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

A STUDY OF THE INSTRUCTIONAL EFFECTIVENESS OF *HOUGHTON MIFFLIN HARCOURT JOURNEYS*

Report Number 368A

June 2009

Advisory Board:

Michael Beck, President
Beck Evaluation & Testing Associates, Inc.

Jennifer M. Conner, Assistant Professor
Indiana University

Keith Cruse, Former Managing Director
Texas Assessment Program

Joseph A. Fernandez, Former Chancellor
New York City Public Schools

Table of Contents

Abstract……2

Overview of the Study 3

Project Background..... 3

Research Questions..... 3

Design of the Study..... 3

Instructional Approach under Study 4

Description of the Research Sample 6

Description of the Assessments 9

Data Analyses 10

Grade 1 Results 11

Control Group/Experimental Group Posttest Analyses 11

Pretest/Posttest Analyses 13

Grade 5 Results 20

Control Group/Experimental Group Posttest Analyses 20

Pretest/Posttest Analyses 22

Conclusions..... 29

Control Group/Experimental Group Posttest Analyses 29

Pretest/Posttest Analyses 30

A CONTROL GROUP/EXPERIMENTAL GROUP STUDY OF THE INSTRUCTIONAL EFFECTIVENESS OF HOUGHTON MIFFLIN HARCOURT JOURNEYS PROGRAM,

ABSTRACT

This paper reports on a quasi-experimental study designed to determine the effect of Houghton Mifflin Harcourt Journeys program, on students' reading skills and strategy use. Twenty-two grade 1and grade 5 teachers from 19 different elementary schools participated in the study as either control or experimental class teachers. The experimental classes included significant numbers of English Language Learners (ELL); however, no ELL students were included in the control classes.

The fifteen teachers participating in the experimental group taught a single unit from HOUGHTON MIFFLIN HARCOURT JOURNEYS PROGRAM. The program included teaching suggestions and program materials for use with ELL students. The seven control group teachers continued to use the reading programs they had been using prior to their involvement in the study. Experimental group teachers administered a pretest prior to beginning instruction using JOURNEYS and a posttest after teaching a single unit from the program; the control group teachers administered the posttest at the same time as the experimental group teachers.

Results of statistical analyses revealed that students who received instruction using the JOURNEYS program had statistically significant reading skills and strategy achievement when compared to students in the control group. Pretest to posttest analyses for the experimental group revealed significant growth in reading skills and strategy. In addition, analyses showing the growth from pretest to posttest for the ELL students indicated growth from pretest to posttest that was statistically significant and equal to the non-ELL students.

Overview of the Study

This report describes an instructional efficacy study that was conducted to determine the impact of *HOUGHTON MIFFLIN HARCOURT JOURNEYS,* on students' reading skills and strategy use.

Project Background

There has never been a greater need to ensure that the reading programs that young students are using are optimally supporting them in developing the literacy skills required for success in high school, college, and in the workplace.

Because of the importance of determining the effectiveness of reading programs, Houghton Mifflin Harcourt contracted with the Educational Research Institute of America (ERIA) to study the effectiveness of *HOUGHTON MIFFLIN HARCOURT JOURNEYS,* using an experimental group and a control group design. This report presents the findings from that study.

Research Questions

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

> Is *HOUGHTON MIFFLIN HARCOURT JOURNEYS, effective* in improving students' reading skills and strategy use?
>
> Is *HOUGHTON MIFFLIN HARCOURT JOURNEYS, effective* in improving the reading skills and strategy use of English Language Learning (ELL) students?

Design of the Study

The study was conducted in grades 1 and 5, with a control group and an experimental group at each grade level. Control group classrooms were identified that matched as closely as possible to the demographic characteristics of the experimental group classrooms.

At grade 1, the experimental group included eight teachers from six public schools located in four different states. The control group included four teachers from four public schools located in two different states.

At grade 5, the experimental group included seven teachers from six public schools located in four different states. The control group included three teachers from three public schools located in the same state.

In total the study included 22 teachers, 19 different schools, and 409 students, which includes both those students enrolled in experimental group classrooms and control group classrooms.

All of the experimental group teachers taught a single unit from *HOUGHTON MIFFLIN HARCOURT JOURNEYS,* a comprehensive reading/language arts program for students in grades K through 6 that includes teaching suggestions and program materials that can be used in classrooms with ELL students. During the course of the two to three week study,

the experimental group teachers used *HOUGHTON MIFFLIN HARCOURT JOURNEYS*, as their primary program for reading instruction. None of the participating teachers had used the program prior to their involvement in the study.

