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 ALGEBRA 1

Big Ideas Learning

Report Number 514

July 2016

Table of Contents

Abstract

The focus of this study was the effectiveness of *Big Ideas Math*® Algebra 1, an Algebra program for secondary school students published by Big Ideas Learning. The study included students from eight different schools in five different states. A total of 15 different teachers participated in this full academic-year study.

The average percentage of students in the eligible for free/reduced price lunch program was about 25% lower than the national average percentage. The percentage of non-Caucasian students was about 30% lower than the national average.

The study was conducted with over 600 students enrolled primarily in grade 9, but some grade 8 and grade 10 students were also enrolled in some of the Algebra classes. 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 for a minimum of 35 minutes per day.

The *Big Ideas Math* Algebra 1 program 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. 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 algebra 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 algebra students and those scoring lowest on the pretests were designated as the lower-scoring algebra students.

The average gain scores for the total group of students were statistically significant. In addition, the average gain scores for the higher- and lower-scoring groups were also statistically significant. The effect sizes for all students as well as the higher- and lower- scoring groups were large. The effect sizes 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 an academic year study of the effectiveness of the *Big Ideas Math* High School Algebra 1 program. 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 Algebra 1* program lead to improved student mathematics achievement?
- Does the *Big Ideas Math Algebra 1* 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 Algebra 1* secondary school students. The schools used the program during the 2015–2016 academic year.

A total of 15 teachers in eight different schools and five different states participated in the study. The classes included primarily grade 9 students, but students in grade 8 and grade 10 were included in some of the classes. Teachers used the program for their math instruction five days per week for a minimum of 35 minutes per day.

Instructional Program

The *Big Ideas Math Algebra 1* program is described by Big Ideas Learning as follows:

> Big Ideas Math Algebra 1 *was written by Ron Larson and Laurie Boswell. The* Big Ideas Math Algebra 1 *Student Edition mirrors the pedagogical philosophy of the authors. Each lesson begins with an Essential Question, followed by Explorations. Each Exploration is followed by a scaffolded instruction Lesson. These lessons give students the opportunity to develop procedural fluency and to use clear, precise mathematical language. These lessons also give teachers opportunities to use class discussion, flexible grouping, and other delivery methods* in their classrooms.

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 Algebra 1* program.

The pretest and posttest included a total of 22 multiple-choice items and 22 open-ended response questions. Only the multiple-choice test items were used to analyze the test reliabilities.

Table 1 provides the statistical results for the administration of the pretest and the posttest. The KR 20 reliability and the standard error of measurement for the two tests indicates that the pretest and

posttest score results were reliable for arriving at decisions regarding the achievement of the students to whom the tests were administered.

Table 1
Pretest/Posttest Test Statistics

Test	Reliability*	S.D.	SEM**
Algebra 1 Pretest	.76	4.08	2.00
Algebra 1 Posttest	.82	4.38	1.86

**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 were 82% for the pretest and 95% for the post-test. The average test-item discriminations across the items for each test were quite high.

[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	Algebra 1 Pretest	Algebra 1 Posttest
Below 0	Poor items	0	1
+.01 to +.10	Weak items	0	0
+.11 to +.20	Acceptable	2	0
+.21 to +.30	Good	2	O
+.30	Excellent	18	21
% of Items Acceptable, Good or Excellent		*82%*	*95%*
Average Test Item Discriminations		*.40*	*.45*

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 4% to 39% with an average of 20%. By comparison, 49.8% of the students enrolled in U.S. public schools were classified as minority.[2]

The percentages of students enrolled in free/reduced lunch programs ranged from 7% to 57% and averaged 26% 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.

Table 3
Demographic Characteristics of the Schools Included in this Study

State	Location	Grades	Enrollment	Non-Caucasian	Free/Reduced Lunch
FL	Rural	9 to 12	337	17%	57%
IA	Rural	9 to 12	146	4%	18%
IA	Urban	7 to 8	555	17%	47%
IL	Suburban	7 to 8	749	19%	12%
IL	Suburban	6 to 8	548	37%	7%
IL	Suburban	6 to 8	639	39%	13%
NJ	Suburban	9 to 12	843	19%	7%
WI	Rural	9 to 12	845	7%	39%
Average			618	20%	26%

[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.

