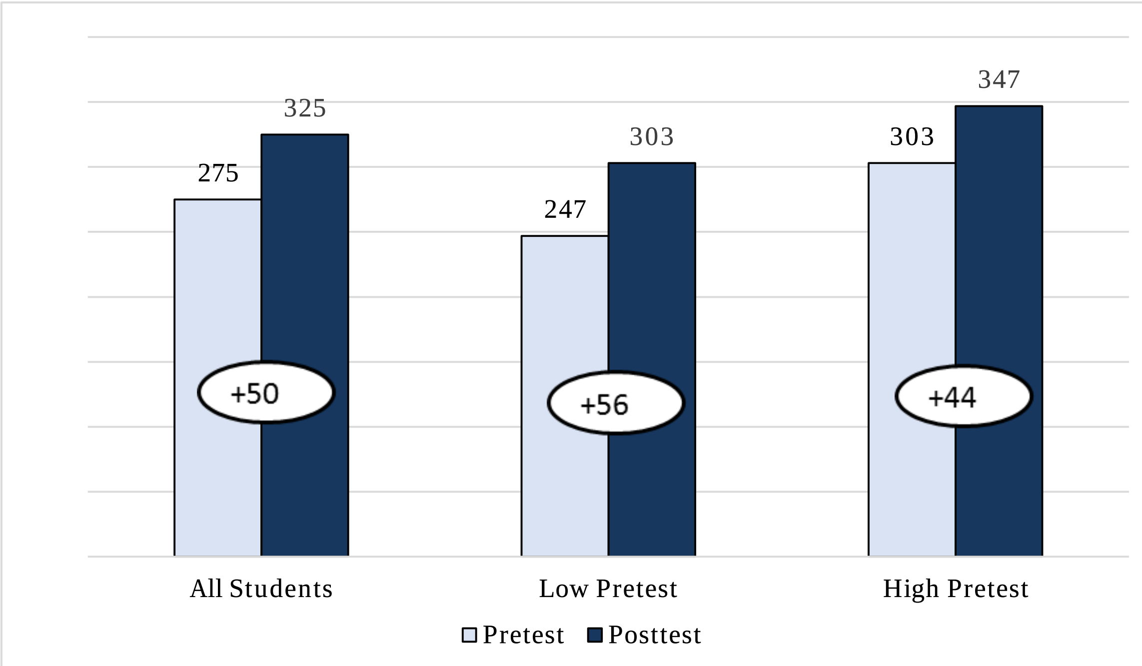
Table 7 shows that both groups made statistically-significant gains. The effect sizes for both groups were substantively important and were classified as large.

Table 7
Grade 7 Paired Comparison t-test Results
Higher- and Lower-Scoring Pretest Groups

Test	***Number of Students***	***Mean Standard Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Higher-Scoring Group						
Pretest	641	303	28.34	29.849	≤.0001	1.17
Posttest	641	347	45.20			
Lower-Scoring Group						
Pretest	641	247	16.4	33.180	≤.0001	1.75
Posttest	641	303	42.33			

Figure 2 provides a graphic representation of the gains achieved by the grade 7 students. In this full-year study, the grade 7 students increased their average standard scores by 50 points. The lower-achieving mathematics students increased their scores by 56 points while the higher-achieving mathematics students increased their scores 44 points.

Figure 2
Grade 7 Pretest/ Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Grade 8 Results

Table 8 shows that the average scores of the 1043 grade 8 students participating in the study increased their average test scores at a statistically-significant level. The effect size was substantively important and was classified as large.

Table 8
Grade 8 Total Group Paired Comparison t-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Pretests	1043	269	29.60	44.547	≤.0001	1.58
Posttests	1043	331	46.87			

The total group of 1043 grade 8 students was divided into two approximately equal-sized groups based on their pretest scores. The 521 students scoring lowest on the pretest were considered to be lower-achieving mathematics students while the 522 scoring highest on the pretest scores were considered to be higher-achieving mathematics students.