Upon completion of their participation in the study, the experimental group teachers filled out a questionnaire that asked them about their use of the program during the study, in order to determine the fidelity with which they used the program materials. The teachers reported using the program for an average of ten school days, with all teachers reporting that they used the program for at least eight days. According to the questionnaire results, teachers used the program for an average of one hour per day, with all teachers reporting that they used the program for at least one half hour per day.

According to their completed questionnaires, teachers at grade 1 and grade 5 who had ELL students in their classrooms used the ELL teaching suggestions and materials "to some extent" or "to a great extent."

All experimental group teachers administered the pretest during the fourth week of April 2009 and administered the posttest in the second week of May 2009. Control group teachers also administered a posttest in the second week of May 2009. All tests and questionnaires were returned to ERIA the week of May 18.

Instructional Approach under Study

Following is a description of the program provided by Houghton Mifflin Harcourt Harcourt:

> HOUGHTON MIFFLIN HARCOURT JOURNEYS *is a brand new fully integrated Reading/Language Arts program. Through the best in new literature from Kindergarten to grade six* HOUGHTON MIFFLIN HARCOURT JOURNEYS, *applies the latest research to help all children become readers. The Journey begins at Kindergarten where children are introduced to the concepts of print through thirty beautiful Big Books and thirty Read Alouds Trade Books. Each read aloud serves to introduce a series of oral vocabulary that begins an extraordinary adventure in building words! The oral vocabulary effort continues throughout the balance of Journeys assuring that every student will have the vocabulary necessary to excel through grade 6 and beyond.*
>
> *Vocabulary is further covered through Vocabulary in Context student pages and Vocabulary in Context Cards all designed to provide important vocabulary routines and build a strong bank of Tier II words for every child.*
>
> *Comprehension in* HOUGHTON MIFFLIN HARCOURT JOURNEYS *also begins with the Read Alouds at each grade level. This is supported by direct skill instruction in the student text for both Comprehension skills and strategies. Both vocabulary and comprehension skills and strategies are reinforced through Leveled Vocabulary Readers, 4 Leveled Readers and 2 weekly selections all designed to help children develop and apply each skill for building fluency.*
>
> *Each leveled reader is leveled through the guidance of Irene Fountas, the leading expert in small group instruction and a consulting author for Journeys. Graphic*

Organizers are a constant that tie each of these pieces together. Critical graphic organizers are introduced early in each lesson and reinforced throughout the week with each piece of literature and each leveled text. For children this means an opportunity to understand graphic organizers in a way that creates a direct application to standardized test effectiveness. Retelling, complete with retelling cards for support, help children beginning at Kindergarten to show the comprehension they know.

At grades 3-6 Houghton Mifflin Harcourt completes the year in Unit 6 with a combination of magazines and trade books designed to review skills in a fun and motivating way. These magazines are a core part of the instruction in HOUGHTON MIFFLIN HARCOURT JOURNEYS.

Daily plans are included at all grade levels for both whole and small group instruction. The differentiated nature of the small group allows each child to progress at their level as they learn grade level skills. Ready Made Workstation Flipcharts provide the perfect answer for what the other children are doing while their teacher is involved in the small group plan. A Weekly To-Do List helps individualize instruction and keeps each child on track.

Description of the Research Sample

Tables 1 and 2 provide demographic summaries of the 19 different schools included in the study. This school data does not provide a description of the make-up of each of the classes that participated in the study. However, the tables do provide general descriptions of each of the schools and, thereby, an estimate of the make-up of the classes that comprised the sample.

At grade 1, students and teachers from eight classrooms in six different schools made up the experimental group. The control group included students and teachers from four classrooms in four different schools. As can be seen in Table 1, the average enrollments of the experimental group schools and the control group schools were about the same. The average percentage of students enrolled in free/reduced lunch programs was somewhat higher in the control group schools than in the experimental group schools, as was the average percentage of minority students.

At grade 5, students and teachers from seven classrooms in six different schools made up the experimental group. The control group included students and teachers from three different classrooms in three different schools. As can be seen in Table 2, the average enrollment of the experimental group schools was somewhat lower than the average enrollment of the control group schools. The average percentage of students enrolled in free/reduced lunch programs and the average percentage of minority students was very similar in the control group schools and the experimental groups schools.

