

Journeys Reading Program

An Efficacy Study

Houghton Mifflin Harcourt

Report Number 523

July 2016

Table of Contents

Abstract	3
Overview of the Study	4
<i>Research Questions</i>	4
<i>Design of the Study</i>	4
<i>Program Overview</i>	5
<i>Description of the Assessments</i>	5
<i>Description of the Study Sample</i>	5
Data Analyses and Results	7
<i>Grade One Results</i>	7
<i>Grade Two Results</i>	9
<i>Grade Three Results</i>	10
<i>Grade Four Results</i>	12
<i>Grade Five Results</i>	13
Conclusions	15
<i>Research Question 1</i>	15
<i>Research Question 2</i>	15

Abstract

The focus of this study was the effectiveness of Journeys ©2017, a reading program for kindergarten to grade 6 students, published by Houghton Mifflin Harcourt. The study included students from 15 different schools in 5 different states. The study employed a unique design. Teachers who schools had chosen the program as their school reading program were asked to select one unit of instruction that would be the focus of the study. The unit selection was left to each teacher. The only consideration was that the unit be one that the teacher would be planning to use near the beginning of the second semester of the 2015-2016 academic year. Length of time to complete the unit depended on the grade level and the teacher's teaching pace. Most teachers took between 6 and 8 weeks to complete the study. Pretest and post-test pairs of assessments were developed for each unit chosen for inclusion in the study.

The study was conducted with over 650 students enrolled in grades 1 to 5. Only those students who took both a pretest and posttest were included in the data analysis. Teachers used the program for their reading instruction five days per week and more than 25 minutes per day.

The study was thus an intensive study of one unit per teacher using the Journeys program. A total of 12 different units were identified by teachers. Pretests and posttests were developed for each unit by reading/language arts curriculum specialists and were based on the standards for each unit included in the study. 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 students. Higher and lower scoring students were identified by the students' pretest scores. Those scoring highest on the pretests were designated as the high scoring reading students and those scoring lowest on the pretests were designated as the lower scoring reading students.

The average gain scores for the total group of students at every grade were statistically significant. At every grade level the scores for the low and high pretest scoring groups the scores increased statistically significantly with the exception of grade one. The high scoring group at grade one had identical average pretest and post-test scores of 93% correct. The pretest scores were so high there was little possibility of increasing their scores from pretesting to post-testing.

At every grade level the lower pretest scoring group effect size was large with the exception of grade four with a medium effect size. For the high pretest scoring group the effect size was small at all grade with the exception of grade 4 where the effect size was medium.

Overview of the Study

Houghton Mifflin Harcourt, School Publishers contracted with Educational Research Institute of America (ERIA) to conduct a study to evaluate the effectiveness of single instructional units of the Journeys Reading Program for grades K to 6. Grade one to five teachers who were already using the program in their classes were contacted and were asked to participate. Each teacher was to select a unit from the program which the teacher would be using near the beginning of the second semester of the 2015-2016 academic year.

A total of 41 teachers from 5 states and 15 different schools agreed to participate in the study. A different pretest/post-test pair of assessments was developed for each unit of study selected by a teacher. The pretests were administered prior to the time the teacher began using the chosen unit and post-tests were administered after instruction for the unit was completed. Teachers took between 6 and 8 weeks to administer the unit of instruction.

Research Questions

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

- Does the implementation of a single unit in the *Journeys Elementary Reading Program* in grades 1 to 5 lead to improved student reading achievement?
- Does the implementation of a single unit in the *Journeys Elementary Reading Program* in grades 1 to 5 lead to improved student reading achievement for higher scoring students as well as for lower scoring students?

Design of the Study

The design of the program called for the implementation of a single unit of the *Journeys Reading Program* for grade one to five students during the second semester of the 2015–2016 academic year. Units of study were chosen by teachers according to their teaching plan for the unit most likely to be used during the early part of the second semester.

A total of 41 teachers in 5 different states participated in the study. The number of teachers at each grade included:

- Grade One: 2 schools, 2 states, 5 teachers
- Grade Two: 5 schools, 3 states, 11 teachers
- Grade Three: 3 schools, 2 states, 5 teachers
- Grade Four: 3 schools, 3 states, 5 teachers
- Grade Five: 8 schools, 4 states, 15 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 in grade one have been teaching for 6 to 10 years, grade two teachers have been teaching 5 years or fewer, grade three teachers more than 15 years, and grade four and five teachers 11 to 15 years.

