An Efficacy Study of Escalate English

Published by Houghton Mifflin Harcourt

Report Number 527
June 2017

Table of Contents

Abstract

To help school students become more effective users of language, *Houghton Mifflin Harcourt* has published, *Escalate English © 2017* for students in grades 4 to 8*. Escalate English* was created to aid those students who may become disabled in classroom academic work due to language difficulty. Escalate English provides focused, inviting, and intellectually challenging activities to encourage student and accelerate their language development.

To evaluate the program's effectiveness, *Houghton Mifflin Harcourt* contracted with the *Educational Research Institute of America* (ERIA) to conduct a full school year study to test the effectiveness of the program. The study was conducted with students in grades 6 to 8 during the 2016-2017 academic year.

Pretest and post-test assessments were developed to assess student growth in mastering the program objectives. The assessments were focused on Vocabulary, How English Works, and Reading Comprehension.

 The results were very positive demonstrating significant growth for students at all three grade levels over the course of the full year study. The increases at all three grades were statistically significant and the effect sizes were medium at grade 6 and large at grades 7 and 8. The results also showed that the *Houghton Mifflin Harcourt Escalate English* program proved effective with both higher and lower pretest scoring students. Both groups of students increased their average scores statistically significantly. The effect sizes at grade 6 were large for the lower pretest scoring students and medium for the higher pretest scoring students. For grade 7 and 8 students, the effect sizes for both the lower and the higher pretest scoring students were large.

Overview of the Study

This report describes a 2016-2017 academic year study with students in grades 6 to 8 to determine the effectiveness of the **Houghton Mifflin Harcourt Escalate English © 2017** program for students in grades 4 to 8. **Escalate English © 2017** focuses on academic relevant and rigorous standards in language arts.

Escalate English provides demanding yet accessible content in a digital and print environment. The program is planned to surround students in rich, topically connected, language activities that are cognitively and linguistically demanding.

To determine the program's effectiveness, *Houghton Mifflin Harcourt* contracted with the *Educational Research Institute of America* (ERIA) to conduct a full year study of the program during the 2016/2017 academic year. **Escalate English** was the primary language instructional program in the tryout classes.

The program is described by the publisher on the Houghton Mifflin Harcourt web site as follows:

> **Escalate English** © 2017 is a language development program for Grades 4–8 designed to help students realize their potential by rapidly increasing their language proficiency and mastery of academic English. Created specifically for students who are—or are at risk of becoming—long-term English learners, **Escalate English** provides language-rich, intellectually challenging experiences to motivate students and accelerate their growth.

Research Questions

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

1. Is **Houghton Mifflin Harcourt Escalate English** effective in increasing the skills and knowledge of grade 6 to 8 students so they can meet the increasing language demands in K-12 education?

2. Is **Houghton Mifflin Harcourt Escalate English** effective in increasing the skill and knowledge of grade 6 to 8 students for both higher pretest scoring and lower pretest scoring students so they can meet the increasing language demands in K-12 education?

Design of the Study

The program's efficacy was evaluated using a pretest/post-test design. The study took place during the 2016/2017 academic year in seven different schools across three states. The number of teachers included:

Grade 6   6 different teachers
Grade 7   7 different teachers
Grade 8   7 different teachers

Pre-tests and post-tests were administered at the beginning and end of the school year. The tests modeled the assessments developed for the *Escalate English* program. The tests carefully matched the standards that were the focus of the instructional program. Pretest and post-test administration was under the direction of the classroom teacher. All tests were returned to ERIA for scoring and analyses.

Timeline and Program Use

The teachers used the ***Houghton Mifflin Harcourt Escalate English*** text as their primary instructional program. The teachers reported using the program an average of 3 days per week and for an average of about 35 minutes per day over the entire academic year. Pretests were administered about the middle of September, 2016 and post-tests were administered about the middle of June, 2017.

Description of the Research Sample

Table 1 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 school and, thereby, an estimate of the make-up of the classes included in the study.

