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

Efficacy Study for Houghton Mifflin Harcourt Mathematics:

Algebra 1, *Geometry*, and *Algebra 2*

Houghton Mifflin Harcourt

Report Number 508
June 2015

Table of Contents

Abstract

The focus of this study was the effectiveness of the Houghton Mifflin Harcourt high school math program which includes Algebra 1, Geometry, and Algebra 2 © 2015. The study included students from 4 different schools in 3 different states. The percentage of students in the study eligible for free/reduced lunch programs is similar to the percentage of students eligible for free/reduced lunch programs for all public schools in the United States. The percentage of non-Caucasian students was much lower than the percentage of non-Caucasian students enrolled in public schools in the United States.

The study was conducted during the second semester of the 2014-2015 academic year. The teachers in the study had used the program during the first semester and they agreed to engage in a study of the program during the second semester. The teachers identified the program modules they would be using during the second semester. Pretest and post-test assessments were developed for the modules taught during the second semester.

Approximately 450 students were included in the study. Teachers used the program as their primary curriculum in either Algebra 1, Geometry or Algebra 2 for the entire second semester. Teachers reported this was the first year of use of the program. The first semester provided an opportunity for the teachers to become familiar with the program and the second semester was the study semester. All of the teachers had at least five years of teaching experience and most had 10 to 15 years of teaching experience.

Pretests and posttests were developed by mathematics curriculum test writers and were reviewed by other mathematics curriculum experts to determine if the items matched the material being taught. In addition to analyzing the gain scores for the total group of students for each of the three subjects, analyses were conducted separately for higher and lower mathematics achieving students. Higher and lower achieving 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 mathematics students.

The average gain scores for the total group of students for each of the three subjects were statistically significant. The effect size for the Algebra 1 students was large. For Geometry and Algebra 2, the effect sizes were medium.

In addition, the average gain scores for the low and high scoring groups for each of the three subject followed the same pattern. The effect sizes for the lower achieving mathematics students were large and the effect sizes for the higher achieving mathematics students were large. *While all groups made statistically significant gains, the lower achieving students made the largest gains.*

Overview of the Study

Houghton Mifflin Harcourt School Publishers contracted with Educational Research Institute of America (ERIA) to conduct a one semester study to evaluate the effectiveness of the high school mathematics programs which included Algebra1, Geometry, and Algebra 2. The study compared pretests which were administered to students the middle of January 2015 and the post-tests were administered the end of May 2015.

Research Questions

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

- Does the implementation of the *AGA Mathematics programs* lead to improvement of student's mathematics achievement?
- Does the implementation of the *AGA Mathematics programs* lead to improvement of mathematics achievement for both low achieving mathematics students as well as high achieving students?

Design of the Study

The design of the program called for the implementation of the AGA mathematics program for students enrolled in Algebra 1, Geometry, or Algebra 2 during the second semester of the 2014–2015 academic year. The schools had been using the program during the first semester and reported that they had not used the program prior to that time. The study assessments only covered the content taught during the second semester.

A total of 9 teachers in 3 different states and 4 different schools participated in the study. The number of teachers for each subject included:

- Algebra 1: 4 schools, 3 states, 5 teachers
- Geometry: 2 schools; 2 states; 3 teachers
- Algebra 2: 3 schools, 3states, 4 teachers

Teachers reported using the program 5 days a week with an average usage time of more than 25 minutes. The majority of the teachers had been teaching 6 to 10 years.

Program Overview

The following information taken from the Houghton Mifflin Harcourt web site provides a basic description of the AGA mathematics program.

> *No other high school math curriculum empowers students to develop the core skills they need like Houghton Miffilin Harcourt Algebra 1, Geometry, and Algebra 2. In keeping with the Common Core State Standards, the new, streamlined Student Editions focus on deeper understanding of math strategies and concepts, with implementation of the Standards for Mathematical Practice. Combination of print texts, eTexts, and robust Interactive Online Editions is ideal for high school math courses.*

Description of the Assessments

The pretests and post-tests used in the study were developed by ERIA mathematics curriculum test development writers. Tests were developed to match the content of the AGA Mathematics program as well as to emphasize the Common Core State Standards (CCSS).

The assessments were developed to assess the content of the modules chosen by the teachers for instruction during the second semester. All test items were multiple choice format and matched the format of the instructional tests used with the program as well as matching the formats of many of the CCSS assessments. The Algebra 1 and Algebra 2 tests included a total of 40 items for both the pretests and the post-tests. The Geometry assessments included a total of 35 pretest and post-test items.