Program Overview

The Journeys 2017 program is described by the publisher as follows:

With Journeys © 2017, readers are empowered by skill mastery; inspired by authentic, award-winning text; and confident that they are building the skills needed for college and careers. A realistically paced close reading routine and online tools empower students to read rigorous texts. Using authentic text to anchor the core instruction, Journeys weaves the skills of close reading into a practical routine that is designed to have students read and reread for a variety of purposes, giving students more time to dig deeper into the text. The Student eBook provides tools that promote close reading such as responding to questions at point-of-use, highlighting text, and taking notes online. Journeys Close Reader consumable resources feature the high-quality paired text from each Student Book lesson and instruction in reading, re-reading, note-taking, and text annotation—empowering students to read any rigorous text.

Description of the Assessments

The pretest and posttest used in the study were developed by ERIA curriculum experts. Tests were developed to match the content of each of the units included in the study.

Table 1 provides a summary of the post-test statistics. The table shows that the reliabilities of the tests are high and provide adequate stability to assess reading achievement.

Table 1
Pretest and Posttest Statistics for the Journeys Students
Grades One to Five

Test	Mean Score	Standard Deviation	KR 20*	SEm**
Grade 1 Unit 4 Post-test	251	48.7	.86	18.22
Grade 1 Unit 5 Post-test	311	40.5	.86	15.15
Grade 2 Unit 4 Post-test	378	50.3	.86	18.82
Grade 2 Unit 5 Post-test	322	35.4	.76	17.34
Grade 3 Unit 4 Post-test	289	44.9	.84	17.96
Grade 3 Unit 5 Post-test	334	37.7	.84	15.08
Grade 3 Unit 6 Post-test	305	21.6	.56***	14.33
Grade 4 Unit 4 Post-test	292	47.5	.84	19.00
Grade 4 Unit 5 Post-test	329	40.7	.80	18.20
Grade 5 Unit 3 Post-test	331	37.4	.86	13.99
Grade 5 Unit 4 Post-test	273	36.7	.75	18.35
Grade 5 Unit 5 Post-test	312	41.9	.81	18.26

*KR 20 stands for Kuder-Richardson 20 measure of internal-test reliability

**SEm stands for Standard Error of Measurement.

***The relatively low reliability is most likely due to the fact that only 15 grade three students were included in the tryout for this unit.

Description of the Study Sample

Table 2 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 an estimate of the make-up of the classes included in the study.

The percentage of students classified as minority students (non-Caucasian) averaged 28% and ranged from 7% to 98%. The percentage of students enrolled in free/reduced lunch programs averaged 35% and ranged from 8% to 95%.

By comparison, the National Center for Educational Statistics reports that approximately 50% of the students enrolled in U.S. public schools are classified as non-Caucasian, and the reported national average for students enrolled in free/reduced lunch programs in public schools is reported as approximately 48%.¹

Table 2
Demographic Description of the Schools Included in the Study

	State	Location	Grades	Enrollment	% non-Caucasian	% Free/Reduced Lunch
1	IL	Suburban	PK to 8	1037	98%	95%
2	KS	Urban	PK to 5	276	50%	59%
3	KS	Urban	K to 5	364	43%	33%
4	NE	Suburban	PK to 6	420	7%	20%
5	NE	Suburban	K to 6	420	N/A*	N/A*
6	NE	Suburban	PK to 6	426	27%	40%
7	NE	Suburban	PK to 6	556	11%	8%
8	NE	Suburban	PK to 6	441	16%	9%
9	NE	Suburban	PK to 6	403	13%	19%
10	NE	Suburban	PK to 6	367	34%	56%
11	NE	Suburban	PK to 6	274	18%	14%
12	NY	Suburban	K to 3	325	22%	14%
13	WA	Suburban	K to 6	587	11%	25%
14	WA	Suburban	K to 6	511	14%	30%
15	WA	Suburban	K to 6	522	34%	69%
	Average			462	28%	35%

¹ *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.

