

***Efficacy Study for On Our Way to English***  
***Published by Houghton Mifflin Harcourt***

*Report Number 505*

*July 2015*

# Table of Contents

---

Abstract .....	2
Overview of the Study .....	3
<i>Research Questions</i> .....	3
<i>Design of the Study</i> .....	3
<i>Program Overview</i> .....	3
Description of the Assessments .....	4
<i>Test Reliability and Standard Error of Measurement</i> .....	4
<i>Test Item Discrimination</i> .....	4
Description of the Study Sample .....	6
Data Analyses and Results .....	7
<i>Grade 2 Results</i> .....	7
<i>Grade 4 Results</i> .....	9
Conclusions .....	11
<i>Research Question 1</i> .....	11
<i>Research Question 2</i> .....	11

## Abstract

---

The focus of this study was the effectiveness of the Houghton Mifflin Harcourt kindergarten to grade 5 language development program for second language learners. The *On Our Way to English* program is generally referred to as OWE. The study included students from 3 different schools in 2 different states. The percentage of students in the study eligible for free/reduced lunch programs is much higher than the percentage of students eligible for free/reduced lunch programs for all public schools in the United States. The percentage of non-Caucasian students is also much higher than the percentage of non-Caucasian students enrolled in public schools in the United States.

The study was conducted during the 2014-2015 academic year. Approximately 70 students were included in the study. Teachers used the program as their primary curriculum with the second language learners for the entire academic year. Teaching experience for the 6 teachers who participated in the study ranged from less than 5 years to more than 15 years.

Pretests and posttests at both grades 2 and 4 included 40 items and were all multiple-choice items. The tests were designed to sample from the full content of the course. In addition to analyzing the gain scores for the total group of students at grade 2 and grade 4, analyses were conducted separately for higher and lower 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 achieving students and those scoring lowest on the pretests were designated as the lower achieving students.

*The average gain scores for the total group of students at both grades 2 and 4 statistically significant and the effect sizes for both grades were large.*

In addition, the average gain scores for the low and high scoring groups at each group were statistically significant. The effect sizes for the lower achieving students at grades 2 and 4 were large. The effect sizes for the higher achieving students were medium at grade 2 and large at grade 4. *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 school year study to evaluate the effectiveness of the *On Our Way to English*<sup>®</sup> (OWE) language program for second language learners in grades kindergarten to grade 5. The study compared students in grades 2 and 4. Pretests were administered to students the middle of September 2014 and the posttests were administered the middle of June 2015.

### Research Questions

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

- Does the implementation of the *On Our Way to English* program lead to improvement of students' skills and understanding of English?
- Does the implementation of the *On Our Way to English* program lead to improvement of students' skills and understanding of English for low achieving students as well as for high achieving students?

### Design of the Study

The design of the program called for the implementation of the OWE program for students enrolled in grades 2 and 4 during the 2014–2015 academic year. Two of the schools had used the program in the previous year in a one semester tryout of the program. The other school had not used the program prior to the 2014–2015 school year. The study included a grade 2 and a grade 4 teacher in each of three different schools in two states.

Teachers reported using the program from 3 to 5 days per week. Class size ranged from 15 to 19 students. However, one teacher reported a class size of fewer than 15 students. Teachers reported teaching experience that ranged from fewer than 5 years to more than 15 years.

### Program Overview

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

***On Our Way to English** is a comprehensive English language development program that provides everything teachers need for effective instruction. Domain- based instruction includes a focus on academic language and vocabulary development; thematic, content-based instruction; differentiated instruction for language and literacy; and a daily instructional routine in oral language, reading, and writing. Engaging online and digital tools motivate English language learners.*

*The program is designed to engage through visuals, relevant topics and meaningful activities. At the heart of **On Our Way to English** is a commitment to bring rich, culturally relevant language learning to every English language learner. The instruction is designed to challenge students to reach new heights through rigorous content specifically written to foster success across the four language domains. Built upon the latest research and the Common Core State Standards, **On Our Way to English** presents language learners with enhanced writing instruction, foundational skills, embedded speaking and listening activities and a myriad of text interaction opportunities that will bring students to the next level of language development.*

## Description of the Assessments

---

The pretest and posttest used in the study were developed to assess Common Core State Standards-based language knowledge and skills. Based on these standards a 40 item multiple-choice pretest and post-test assessment was developed for each grade. The assessments focusing on the skills, strategies, and knowledge necessary for effective English language understanding and performance.