The modules were chosen for instruction by the teachers and the assessments were based on those instructional modules.

- The Algebra 1 items focused on modules 7 to 18.
- The Algebra 2 items focused on modules 4 to 7.
- The Geometry assessments focused on modules 4 to 12.

Test Reliability and Standard Error of Measurement

Table 1 provides the test statistics. The table shows that the reliabilities of the post-tests are high and provide adequate stability to assess mathematics achievement. Of particular importance is the fact that the test reliabilities are higher for the post-tests than for the pretests. This is almost certainly the result of instruction which would result in less random guessing on the post-tests than on the pretests. Geometry does not show the same increase; however, both the pretest and the post-test reliabilities for Geometry are high. In addition, the reliabilities for all 3 of the tests are adequate for assessing gain scores.

Table 1
Pretest and Posttest Statistics for the *Algebra 1, Algebra 2, Geometry* Students

Test	*Mean Score*	*Standard Deviation*	*KR 20*	*SEm**
Algebra 1 Pretest	279	43.5	.77	24.6
Algebra 1 Post-test	321	47.3	.81	25.9
Algebra 2 Pretest	283	41.4	.54	28.1
Algebra 2 Post-test	317	52.1	.70	28.5
Geometry Pretest	282	50.3	.85	19.5
Geometry Post-test	318	42.5	.81	18.5

**SEm stands for 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..

The calculation of 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 good items there are on a test.

A scale that can be used to evaluate the discrimination of test items and the number of items for each of the three tests used in this study is provided in Table 2. The table shows that all three of the tests have a large percentage of acceptable, good or excellent test items Algebra 1 (85%), Algebra 2 (85%) and Geometry (94%). The average test item discrimination for Algebra 1 is excellent. For Algebra 2 the average is good and for Geometry the average is excellent.

Table 2
Test Item Discrimination for AGA Post-test Assessments

		Number of Test Items in each Category		
Item Discrimination Values	**Interpretation of Discrimination Values**	**Algebra 1 (Total 40 Items)**	**Algebra 2 (Total 40 Items)**	**Geometry (Total 35 items)**
Below 0	Poor test items (should be replaced)	3 items	3 items	1 item
+.01 to +.10	Weak test items (revise items)	3 items	3 items	1 item
+.11 to +.20	Acceptable	7 test items	6 test items	0 items
+.21 to +.30	Good items	4 test items	4 test items	9 items
+.30 and above	Excellent test items	23 test items	24 test items	24 test items
Average Discrimination		+.34	+.27	+.36

[1] It should be noted that item discrimination is based not just on the quality of the test item but also on the effects of instruction and the performance level of students to whom the test is being administered.

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 percentage of students classified as minority students (non-Caucasian) ranged from 3% to 49% with an average of 15%. By comparison, approximately 50% of the students enrolled in U.S. public schools were classified as non-Caucasian.[2]

The percentage of students enrolled in free/reduced lunch programs ranged from 0% to 51% and averaged 33% across the sample of schools. By comparison, the reported national average for students enrolled in free/reduced lunch programs in public schools was reported as 48%.[2]

Table 3
Demographic Description of the Schools Included in the Study

	State	Location	Grades	Course(s) Taught	Enrollment	% Minority	% Free/Reduced Lunch
1	OH	Rural	9-12	AGA[3]	544	3%	0
2	OH	Rural	7-12	AGA	402	6%	34%
3	MO	Rural	9-12	AGA	356	3%	45%
4	NY	Urban	9-12	A-1[4]	1,252	49%	51%
AVERAGES					639	15%	33%

[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.

[3] AGA indicates the school included students enrolled in Algebra 1, Geometry, and Algebra 2.

[4] A-1 indicates the school included only students enrolled in Algebra 1.

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 using students' standard scores.

For most of the comparisons, paired comparison *t*-tests were used to determine if differences in pretest and post test scores were significantly different. The comparisons were conducted for differences between the pretests administered in mid-January 2015 and the post-tests administered at the end of May 2015. 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 effect sizes in this report use the following guidelines:

.20 to .49 = small
.50 to .79 = medium
.80+ = large

Algebra 1 Results

Table 4 shows that the average scores of the 199 students participating in the study increased at a statistical significant level. The effect size was large.