Table 1
Demographic Characteristics of Grade 1
Experimental Group and Control Group Schools Included in the Study

Location	Grades	Students Enrolled	% Students Free/Reduced Lunch Programs	% Minority	% Students With Special Education Needs
EXPERIMENTAL GROUP SCHOOLS					
Urban Fringe Mid-Size City	K to 5	230	90%	73%	19%
Urban Fringe Mid-Size City	K to 6	480	75%	79%	11%
Mid-Size Central City	K to 5	289	78%	83%	20%
Mid-Size Central City	PK to 6	554	99%	95%	19%
Mid-Size Central City	PK to 6	533	70%	70%	21%
Urban Fringe Large City	K to 5	207	56%	47%	19%
	Averages	***382***	***78%***	***75%***	***18%***
CONTROL GROUP SCHOOLS					
Mid-Size Central City	K to 5	289	78%	83%	20%
Mid-Size Central City	PK to 6	554	99%	95%	19%
Mid-Size Central City	PK to 6	533	70%	70%	21%
Urban Fringe Mid-Size City	K to 5	230	90%	73%	19%
	Averages	***401***	***84%***	***80%***	***20%***

Table 2
Demographic Characteristics of Grade 5
Experimental Group and Control Group Schools Included in the Study

Location	Grades	Students Enrolled	% Students Free/Reduced Lunch Programs	% Minority	% Students With Special Education Needs
EXPERIMENTAL GROUP SCHOOLS					
Urban Fringe Mid-Size City	K -06	480	35%	39%	11%
Urban Fringe Mid-Size City	K -05	230	90%	73%	19%
Mid-Size Central City	K -06	612	49%	52%	19%
Mid-Size Central City	K -05	289	78%	83%	20%
Mid-Size Central City	K -05	187	49%	50%	27%
Urban Fringe Large City	K -05	207	35%	42%	19%
	Averages	***334***	***56%***	***57%***	***19%***
CONTROL GROUP SCHOOLS					
Mid-Size Central City	K -06	612	49%	52%	19%
Mid-Size Central City	K -05	289	58%	53%	20%
Mid-Size Central City	Pk-06	533	70%	70%	21%
	Averages	***478***	***59%***	***58%***	***20%***

Description of the Assessments

The pretest and posttest assessments used for the study were developed by researchers at ERIA. There was a different assessment developed for each grade level. Test items on the pretest were scrambled on the posttest. Each test was developed to align with the instruction and learning outcomes of the unit being taught. Table 3 shows the number of test items included on both tests. The control group was only administered the posttest.

Table 3
Number of Test Items Included in Each of the Subtests and on the Total Test
Grade 1 and Grade 5

Grade 1	
Subtests	*Number of Items*
Vocabulary	9
Multiple Meanings	5
Phonics	10
Grammar	5
Comprehension	8
Total	***37***
Grade 5	
Subtests	*Number of Items*
Vocabulary	13
Grammar	12
Comprehension	10
Total	***35***

Table 4 provides the test reliabilities for the pretests and posttests administered at both grades. As can be seen in the tables, the total score reliabilities at both grades for both the pretests and posttests were quite high. Most notably, the posttest reliabilities for the four tests averaged .84 and none of the reliabilities was below .82. The reliabilities indicate that confidence can be placed in the results of the tests and the data analyses which use the test results.

Table 4
Experimental and Control Groups Pretest and Posttest Reliability Statistics
Grade 1 and Grade 5

Test	Group	Pretest Reliability*	Posttest Reliability*
Grade 1	Experimental	.89	.88
Grade 1	Control	**	.90
Grade 5	Experimental	.79	.82
Grade 5	Control	**	.83

**Kuder-Richardson 20*
***The Control Group was administered a posttest only*

Data Analyses

Data analyses were conducted separately for each grade. The <.05 level of significance was used as the level at which increases would be considered statistically significant for all of the statistical tests.

Statistical analyses were conducted at each grade level to compare the control groups to the experimental groups:

- An Analysis of Variance (ANOVA) was used to compare the posttest average percent correct scores of the control group to the posttest average percent correct scores of the experimental group.

Statistical analyses were conducted at each grade level to compare the pretest and posttest scores of the experimental groups:

- A paired comparison *t*-test was used to compare the percent correct pretest scores to the percent correct posttest scores for the entire sample.
- Paired comparison *t*-tests were used to compare the percent correct pretest and posttest scores of the ELL students, as well as the percent correct pretest and posttest scores of the non-ELL students.
- The total experimental group at each grade level was divided into two groups based on their pretest scores – those that scored highest on the pretests and those that scored lowest on the pretests. Paired comparison *t*-tests were used to compare the percent correct pretest and posttest scores of the students in the lower scoring pretest group, as well as the percent correct pretest and posttest scores of the higher scoring pretest group.