Sixty-one percent of the students enrolled in the 7 schools were minority students. Fifty-six percent of the students in the schools were enrolled in National School Lunch Programs. The National Center for Educational Statistics reports that minority enrollment in public elementary and secondary schools at 50.5% and 48% for National School Lunch Programs.

Table 1
Schools Included in the Study: Demographic Characteristics

School	State	Location	Grades	Enrollment	% Minority	% NSLP*
1	CO	City	6 to 8	756	77%	77%
2	CO	City	4 to 8	445	79%	84%
3	TX	Suburban	6 to 8	869	91%	82%
4	CO	City	6 to 12	850	99%	95%
5	SC	Suburban	6 to 8	680	30%	23%
6	SC	Rural	6 to 8	754	18%	11%
7	SC	Suburban	6 to 8	737	32%	22%
Averages				727	61%	56%
National Averages					50.5%	48%

*__National Center for Educational Statistics__, Table 204.20 *Number and percentage of public school students participating in programs for National School Lunch Program, by district: Selected years, 2003-04 through 2013-14*

Description of the Assessments

The pretests and post-tests used in the study were developed to assess the language skills and understanding of students. Based on these standards multiple-choice assessment pre/post-tests were developed focusing on students' abilities to understand vocabulary, understand how English works, and reading comprehension as taught in the *Escalate English* program.

Table 2 provides the statistical results for the administration of the pretests and the post-tests for grades 6, 7, and 8. The KR 20 reliability and the Standard Error of Measurement for the post-test indicates both the pretest score results and the post-test score results were reliable for arriving at decisions regarding the achievement of the students to whom the tests were administered. The increase in the reliability of the tests from pretesting to post-testing at all three grade levels indicates the effect in learning and the decrease in guessing answers.

Table 2
Pretest and Post-Test Test Statistics

Test	Reliability*	SEM**
Grade 6 Pretest	.69	22.8
Grade 6 Post-test	.87	19.1
Grade 7 Pretest	.74	20.9
Grade 7 Post-test	.87	17.7
Grade 8 Pretest	.65	18.3
Grade 8 Post-test	.84	20.8

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.

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

All tests have a range of item discriminations. The average discrimination of all the items on a test should be above +.15. The highest discriminations are rarely above +.50.

A scale that can be used to evaluate the discrimination of test items and the number of items for each of the tests used in this study is provided in Table3. The table shows that for the grade 6, grade 7, and grade 8 post-tests a large percentage of the test items are acceptable, good or excellent test items. The average test item discriminations for all three grades can be described as excellent.

Table 3
Test Item Discrimination for Escalate English Post-test Assessments

Item Discrimination	Discrimination Values	Test Items in each Category		
		Grade 6 Post-test	Grade 7 Post-test	Grade 8 Post-test
Below 0	Poor test items (should be replaced)	2	0	0
+.01 to +.10	Weak test items (revise items)	1	0	2
+.11 to +.20	Acceptable	2	6	2
+.21 to +.30	Good items	2	0	5
+.30	Excellent test items	30	33	31

The table shows that for the grade 6 test 34 out of 37 test items were acceptable or above and 30 of those 34 items were rated as excellent. At grade 7, 33 of 39 items were considered acceptable or above and 33 were rated as excellent. Of the 40 grade 8 items, 38 were acceptable or above and 31 of those items were rated as excellent.

Data Analyses

Standard scores were developed to provide a more normal distribution of scores. The standard scores were a linear transformation of the raw scores. A mean raw score was translated to a mean standard score of 300 and the standard deviation of the raw scores was translated to 50. Standard scores were then used for the statistical analyses.

Data analyses and descriptive statistics were computed for the standard scores from the *Escalate English* assessments. The ≤.05 level of significance was used as the level at which increases would be considered statistically significant for all the statistical tests.

The following statistical analyses were conducted to compare students' pretest scores to post-test scores:

- A paired comparison *t*-test was used to compare the pretest mean standard scores with the post-test mean standard scores for all students.
- The students were split into two groups based on pretest scores. Paired comparison *t*-tests were used with the group that scored higher and the group that scored lower on the pretest to determine if the program was equally effective with students who had lower and higher pretest scores.