### 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 English 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.

**Table 1**  
**Pretest and Posttest Statistics for the Grade 2 and Grade 4 Assessments**

<i>Test</i>	<i>Mean Score</i>	<i>Standard Deviation</i>	<i>KR 20</i>	<i>SEm*</i>
Grade 2 Pretest	281	42.5	.83	17.52
Grade 2 Posttest	319	50.2	.89	16.65
Grade 4 Pretest	276	39.7	.79	18.19
Grade 4 Posttest	324	48.5	.87	16.65

\*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. Test item discrimination is an easy concept to understand.

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.<sup>1</sup> 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. 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 two tests used in this study is provided in Table 2. The table shows that both the grade 2 and grade 4 posttests have a large percentage of acceptable, good or excellent test items grade 2 (95%) grade (85%).

---

<sup>1</sup> Item discrimination is determined by 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.

**Table 2**  
**Test Item Discrimination for OWE Post-test Assessments**

		Number of Test Items in each Category	
Item Discrimination Values	Interpretation of Discrimination Values	Grade 2 Posttest	Grade 4 Posttest
<i>Below 0</i>	Poor test items (should be replaced)	1 items	2 items
<i>+.01 to +.10</i>	Weak test items (revise items)	1 items	3 items
<i>+.11 to +.20</i>	Acceptable	2 test items	4 test items
<i>+.21 to +.30</i>	Good items	3 test items	3 test items
<i>+.30 and above</i>	Excellent test items	33 test items	28 test items
<i>Average Discrimination</i>		+.43	+.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 percentage of students classified as minority students (non-Caucasian) ranged from 50% to 98% with an average of 82%. By comparison, approximately 50% of the students enrolled in U.S. public schools were classified as non-Caucasian.<sup>2</sup>

The percentage of students enrolled in free/reduced lunch programs ranged from 62% to 100% and averaged 87% 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%.<sup>2</sup>

**Table 3**  
**Demographic Description of the Schools Included in the Study**

	State	Location	Grades	Enrollment	% Minority	% Free/Reduced Lunch
1	CT	Urban	PK-08	1063	97%	100%
2	CT	Urban	PK-08	900	98%	100%
3	VA	Suburban	PK-05	483	50%	62%
<b>AVERAGES</b>				815	82%	87%

---

<sup>2</sup> *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 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 at the beginning of September 2014 and the post-tests administered at the middle of June 2015. The  $\leq .05$  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

### Grade 2 Results

Table 4 shows that the average scores of the 28 grade 2 students participating in the study increased at a statistical significant level. The effect size was large.

**Table 4**  
**Grade 2 Paired Comparison *t*-test Results**  
**Pretest/Posttest Standard Score Comparisons**

	<i>Number Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Pretests	28	281	42.5	4.696	$\leq .0001$	.81
Post-tests	28	319	50.2			

The total group of 28 OWE students was divided into two equal sized groups based on their pretest scores. The 14 students scoring lowest on the pretest were considered to be lower OWE achieving students while the 14 students scoring highest on the pretest were considered to be the higher OWE achieving students.