An effect-size analysis was computed for each of the paired comparisons. 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 1 Results

Control Group/Experimental Group Posttest Analyses

Researchers at ERIA conducted an Analysis of Variance (ANOVA) to determine if the differences in posttest scores between the control group and the experimental group at grade 1 were significantly different. The total test included 37 items (worth one point each) which was an adequate length to conduct an ANOVA. The .05 level of significance was used as the level at which differences would be considered statistically significant. For these analyses, 137 students were included in the experimental group and 65 students were included in the control group.

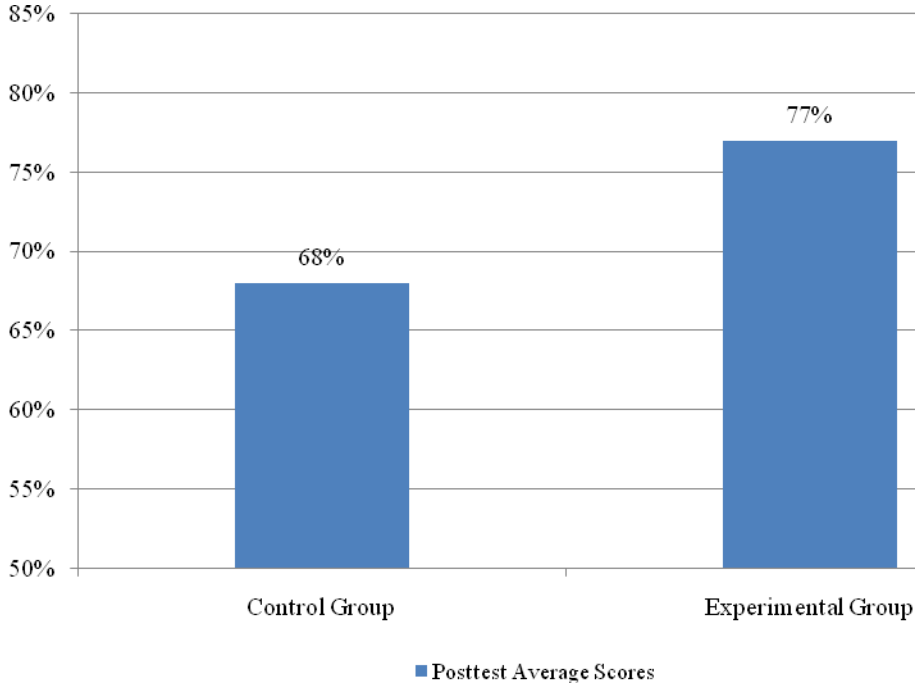
Table 5 presents the results of the ANOVA performed to determine if posttest scores of the grade 1 control group were significantly different from the posttest scores of the grade 1 experimental group. The average percent correct score on the posttest for the control group was 67.9% and for the experimental group was 77.4%, a difference that was statistically significant at the .0001 level. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The effect size was medium.

Table 5
ANOVA Results Comparing the Total Test Percent Correct Scores of the Control Group and the Experimental Group on the Posttest
Grade 1

Test	***Group***	***Number of Students***	***Mean Percent Correct Score***	***SD***	**F Test**	***Significance***	***Effect Size***
Posttest	Control	65	67.9%	20.2%	13.281	<.0001	.52
Posttest	Experimental	137	77.4%	15.8%			

Figure 1 provides a comparison of the average posttest scores of the students in the experimental and control groups. The experimental group averaged 9% higher than the control group on the posttests.

Figure 1
Comparison of Posttest Average Scores for Control and Experimental Groups
Grade 1



Pretest/Posttest Analyses

Total Experimental Group Pretest to Posttest Comparisons

Table 7 shows the average pretest and posttest scores for the grade 1 experimental group. The differences were analyzed using a paired comparison *t*-test to determine if the students in the experimental group (including both the ELL and non-ELL students) made significant gains. The increase from pretest to posttest in the mean percent correct scores of the experimental group was significant at the.0001. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The effect size was medium.