Descriptive statistics were also used to compare pretest and post-test standard test scores for the total group as well as the higher and lower pretest score groups.

An effect-size analysis was computed for each of the paired *t*-tests. Cohen's *d* statistic was used to determine the effect size. This statistic provides an indication of the strength of the effect of the treatment regardless of the statistical significance. Cohen's *d* statistic is interpreted as follows:

.2 = small effect
.5 = medium effect
.8 = large effect

Grade 6 Analyses

A paired comparison *t*-test to determine if there was a difference from pretest standard scores to post-test standard scores was statistically significant. For this analysis, pretest and post-test scores for 74 students were analyzed. Students who did not take both the pretest and the post-test were not included.

Table 4 shows that the average standard score on the pretest was 286, and the average standard score on the post-test was 314. The increase was statistically significant (≤.0001). The effect size was medium.

Table 4
Grade 6 Paired Comparison *t*-test Results
Pretest/Posttest Comparison of Standards Scores

Test	Number of Students	Mean Standard Score	SD	t-test	Significance	Effect Size
Pretest	74	286	41.5	4.876	≤.0001	.60
Posttest	74	314	53.7			

Higher and Lower Scoring Students

An additional analysis was conducted to determine if students who scored lower on the pretest made gains as great as those students who scored higher on the pretest. For this analysis students were ranked in order based on their pretest scores. The group of 74 students was divided into two equal sized groups of 37 students. The first group included those students who scored lower on the pretest with a mean standard score of 250 with scores ranging from 208 to 274. The higher scoring group scored a mean standard score on the pretest of 321 with scores ranging from 283 to 375.

Pretest-to-posttest comparisons are shown in Table 5 for the lower and higher pretest scoring students. Scores were analyzed using a paired comparison *t*-test to determine if both groups made significant gains.

For both the higher and the lower scoring groups, the average scores increased statistically significantly. The effect size for the lower pretest scoring group was large and for the higher pretest scoring the effect size was medium.

Table 5
Grade 6 Paired Comparison *t*-test Results for Pretest/Post-test Standard Scores
for the High- and Low-Scoring Pretest Groups

Test Form	Number of Students	Mean Standard Score	SD	t-test	Significance	Effect Size
Lower Scoring Group						
Pretest	37	250	18.7	4.115	≤.0001	1.02
Post-test	37	288	49.1			
Higher Scoring Group						
Pretest	37	321	23.9	2.730	≤.01	.55
Post-test	37	341	45.0			

Figure 1 provides a graphic representation of the gains achieved by the grade 6 students. The average scores for the total group increased 28 standard score points. The low pretest scoring students increased their average standard scores by 38 points which was almost an increase 100% higher than the high pretest scoring students whose average standard scores increased by 20 points.
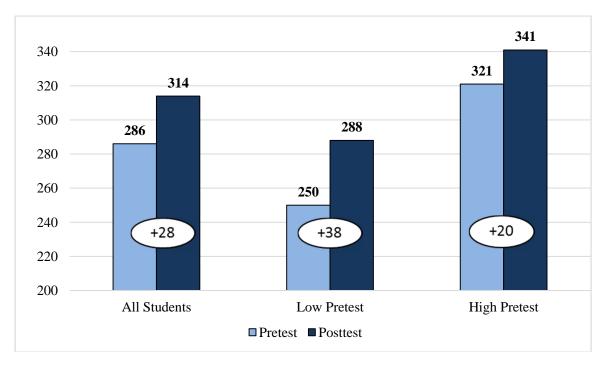
Figure 1
Grade 6 Pretest/Post-test Gain Comparison
All Students, Low Pretest Students, High Pretest Students

Grade 7 Analyses

A paired comparison *t*-test to determine if there was a difference from pretest standard scores to post-test standard scores was statistically significant. For this analysis, pretest and post-test scores for 81 students were available. Students who did not take both the pretest and the post-test were not included.