Table 4
Algebra 1 Paired Comparison *t*-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	***Mean Standard Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Pretests	199	279	43.5	12.910	≤.0001	.92
Post-tests	199	321	47.3			

The total group of 199 Algebra 1 students was divided into two approximately equal sized groups based on their pretest scores. The 100 students scoring lowest on the pretest were considered to be lower Algebra achieving students while the 99 students scoring highest on the pretest scores were considered to be the higher Algebra achieving students.

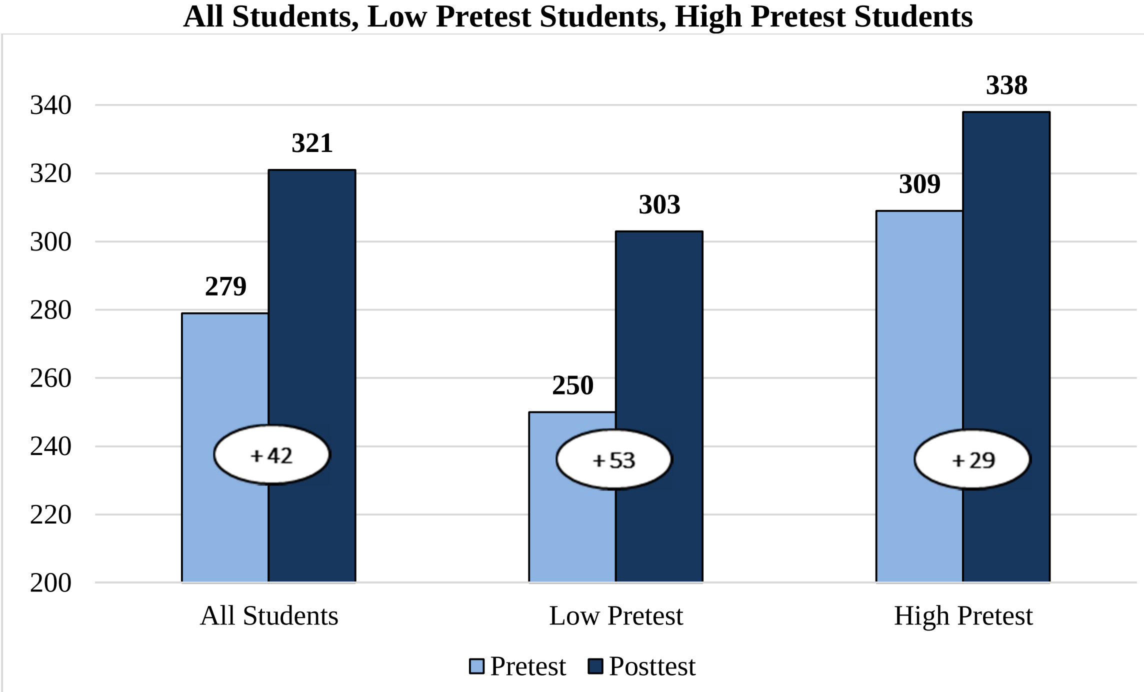
Table 5 shows that both groups made statistically significant gains. The effect sizes for both large.

Table 5
Algebra 1 Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

Test	*Number of Students*	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Lower Scoring Group						
Pretest	100	250	28.4	10.741	≤.0001	1.33
Posttest	100	303	48.9			
Higher Scoring Group						
Pretest	99	309	35.1	7.945	≤.0001	.80
Posttest	99	338	38.3			

Figure 1 provides a graphic representation of the gains achieved by the Algebra 1 students. In one semester, the Algebra 1 students increased their average standard scores by 42 points. The low achieving mathematics students increased their average standard scores by 53 points which was about 50% higher than the increase of the high achieving students whose average standard scores increased by 29 points.

Figure 1
Algebra 1 Pretest Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Algebra 2 Results

Table 6 shows that the average scores of the 150 students participating in the study increased at a statistical significant level. The effect size was substantively important and is classified as medium.

Algebra 2 Total Group Paired Comparison *t*-test Results
Pretest/Posttest Standard Score Comparisons

	Number Students	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Pretests	150	283	41.4	8.453	≤.0001	.72
Post-tests	150	317	52.1			

The total group of 150 Algebra 2 students was divided into two equal sized groups based on their pretest scores. The 75 students scoring lowest on the pretest were considered to be lower Algebra 2 achieving students while the 75 students scoring highest on the pretest scores were considered to be the higher Algebra 2 achieving students.