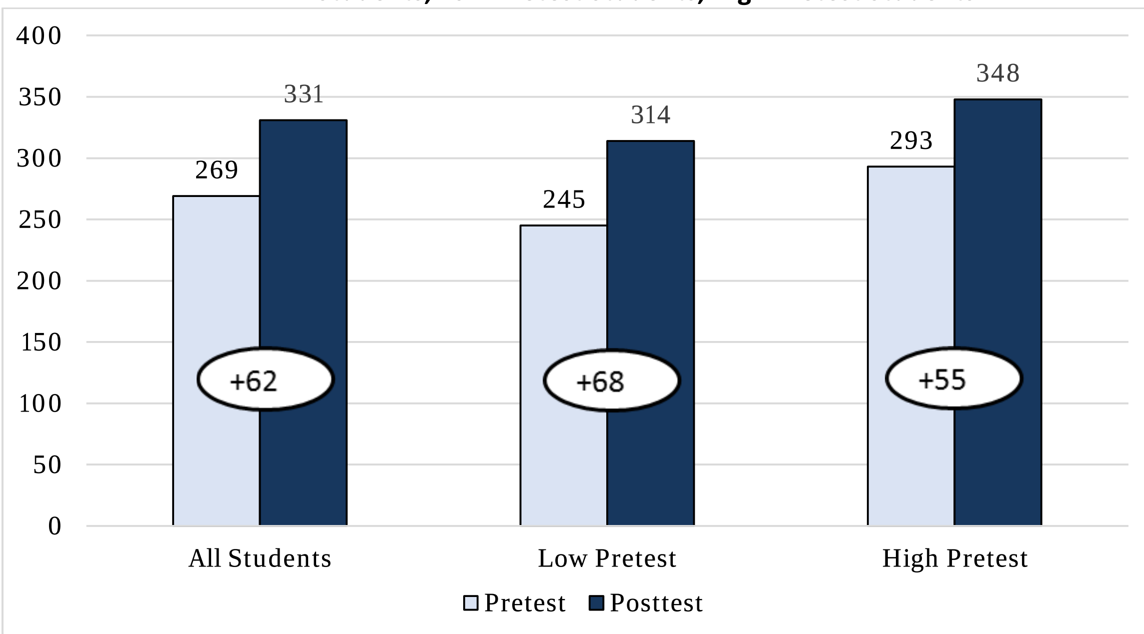
Table 9 shows that both groups made statistically-significant gains. The effect sizes for both groups were substantively important and were classified as large.

Table 9
Grade 8 Paired Comparison t-test Results
Higher- and Lower-Scoring Pretest Groups

Test	*Number of Students*	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Higher-Scoring Group						
Pretest	522	293	18.66	31.832	≤.0001	1.74
Posttest	522	348	40.66			
Lower-Scoring Group						
Pretest	521	245	16.75	32.241	≤.0001	1.96
Posttest	521	314	46.79			

Figure 3 provides a graphic representation of the gains achieved by the grade 8 students. The grade 8 students increased their average standard scores by 62 points. The lower-achieving mathematics students increased their scores by 68 standard score points while the higher-achieving mathematics students increased their scores by 55 points.

Figure 3
Grade 8 Pretest /Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Conclusions

This study sought to determine the effectiveness of the *Big Ideas Math* Middle School mathematics program by comparing growth on reliable and valid pretests and posttests. The study took place during the 2015-2016 academic year and was carried out in five states including ten different schools. The average demographic characteristics of the students in the sample showed there were about 20% fewer students in the sample who were eligible for free/reduced lunch programs than there were reported in the national statistics. The percentage of non-Caucasian students was about 35% lower than the national average.

Two research questions guided the study and the conclusions for each are reported below.

Research Question 1

- Does the implementation of the *Big Ideas Math* Middle School program lead to improved student mathematics achievement?

Across all three grades, mathematics student growth was statistically significant. The effect sizes at all three grades were above a substantively important level and were large at all three grade levels.

Research Question 2

- Does the *Big Ideas Math* Middle School program lead to differential effects on student achievement as a function of student ability level?

Across all three grades, mathematics student growth for the higher-achieving and lower- achieving students was statistically significant. The effect sizes at all three grades for both the high- and low-group students were above a substantively important level and were large at all three grade levels.

On the basis of this study, both research questions can be answered positively:

With the *Big Ideas Math* Middle School program students in grade 6, 7, and 8 showed significant growth and the effect sizes were very large.

The *Big Ideas Math* program for middle school students showed significant growth for both higher-ability and lower-ability students in grades 6, 7, and 8. The effect sizes for both groups at grades 6, 7, and 8 were very large.