Table 6 shows that the average standard score on the pretest was 279, and the average standard score on the posttest was 321. The increase was statistically significant (≤.0001). The effect size was large.

Table 6
Grade 7 Paired Comparison *t*-test Results
Pretest/Post-test Comparison of Standards Scores

Test	Number of Students	Mean Standard Score	SD	t-test	Significance	Effect Size
Pretest	81	279	41.4	8.329	≤.0001	.92
Post-test	81	321	49.2			

Higher and Lower Scoring Students

An additional analysis was conducted to determine if students who scored lower on the pretest made gains as great as those students who scored higher on the pretest. For this analysis students were ranked in order based on their pretest scores. The group of 81 students was divided into two approximately equal sized groups. The first group included 40 students who scored lower on the pretest with a mean standard score of 247 with scores ranging from 155 to 274. The higher scoring group included 41 students who scored a mean standard score on the pretest of 310 with scores ranging from 274 to 386.

Pretest-to-posttest comparisons are shown in Table 7 for the lower and higher pretest scoring students. Scores were analyzed using a paired comparison *t*-test to determine if both groups made significant gains.

For both the higher and the lower scoring groups, the average scores increased statistically significantly. The effect sizes for both the lower and higher pretest scoring groups were large.
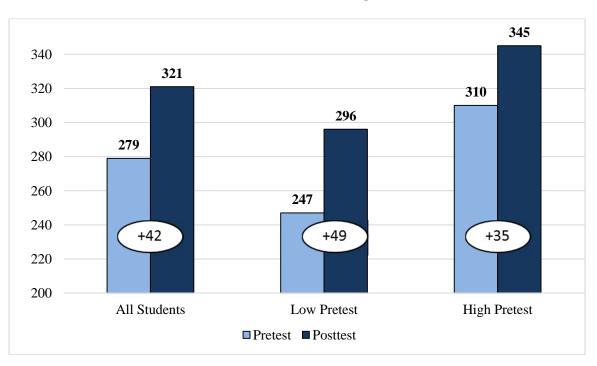
Table 7
Grade 7 Paired Comparison *t*-test Results for Pretest/Post-test Standard Scores
for the High- and Low-Scoring Pretest Groups

Test Form	Number of Students	Mean Standard Score	SD	t-test	Significance	Effect Size
Lower Scoring Group						
Pretest	40	247	26.6	6.920	≤.0001	1.40
Post-test	40	296	41.7			
Higher Scoring Group						
Pretest	41	310	26.1	4.950	≤.0001	.97
Post-test	41	345	44.0			

Figure 2 provides a graphic representation of the gains achieved by the Grade 7 students. The average scores for the total group increased 42 standard score points. The low pretest scoring students increased their average standard scores by 49 points and the high pretest scoring students average standard scores increased by 35 points.

Figure 2
Grade 7 Pretest/Post-test Gain Comparison
All Students, Low Pretest Students, High Pretest Students

Grade 8 Analyses

A paired comparison *t*-test to determine if there was difference from pretest standard scores to post-test standard scores was statistically significant. For this analysis, pretest and post-test scores for 29 students were included. Students who did not take both the pretest and the post-test were not included.

Table 8 shows that the average standard score on the pretest was 275, and the average standard score on the post-test was 325. The increase was statistically significant (≤.0001). The effect size was large.

Table 8
Grade 8 Paired Comparison *t*-test Results
Pretest/Posttest Comparison of Standards Scores

Test	Number of Students	Mean Standard Score	SD	t-test	Significance	Effect Size
Pretest	29	275	31.5	6.107	≤.0001	1.51
Post-test	29	325	52.7			

Higher and Lower Scoring Students

An additional analysis was conducted to determine if students who scored lower on the pretest made gains as great as those students who scored higher on the pretest. For this analysis students were ranked in order based on their pretest scores. The group of 29 students was divided into two approximately equal sized groups. The first group included 14 students who scored lower on the pretest with a mean standard score of 249 with scores ranging from 215 to 271. The higher scoring group included 15 students who scored a mean standard score on the pretest of 299 with scores ranging from 271 to 345.