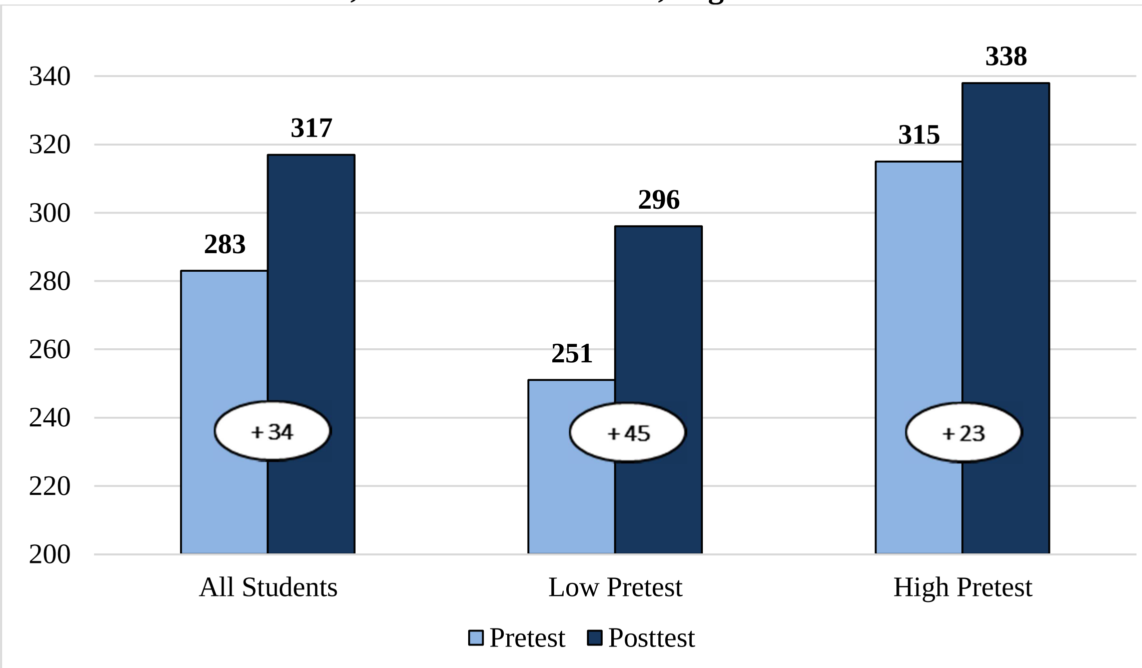
Table 7shows that both groups made statistically significant gains. The effect sizes for both groups were substantively important and are classified as large.

Table 7
Algebra 2 Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

Test	*Number of Students*	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Lower Scoring Group						
Pretest	75	251	24.2	8.134	≤.0001	1.08
Posttest	75	296	43.5			
Higher Scoring Group						
Pretest	75	315	27.4	4.092	≤.0001	.55
Posttest	75	338	51.9			

Figure 2 provides a graphic representation of the gains achieved by the Algebra 2 students. In one semester, the algebra 2 students increased their average standard scores by 34 points. The low achieving mathematics students increased their average standard scores by 45 points which was about 50% higher than the increase of the high achieving students whose average standard scores increased by 23 points.

Figure 2
Algebra 2 Pretest Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Geometry Results

Table 8 shows that the average scores of the 182 students participating in the study increased at a statistical significant level. The effect size was substantively important and is classified as large.

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

	Number Students	*Mean Standard Score*	*SD*	*t*-test	*Significance*	*Effect Size*
Pretests	182	282	50.3	13.166	≤.0001	.77
Post-tests	182	318	42.5			

The total group of 182 Geometry students was divided into two approximately equal sized groups based on their pretest scores. The 91 students scoring lowest on the pretest were considered to be lower Geometry achieving students while the 91 students scoring highest on the pretest scores were considered to be the higher Geometry achieving students.