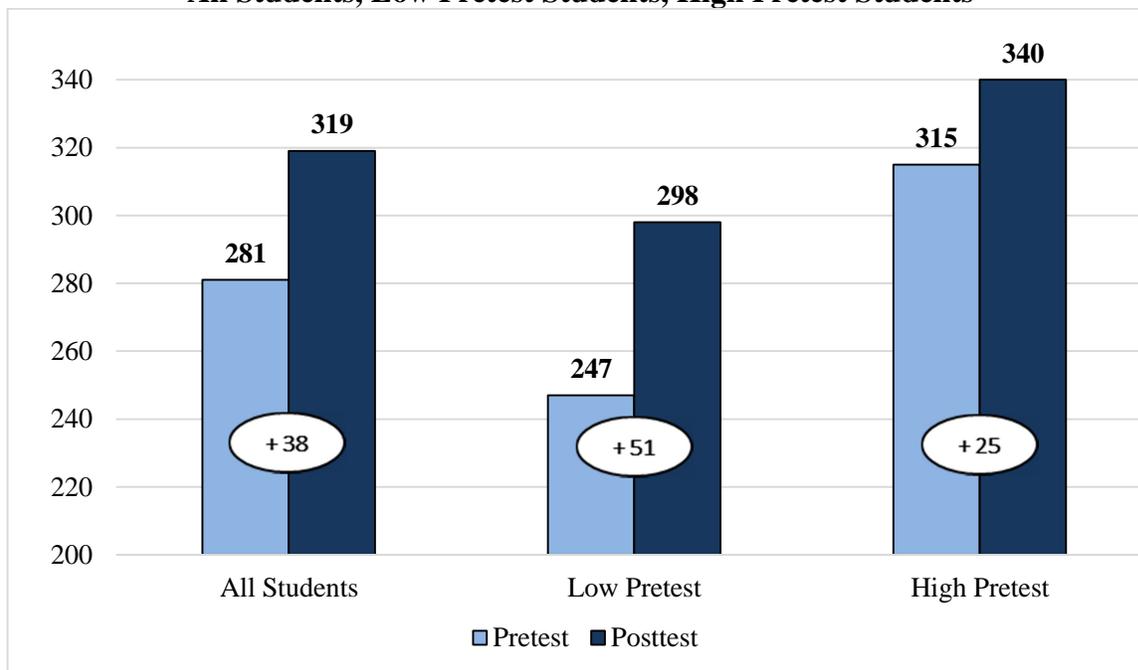
Table 5 shows that both groups made statistically significant gains. The effect sizes for the lower scoring group was large and the effect size for the higher scoring group was medium.

**Table 5**  
**Grade 2 Paired Comparison *t*-test Results**  
**High- and Low-Scoring Pretest Groups**

<i>Test</i>	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
<b>Lower Scoring Group</b>						
Pretest	14	247	14.4	4.449	≤.001	1.48
Posttest	14	298	46.6			
<b>Higher Scoring Group</b>						
Pretest	14	315	33.1	2.337	≤.03	.63
Posttest	14	340	45.7			

Figure 1 provides a graphic representation of the gains achieved by the grade 2 students. The average scores for the total group increased 38 standard score points. The low pretest scoring students increased their average standard scores by 51 points which was an increase of 100% higher than the high pretest scoring students whose average standard scores increased by 25 points.

**Figure 1**  
**Grade 2 Pretest Posttest Gain Comparison**  
**All Students, Low Pretest Students, High Pretest Students**



## Grade 4 Results

Table 6 shows that the average scores of the 41 grade 4 students participating in the study increased at a statistical significant level. The effect size was large.

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

	<i>Number Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Pretests	41	276	39.7	7.150	≤.0001	1.08
Post-tests	41	324	48.5			

The total group of 41 OWE students was divided into two approximately equal sized groups based on their pretest scores. The 20 students scoring lowest on the pretest were considered to be lower OWE achieving students while the 21 students scoring highest on the pretest scores were considered to be the higher OWE achieving students.