Paired comparison *t*-tests were used to determine if differences in pretest and post test scores were significantly different. 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 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’.” 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

Grade One Results

Table 3 shows that the average scores of the 116 grade one students participating in the study increased at a statistically significant level. The effect size was substantively important and is classified as medium.

Table 3
Grade One Total Group Paired Comparison *t*-test Results
Pretest/Posttest Standard Score Comparisons

	<i>Number Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t</i> -test	<i>Significance</i>	<i>Effect Size</i>
Pretests	116	287	52.90	6.247	$\leq .0001$.56
Post-tests	116	314	43.06			

The total group of 116 grade one students was divided into two equal sized groups based on their pretest scores. The 58 students scoring lowest on the pretests were considered to be lower reading achievement students while the 58 scoring highest on the pretests were considered to be higher reading achievement students.

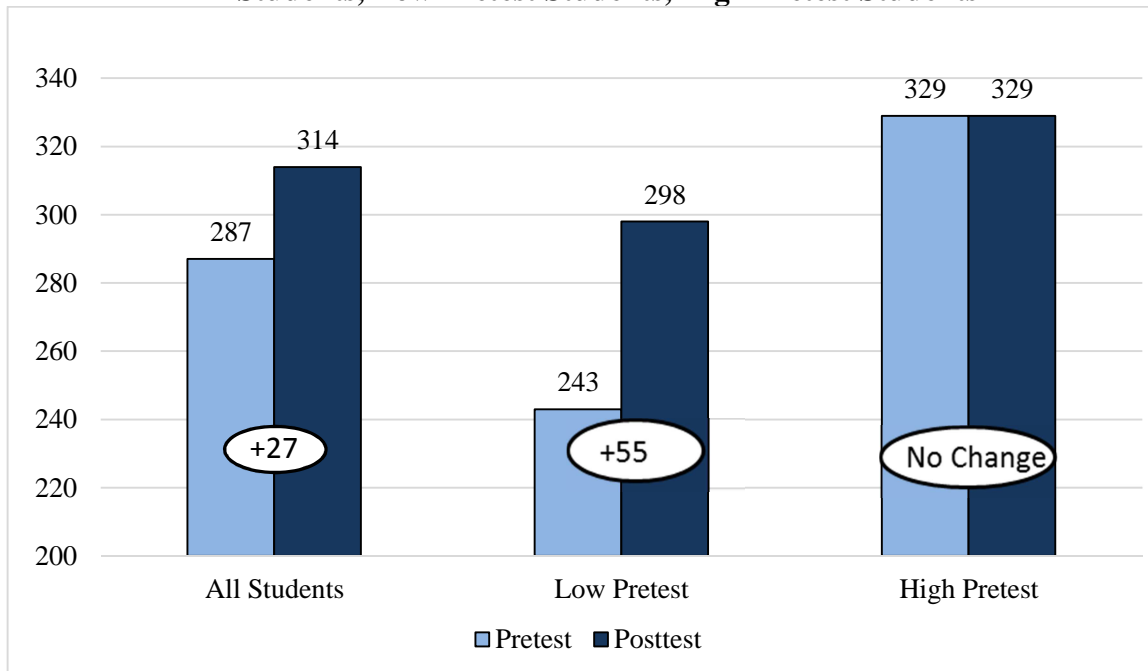
Table 4 shows that the low pretest scoring group made statistically significant gains while the high pretest scoring group did not increase due to the fact that they scored at an average of 93% correct on both the pretests and post-tests. In comparison the low pretest students scored an average percent correct of 65% on the pretests and 83% correct on the post-tests. The effect size for the low scoring group was substantively important and was classified as large.

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

	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Lower Scoring Group						
Pretest	58	243	39.5	9.524	≤.0001	1.21
Posttest	58	298	49.6			
Higher Scoring Group						
Pretest	58	329	17.2	.036	Non-Significant	--
Posttest	58	329	27.7			

Figure 1 provides a graphic representation of the gains achieved by the grade one students. In an eight-week period, using assessments focused on just one unit of instruction, the grade one students increased their average standard scores by 27 points. The low achieving reading students increased their average standard scores by 55 points and due to very high pretest and post-test average scores the high achieving students did not change.

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



Grade Two Results

Table 5 shows that the average scores of the 261 grade two students participating in the study increased at a statistically significant level. The effect size was substantively important and was classified as small.