Table 7
Paired Comparison *t*-test Results Comparing the
Experimental Group's Pretest and Posttest Total Test Percent Correct Scores
Grade 1

Test	***Group***	***Number Students***	***Mean Percent Correct Score***	***SD***	**t-*test***	***Significance***	***Effect Size***
Pretest	Experimental	137	68.0%	19.3%	6.369	<.0001	.53
Posttest	Experimental	137	77.4%	15.8%			

ELL and Non-ELL Experimental Group Pretest/Posttest Comparisons

Table 8 presents the results of two paired comparison *t*-tests conducted to determine whether the pretest to posttest gains of both the ELL experimental group students and the non-ELL experimental group students were significant. Both groups' average scores increased in a way that was statistically significant. The increase of the non-ELL students' scores was significant at the .0001 level. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The increase of the ELL students' scores was significant at the .005 level. This level of significance indicates that such a difference would have occurred by chance less than 5 times out of 1,000 repetitions. The effect sizes for both groups were medium, although the effect size for the ELL students was larger than the effect size for the non-ELL students.

Table 8
Paired Comparison *t*-test Results Comparing the ELL and Non-ELL Experimental Groups' Pretest and Posttest Percent Correct Scores on the Total Test
Grade 1

Test	***Group***	***Number of Students***	***Mean Percent Correct Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Experimental ELL Students							
Pretest	ELL Only	33	57.0%	18.9%	2.994	<.005	.75
Posttest	ELL Only	33	70.6%	17.2%			
Experimental Non-ELL Students							
Pretest	Non-ELL	104	71.5%	18.2%	6.194	<.0001	.50
Posttest	Non-ELL	104	79.7%	14.8%			

Figures 2 and 3 present the percentages of ELL and non-ELL students in the grade 1 experimental group scoring below 50%, from 50% to 84%, and 85% or higher on the pretests and posttests.

Figure 2 shows that for the ELL students the percentage of students scoring at the lowest level declined from pretest to posttest and that the percentage of students scoring at the middle and high levels increased by 30% from pretest to posttest while the percentage scoring at the lowest level decreased by 30%.

Figure 2
Percentage of ELL Students in the Experimental Group Scoring at Various Levels on the Pretest and Posttest
Grade 1

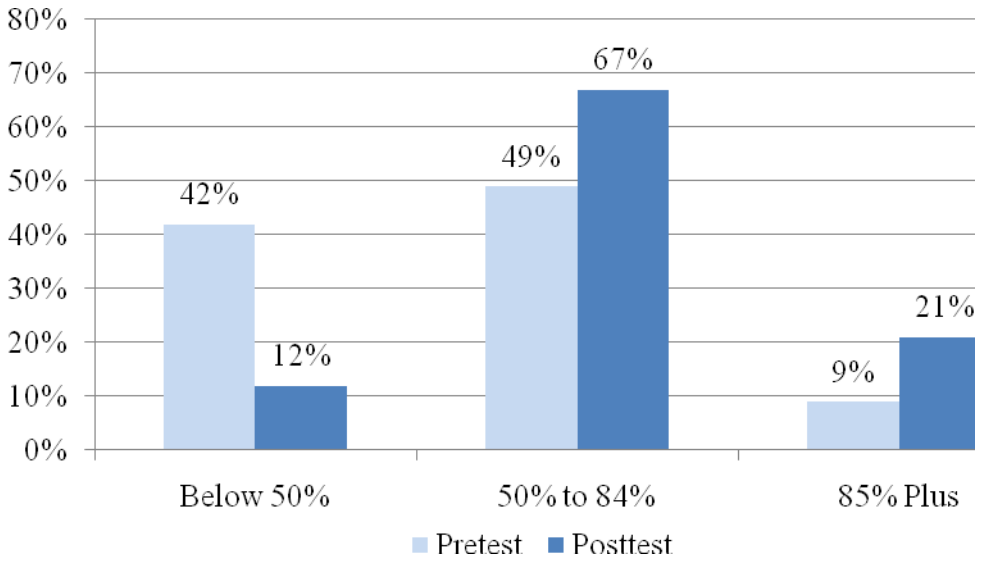
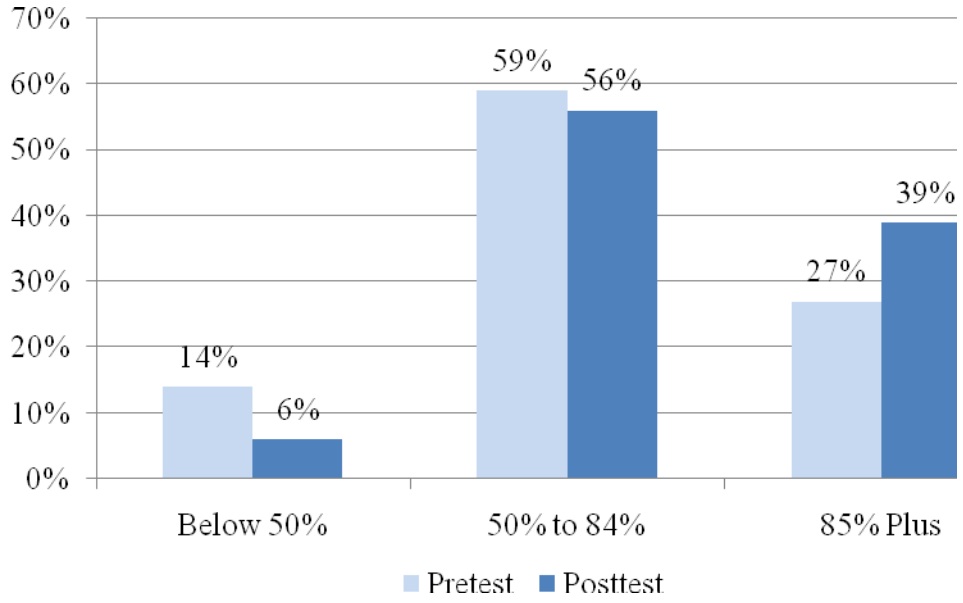