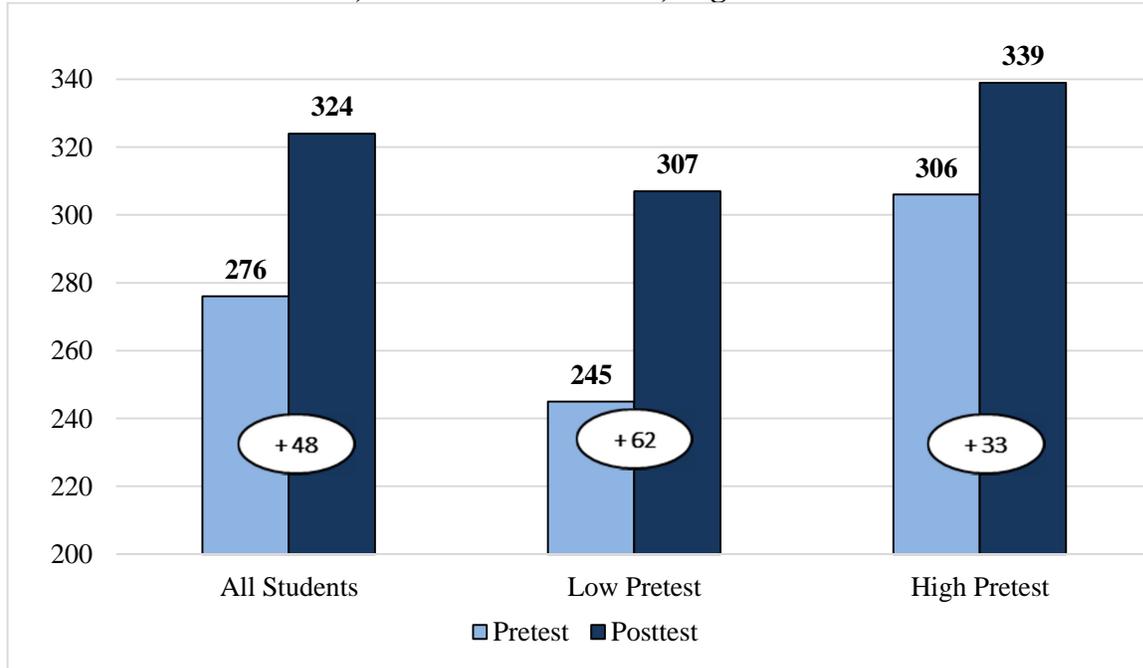
Table 7 shows that both groups made statistically significant gains. The effect sizes for both groups were large.

**Table 7**  
**Grade 4 Paired Comparison *t*-test Results**  
**High- and Low-Scoring Pretest Groups**

<i>Test</i>	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
<b>Lower Scoring Group</b>						
Pretest	20	245	10.8	5.961	≤.0001	1.85
Posttest	20	307	46.1			
<b>Higher Scoring Group</b>						
Pretest	21	306	33.8	4.609	≤.0001	.81
Posttest	21	339	46.6			

Figure 2 provides a graphic representation of the gains achieved by the grade 4 students. The average scores for the total group increased 48 standard score points. The low pretest scoring students increased their average standard scores by 62 points which was an increase of nearly 100% higher than the high pretest scoring students whose average standard scores increased by 33 points.

**Figure 2**  
**Grade 4 Pretest Posttest Gain Comparison**  
**All Students, Low Pretest Students, High Pretest Students**



## Conclusions

---

This study sought to determine the effectiveness of the On Our Way to English program for elementary grade students. The study was conducted with grade 2 and grade 4 students and took place during the 2014-2015 academic year. The student population included a much larger percentage of students eligible for free-reduced price lunch programs than the national average. The percentage of non-Caucasian student was also much larger than the national average.

Analyses of the pretest and posttest assessments used for both grades 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 *On Our Way to English* program lead to improvement of students' skills and understanding of English?

The results for both grade 2 and grade 4 students showed statistically significant growth from pretesting to post-testing with large effect sizes at both grades.

### Research Question 2

- Does the implementation of the *On Our Way to English* program lead to improvement of students' skills and understanding of English for low achieving students as well as for high achieving students?

For both grades included in the study English language achievement increased statistically significantly for both the high achieving and low achieving students. The effect sizes for the lower achieving students in all two grades was large. The effect size for the higher achieving students at grade 2 was medium while the effect size for the higher achieving students at grade 4 was large.

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

The OWE English language program resulted in statistically significant increases for students at grades 2 and 4 and the effect sizes were large.

The OWE English language program resulted in statistically significant growth for both higher ability and lower ability students for both grades. The effect sizes for the lower achieving students were large at both grades. The effect sizes for the higher achieving students were medium at grade 2 and large at grade 4.