Table 5
Grade Two Total Group 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	261	288	52.9	8.906	≤.0001	.49
Post-tests	261	312	44.0			

The total group of 261 grade two students was divided into two approximately equal sized groups based on their pretest scores. The 130 students scoring lowest on the pretests were considered to be lower reading achievement students while the 131 scoring highest on the pretests were considered to be higher reading achievement students.

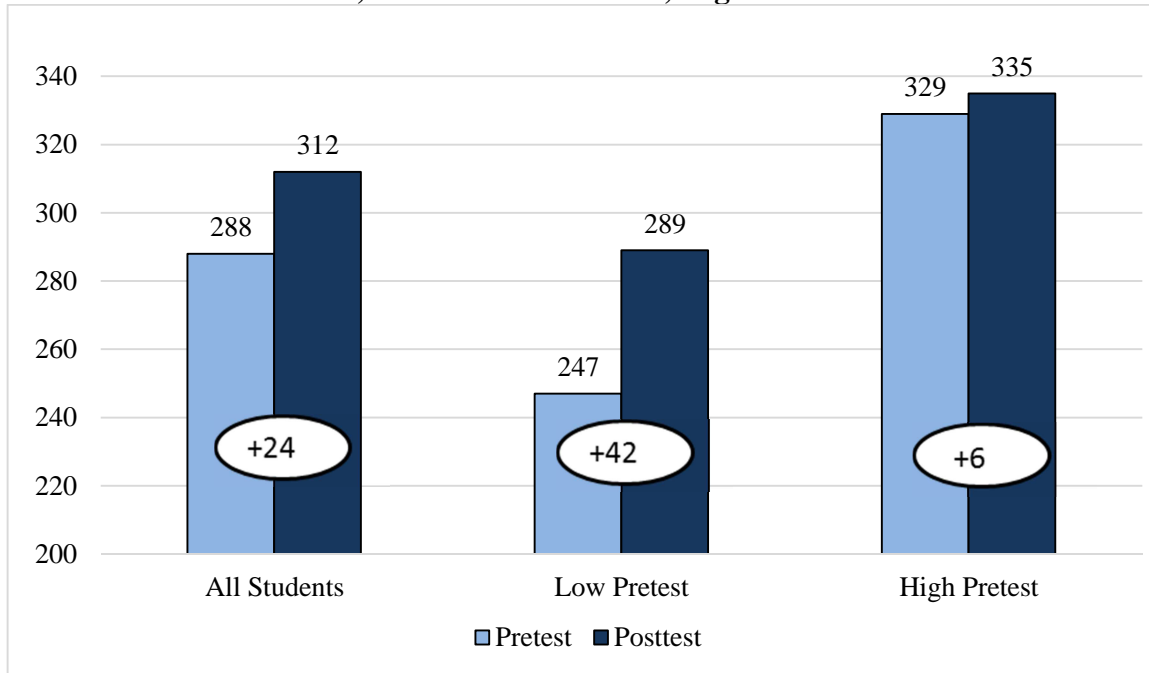
Table 6 shows that the low pretest scoring group and the high pretest scoring group made statistically significant gains. The effect sizes for the low scoring group was substantively important and was classified as large. The effect size for the high scoring group was substantively important and was classified as small.

Table 6
Grade Two Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Lower Scoring Group						
Pretest	130	247	43.5	10.010	≤.0001	.94
Posttest	130	289	45.5			
Higher Scoring Group						
Pretest	131	329	18.0	2.315	≤.02	.25
Posttest	131	335	27.4			

Figure 2 provides a graphic representation of the gains achieved by the grade two students. In an eight- week period, using assessments focused on just one unit of instruction, the grade two students increased their average standard scores by 24 points. The low achieving reading students increased their average standard scores by 42 and the high achieving students increased their scores by 6 points.

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



Grade Three Results

Table 7 shows that the average scores of the 132 grade three students participating in the study increased at a statistically significant level. The effect size was substantively important and is classified as small.

Table 7
Grade Three Total Group 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	132	290	53.0	6.958	≤.0001	.40
Post-tests	132	310	44.9			

The total group of 132 grade three students was divided into two equal sized groups based on their pretest scores. The 66 students scoring lowest on the pretest were considered to be lower reading achievement students while the 66 scoring highest on the pretest scores were considered to be higher reading achievement students.