Figure 3 shows that for the non-ELL students the percentage of students scoring at the lowest levels declined from pretest to posttest and that the percentage of students scoring at the high levels increased from pretest to posttest by more than 10%.

Figure 3
Percentage of Non-ELL Students Scoring at Various Levels on the Pretest and Posttest
Grade 1



Experimental Group Pretest/Posttest Comparisons for High and Low Pretest Scorers

Table 9 provides a comparison of the pretest and posttest scores for the high and low pretest scorers. A paired comparison *t*-test was used to determine whether the pretest to posttest gains of the students who scored the lowest on the pretests were as large as those who scored the highest on the pretests. The students were ranked from lowest to highest based on their pretest scores. The lowest 68 students were considered the low pretest group and the highest 69 students were considered the high pretest group. The pretest to posttest gain made by the low pretest student group was significant at the .0001 level. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The effect size for the low pretest group was large.

The pretest to posttest gain made by the high pretest student group was not statistically significant. Part of the reason for the non-significant gain was that the high scoring students scored fairly high on the pretests and there was therefore little margin for gain from pretest to posttest.

Table 9
Paired Comparison *t*-test Results Comparing the Lower and Higher Scoring Pretest Groups' Pretest and Posttest Percent Correct Scores on the Total Test
Grade 1

Test	***Group***	***Number of Students***	***Mean Percent Correct Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Low Pretest Students							
Pretest	Total Experimental	68	52.1%	14.0%	7.690	<.0001	1.16
Posttest	Total Experimental	68	69.7%	16.1%			
High Pretest Students							
Pretest	Total Experimental	69	83.7%	7.1%	1.051	Non-Significant	--
Posttest	Total Experimental	69	85.1%	11.2%			

Figures 4 and 5 compare the percentage of grade 1 students in the lower and in the higher scoring pretest groups scoring below 50%, from 50% to 84%, and 85% or higher on the pretests and posttests.

Figure 4 shows that for the lower scoring pretest group, the percentage of students scoring at the lowest levels declined by 30% from pretest to posttest. The percentage of lower scoring pretest students scoring at the middle and higher levels both increased.

Figure 4
Percentage of Low Pretest Scoring Students Scoring at Various Levels on the Pretest and Posttest
Grade 1