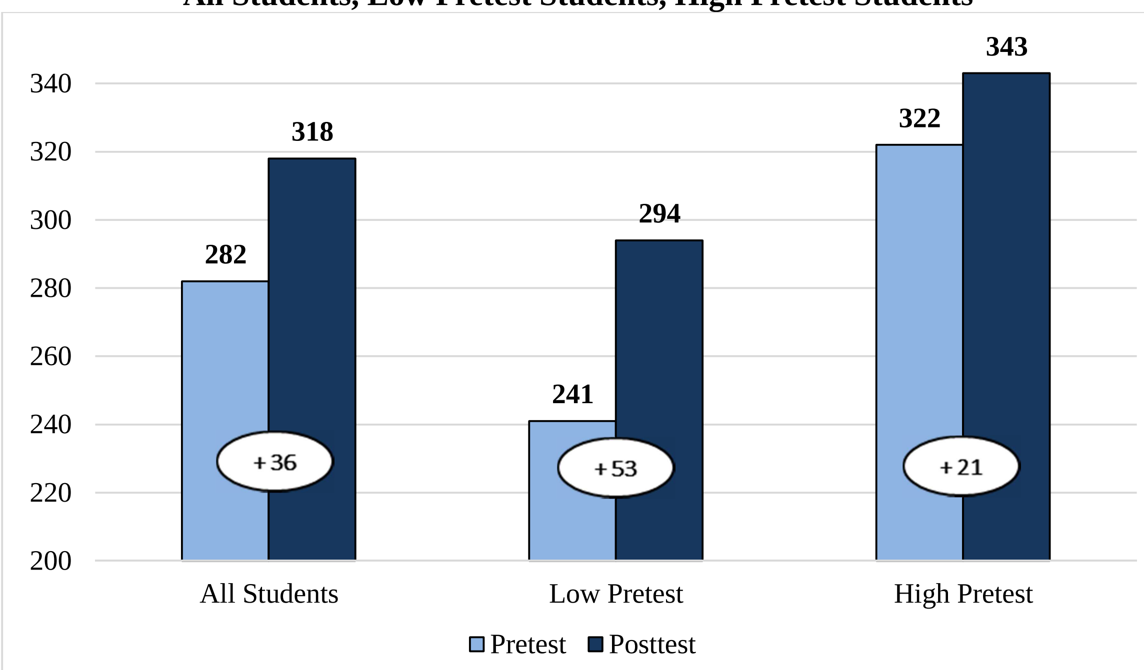
Table 9 shows that both groups made statistically significant gains. The effect sizes for both groups were substantively important and are classified as large.

Table 9
Geometry Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

Test	*Number of Students*	*Mean Standard Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Lower Scoring Group						
Pretest	91	241	20.4	12.944	≤.0001	1.87
Posttest	91	294	34.5			
Higher Scoring Group						
Pretest	91	322	36.9	6.953	≤.0001	.58
Posttest	91	343	35.6			

Figure 3 provides a graphic representation of the gains achieved by the Geometry students. In one semester, the Geometry students increased their average standard scores by 36 points. The low achieving mathematics students increased their average standard scores by 53 points which was more than 50% higher than the increase of the high achieving students whose average standard scores increased by 21 points.

Figure 3
Geometry Pretest Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Conclusions

This study sought to determine the effectiveness of the AGA mathematics program for Algebra 1, Geometry, and Algebra 2 comparing growth on reliable and valid pretests and posttests. The study took place during the second semester of the 2014-2015 academic year and was carried out in 3 states and included 4 different schools and 9 teachers. The student population included a slightly larger percentage of students eligible for free-reduced price lunch programs than the national average. The percentage of non-Caucasian student was about 35% lower than the national average.

Analyses of the mathematics assessments used for all three subjects indicated the tests were reliable and demonstrated that the test items were good at discriminating between those students who scored high on each test and those who scored low on each test.

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

Research Question 1

- Does the implementation of the *AGA Mathematics programs* lead to improved mathematics achievement?

For all three courses, Algebra 1, Geometry, and Algebra 2, included in the study mathematics achievement growth from pretesting to post-testing was statistically significant. The effect size for Algebra 1 was large. For Geometry and Algebra 2 the effect size was medium.

Research Question 2

- Does the implementation of the *AGA Mathematics programs* lead to improvement of mathematics achievement for both low achieving mathematics students as well as high achieving students?

For all three courses, Algebra 1, Geometry, and Algebra 2, included in the study mathematics achievement increased statistically significantly for both the high achieving and low achieving students. The effect sizes for the lower achieving students in all three programs was large. The effect size for the lower achieving Algebra 1 program was also large. For Algebra 2 and Geometry the effect sizes for the higher achieving students was medium.

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

The AGA Mathematics programs resulted in statistically significant increases for students enrolled in Algebra 1, Geometry, and Algebra 2. The effect sizes were either large or medium.

The AGA Mathematics programs resulted in statistically significant growth for both higher ability and lower ability students for all three subjects. The effect sizes for the lower achieving mathematics students was large for all 3 subjects. The effect sizes for the higher achieving students in Algebra were large and for Geometry and Algebra 2, the effect sizes for the higher achieving students were medium.