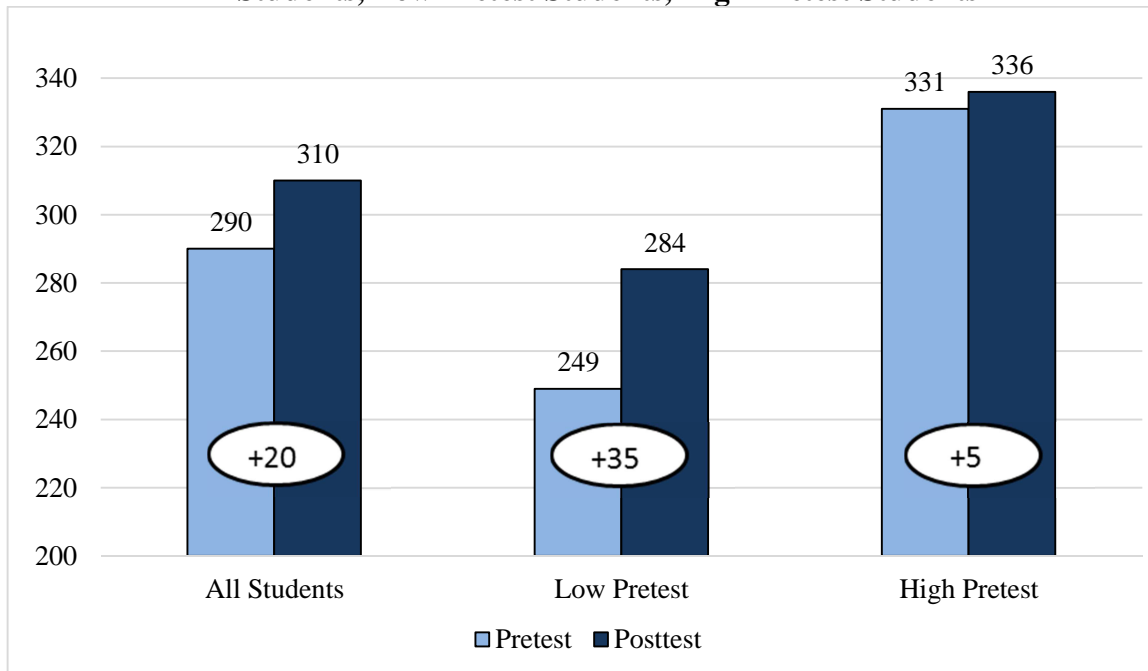
Table 8 shows that both groups made statistically significant gains. The effect sizes for both groups were substantively important. The low pretest groups effect size was classified as large and the effect size for the higher pretest scoring group was classified as small.

Table 8
Grade Three Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Lower Scoring Group						
Pretest	66	249	40.3	7.839	≤.0001	.81
Posttest	66	284	44.3			
Higher Scoring Group						
Pretest	66	331	24.9	1.953	≤.05	.18
Posttest	66	336	26.4			

Figure 3 provides a graphic representation of the gains achieved by the grade three students. In an eight- week period, using assessments focused on just one unit of instruction, the grade three students increased their average standard scores by 20 points. The low achieving reading students increased their average standard scores by 35 points and the high achieving students increased their scores by 5 points.

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



Grade Four Results

Table 9 shows that the average scores of the 117 grade four students participating in the study increased their average test scores at a statistically significant level. The effect size was substantively important and is classified as small.

Table 9
Grade Four Total Group 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	117	290	50.2	8.301	≤.0001	.42
Post-tests	117	310	47.8			

Based on their pretest scores, the total group of 117 grade four students was divided into two approximately equal sized groups. The 59 students scoring lowest on the pretest were considered to be lower reading achievement students while the 58 students scoring highest on the pretest scores were considered to be higher reading achievement students.