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 all three 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 Algebra 1* September 2015 pretest and the *Big Ideas Math Algebra 1* June 2016 posttest. 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

Pretest/Posttest Comparison

Table 4 shows that the average scores of the 638 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
Total Group Paired Comparison t-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	***Mean Standard Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Pretests	638	274	33.6	32.513	≤.0001	1.25
Posttests	638	327	49.7			

The total group of 638 students was divided into two equal-sized groups based on their pretest scores. The 319 students scoring lowest on the pretest were considered to be lower-achieving mathematics students while the 319 scoring highest on the pretest scores were considered to be higher-achieving mathematics students.

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

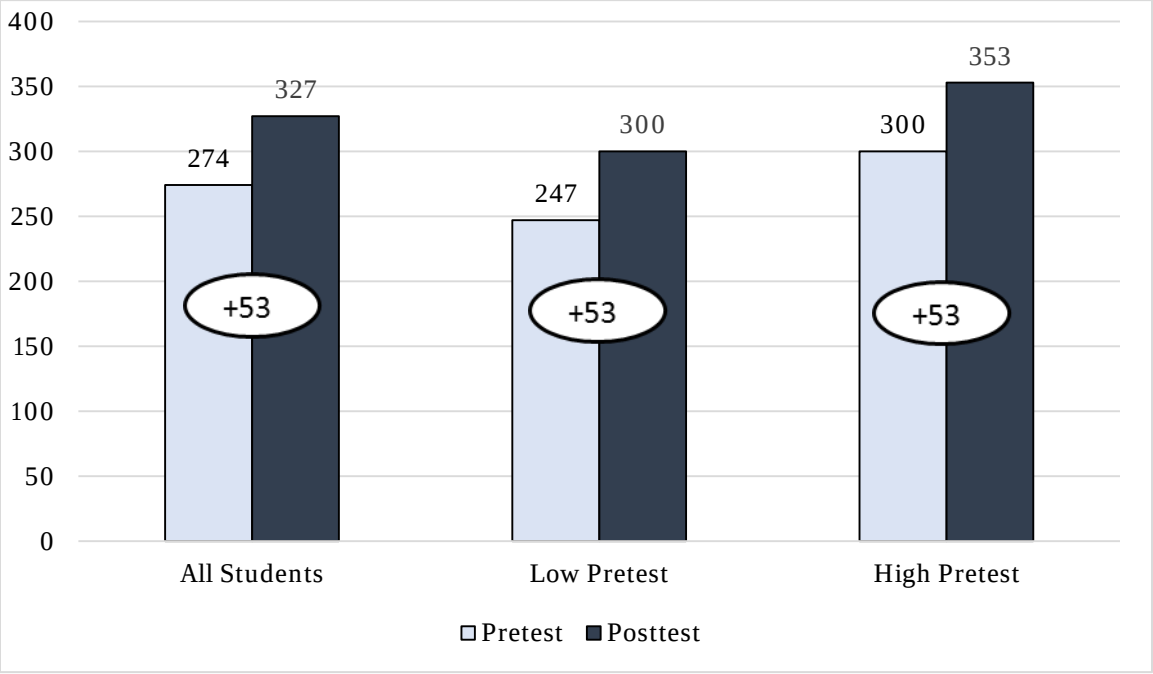
Table 5 shows that both groups made statistically significant gains. The effect sizes for both groups were substantively important and were classified as large.

Table 5
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	319	300	23.2	25.256	≤.0001	1.56
Posttest	319	353	41.8			
Lower-Scoring Group						
Pretest	319	247	18.0	21.250	≤.0001	1.62
Posttest	319	300	42.8			

Figure 1 provides a graphic representation of the gains achieved by the students. In this full-year study, the students increased their average standard scores by 53 points. The higher-and lower-achieving mathematics students increased their scores by the same number of points as the total group.

Figure 1
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 Algebra 1* program by comparing growth on reliable and valid pretests and posttests administered at the beginning and end of the 2015-2016 academic year. The study was carried out in five states and included eight different schools and 15 different teachers. The average demographic characteristics of the students in the sample showed there were about 30% 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 25% 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 Algebra 1* program lead to improved student mathematics achievement?

Student growth was statistically significant, and the effect size was above a substantively important level and was quite large.

Research Question 2

- Does the *Big Ideas Math Algebra 1* program lead to differential effects on student achievement as a function of student ability level?

Algebra score increases were statistically significant for both the higher-achieving and lower-achieving students. The effect sizes for both the higher- and lower-group students were above a substantively important level and were quite large for both groups.

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

The *Big Ideas Math Algebra 1* program for secondary students, showed scores resulted in statistically-significant growth and the effect size was very large.

The Big Ideas Math Algebra 1 program for secondary students showed significant growth for both higher-ability and lower-ability students. The effect sizes for both groups were very large.