Pretest-to-post-test comparisons are shown in Table 9 for the lower and higher pretest scoring students. Scores were analyzed using a paired comparison *t*-test to determine if both groups made significant gains.

For both the higher and the lower scoring groups, the average scores increased statistically significantly. The effect size for the both the lower and higher pretest scoring group was large.

Table 9
Grade 8 Paired Comparison *t*-test Results for Pretest/Post-test Standard Scores
for the High- and Low-Scoring Pretest Groups

Test Form	Number of Students	Standard Score	SD	t-test	Significance	Effect Size
Lower Scoring Group						
Pretest	14	249	18.9	4.306	≤.001	1.37
Post-test	14	303	53.0			
Higher Scoring Group						
Pretest	15	299	18.6	4.200	≤.001	1.38
Post-test	15	346	44.4			

Figure 3 provides a graphic representation of the gains achieved by the Grade 8 students. The average scores for the total group increased 28 standard score points. The low pretest scoring students increased their average standard scores by 38 points which was almost an increase 100% higher than the high pretest scoring students whose average standard scores increased by 20 points.
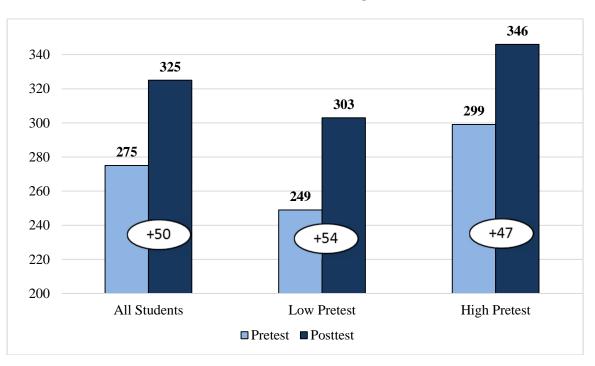
Figure 3
Grade 8 Pretest/Post-test Gain Comparison
All Students, Low Pretest Students, High Pretest Students

Conclusions

This study sought to determine the effectiveness of **Houghton Mifflin Harcourt Escalate English © 2017,** a grade 4 to 8 language development program. The study was carried out with classes at grades 6, 7, and 8. Twenty teachers in three different states were using the program for the first time and received no special instruction in using the program.

Two research questions guided the study:

Question 1: Is Houghton Mifflin Harcourt Escalate English *effective in increasing the skills and knowledge of grade 6 to 8 students so they can meet the increasing language demands in K-12 education?*

Pretests and post-tests were developed to match the standards of the Escalate English program. The assessments covered the objectives of the program and focused on three major areas of instruction: Vocabulary, How English Works, and Reading Comprehension. At all three grade levels analyses of students' scores showed that the scores were statistically significantly. The effect size for the increase was medium at grade 6 and large at grades 7 and 8.

Question 2: Is Houghton Mifflin Harcourt Escalate English *effective in increasing the skill and knowledge of grade 6 to 8 students at higher pretest scoring and lower pretest scoring students so they can meet the increasing language demands in K-12 education?*

At grades 6, 7, and 8 the analysis of the low scoring and high scoring pretest students showed that both groups increased statistically significantly. The effect size for the grade 6 higher pretest scoring group was medium while the effect size for the lower pretest scoring group was large. For grade 7 and 8 students the effect size was large for both the higher and scoring pretest group.

Based on this study, both research questions can be answered positively.

- *The Houghton Mifflin Harcourt Escalate English program is effective in increasing the language skills and knowledge of grade 6 to 8 students so they can meet the increasing language demands in K-12 education.*

- *The Houghton Mifflin Harcourt Escalate English program is effective in increasing the language skills and knowledge of lower performing as well as higher performing grade 6 to 8 students to analyze complex texts, determine evidence, reason critically, and communicate thoughtfully.*