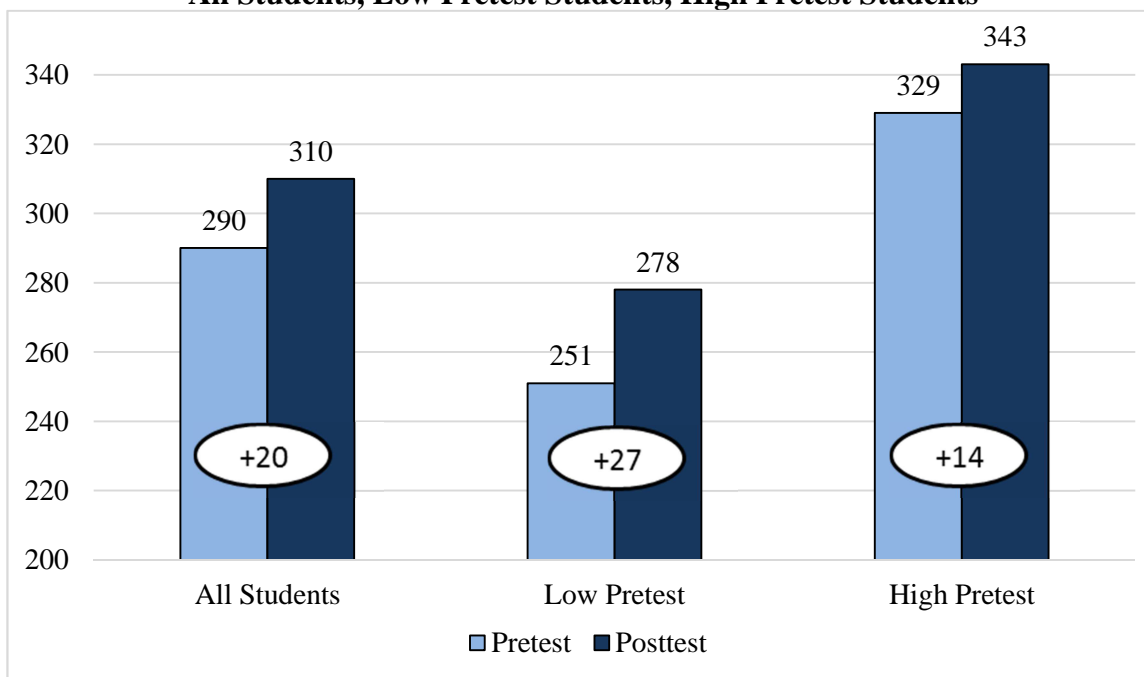
Table 10 shows that both groups made statistically significant gains. The effect sizes for both groups were substantively important and are classified as medium.

Table 10
Grade Four Paired Comparison *t*-test Results
High- and Low-Scoring Pretest Groups

	<i>Number of Students</i>	<i>Mean Standard Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Lower Scoring Group						
Pretest	59	251	40.6	7.002	≤.0001	.65
Posttest	59	278	40.9			
Higher Scoring Group						
Pretest	58	329	19.1	4.817	≤.0001	.62
Posttest	58	343	27.1			

Figure 4 provides a graphic representation of the gains achieved by the grade four students. In an eight- week period, using assessments focused on just one unit of instruction the grade four students increased their average scores by 20 standard score points. The low achieving reading students increased their average scores by 27 standard score points while the high achieving reading students increased their average scores 14 standard score points.

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



Grade Five Results

Table 11 shows that the average scores of the 73 grade five students participating in the study increased their average test scores at a statistically significant level. The effect size was substantively important and is classified as medium.

Table 11
Grade Five Total Group 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	73	286	56.7	5.670	≤.0001	.60
Post-tests	73	315	37.3			

Based on their pretest scores, the total group of 73 grade five students was divided into two approximately equal sized groups. The 37 students scoring lowest on the pretest were considered to be lower reading achievement students while the 36 students scoring highest on the pretest scores were considered to be higher reading achievement students.