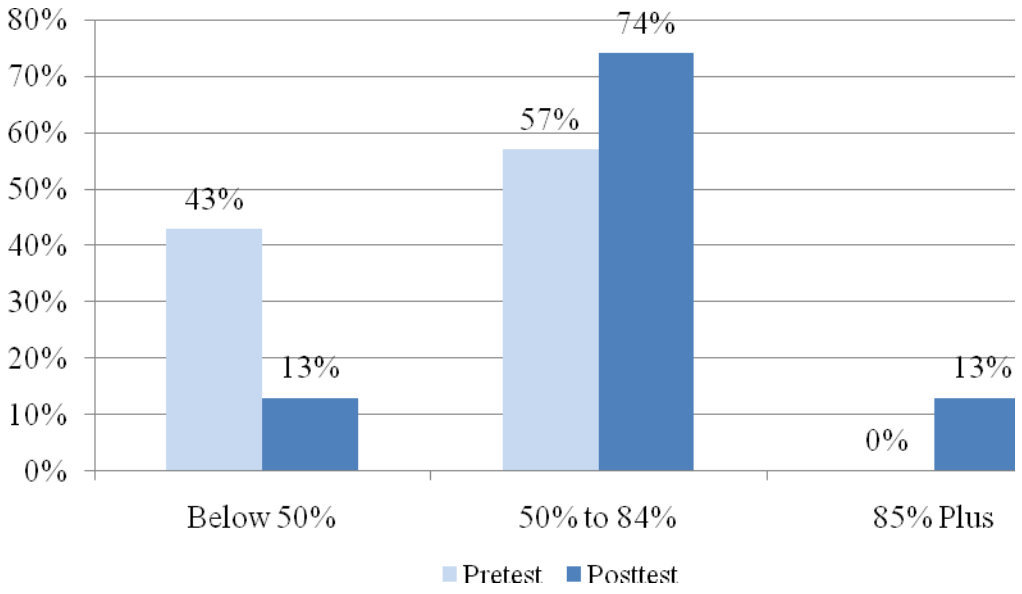
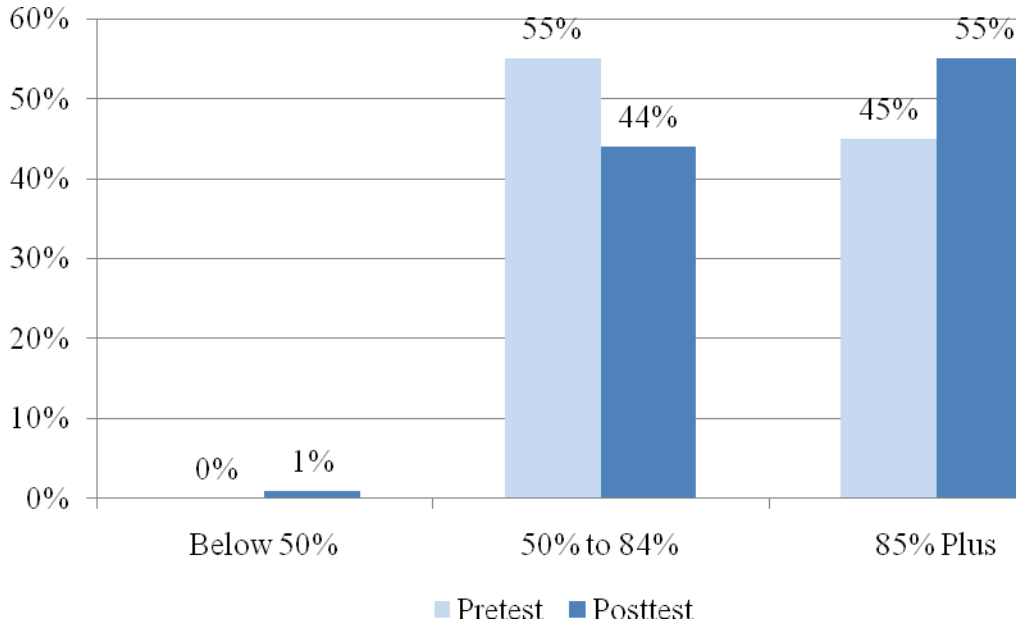


Figure 5 shows that for the higher scoring pretest group, the percentage of students scoring at the highest level increased by 10% from pretest to posttest.

Figure 5
Percentage of High Pretest Scoring Students Scoring at Various Levels on the Pretest and Posttest
Grade 1



Grade 5 Results

Control Group/Experimental Group Posttest Analyses

Researchers at ERIA conducted an ANOVA to determine if the differences in posttest scores between the control group and the experimental group at grade 5 were significantly different. The total test included 35 items (worth one point each) which was an adequate length to conduct an ANOVA. The .05 level of significance was used as the level at which differences would be considered statistically significant. For these analyses, 133 students were included in the experimental group and 74 students were included in the control group.