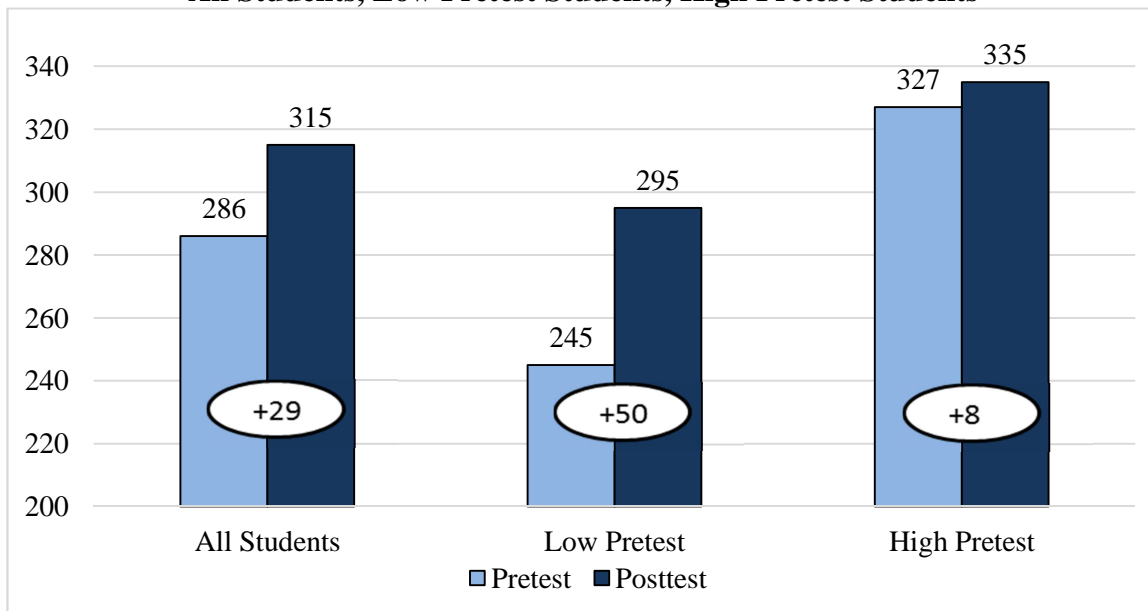
Table 12 shows that both groups made statistically significant gains. The effect sizes for both groups were substantively important. The effect size for the lower scoring group was classified as large and the effect size for the higher scoring group was classified as small.

Table 12
grade five 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>
Lower Scoring Group						
Pretest	37	245	52.4	6.019	≤.0001	1.07
Posttest	37	295	38.2			
Higher Scoring Group						
Pretest	36	327	17.7	2.242	≤.03	.38
Posttest	36	335	23.0			

Figure 5 provides a graphic representation of the gains achieved by the grade 5 students. In an eight-week period, using assessments focused on just one unit of instruction, the grade five students increased their average scores by 29 standard score points. The low achieving reading students increased their average scores by 50 standard score points while the high achieving reading students increased their average scores 8 standard score points.

Figure 5
Grade Five Pretest Posttest Gain Comparison
All Students, Low Pretest Students, High Pretest Students



Conclusions

This study sought to determine the effectiveness of the Journeys Reading Program from grades 1 to 5 based on a single unit of instruction selected independently by each of 41 teachers. The study took place during the second semester of the 2015-2016 academic year and was carried out in 5 states and included 15 different schools and 41 teachers. The student population included a somewhat smaller percentage of non-Caucasian students (28%) than the national average (50%). The percentage of students eligible for free-reduced price lunch programs (35%) was much smaller than the national average (48%).

Research Question 1

- Does the implementation of a single unit in the Journeys Elementary Reading Program in grades one to five lead to improved student reading achievement?

At all five grades reading achievement growth from pretesting to post-testing increased statistically significantly. The effect size at each grades was above a substantively important level.

Research Question 2

- Does the implementation of a single unit in the Journeys Elementary Reading Program in grades one to five lead to improved student reading achievement for higher scoring students as well as for lower scoring students?

At all five grades included in the study reading achievement growth for the high achieving and low achieving students increased statistically significantly. At all grade levels both the high and low pretest scoring students effect sizes were above a substantively important level. The exception was grade one at which the grade one students scored at a 93% correct level and thus any gain was highly unlikely.

The effect sizes for the lower pretest scoring groups at grades one to five were large with the exception of grade four with a medium effect size. The effect sizes for the high pretests scoring students at grades 2, 3, and 5 were small. At grade one the effect size for the high pretest scoring group was non-existent due to pretest average scores of 93% correct. At grade four the effect size for the high pretest scoring group was medium.

For this relatively brief highly focused study, both research questions can be answered positively:

The Journeys Reading Program produced statistically significant increases based on pretest/post-test scores across 12 separate units of study.

The Journeys Reading Program produced statistically significant growth for both higher ability and lower ability students at all grades for all 12 units with the exception of high pretest scoring students at grade one.