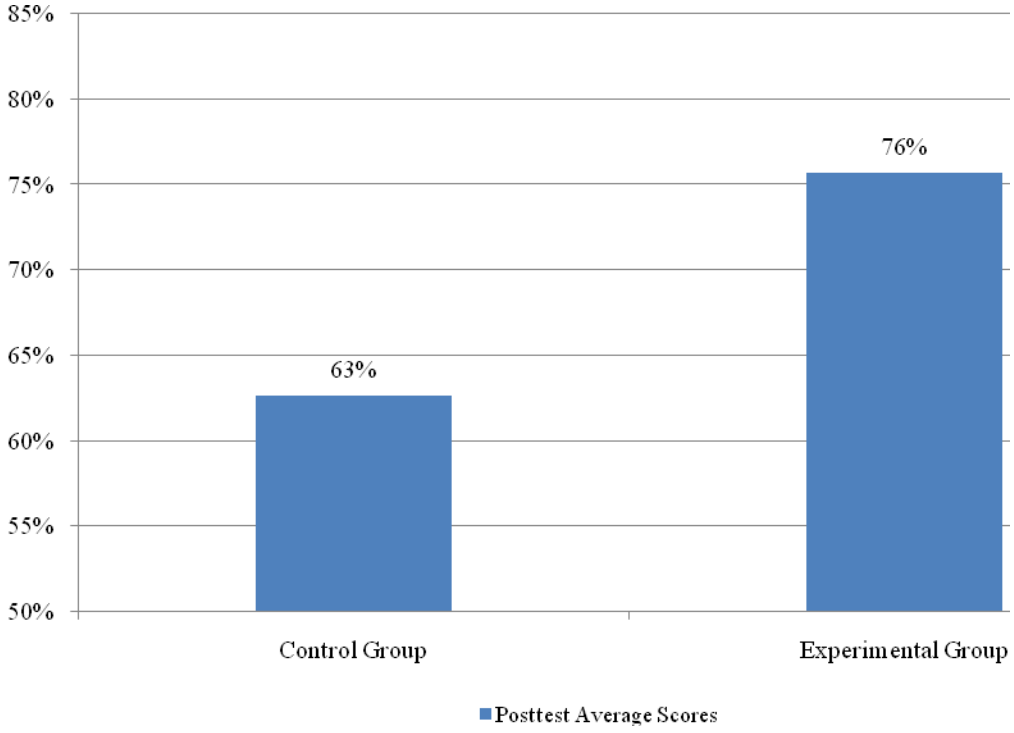
Table 10 presents the results of the ANOVA performed to determine if posttest scores of the grade 5 control group were significantly different from the posttest scores of the grade 5 experimental group. The average percent correct score on the posttest for the control group was 62.7% and for the experimental group was 73.9%, a difference that was statistically significant at the .0001 level. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The effect size was medium.

Table 10
ANOVA Results Comparing the Total Test Percent Correct Scores of the Control Group and the Experimental Group on the Posttest
Grade 5

Test	*Group*	*Number Students*	*Mean Score*	*SD*	**F Test**	*Significance*	*Effect Size*
Posttest	Control	74	62.7%	16.7%	24.271	<.0001	.69
Posttest	Experimental	133	73.9%	15.0%			

Figure 6 provides a comparison of the average posttest scores of the students in the experimental and control groups. The experimental group averaged 13% higher than the control group on the posttests.

Figure 6
Comparison of Posttest Average Scores for
Control and Experimental Groups of Grade 5 Students



Pretest/Posttest Analyses

Total Experimental Group Pretest to Posttest Comparisons

Table 12 shows the average pretest and posttest scores for the grade 5 experimental group. The differences were analyzed using a paired comparison *t*-test to determine if the students in the experimental group (including both the ELL and non-ELL students) made significant gains. The increase from pretest to posttest in the mean percent scores of the experimental group was significant at the .0001. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The effect size was medium.

Table 12
Paired Comparison *t*-test Results Comparing the Experimental Group's Pretest and Posttest Total Test Percent Correct Scores
Grade 5

Test	*Group*	*Number Students*	*Mean Percent Correct Score*	*SD*	t-*test*	*Significance*	*Effect Size*
Pretest	Experimental	133	66.8%	14.8%	7.368	<.0001	.51
Posttest	Experimental	133	73.9%	15.0%			

ELL and Not-ELL Experimental Group Pretest/Posttest Comparisons

Table 13 presents the results of two paired comparison *t*-tests conducted to determine whether the pretest to posttest gains of both the ELL experimental group students and the non-ELL experimental group students were significant. Both groups' average scores increased in a way that was statistically significant. The increase of the non-ELL students' scores was significant at the .0001 level. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The increase of the ELL students' scores was significant at the .001 level. This level of significance indicates that such a difference would have occurred by chance less than once out of 1,000 repetitions. The effect size for ELL group was medium and small for the non-ELL group.

Table 13
Paired Comparison *t*-test Results Comparing the ELL and Non-ELL Experimental Groups' Pretest and Posttest Percent Correct Scores on the Total Test
Grade 5

Test	***Group***	***Number of Students***	***Mean Percent Correct Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Experimental ELL Students							
Pretest	ELL Only	25	54.6%	15.6%	3.656	<.001	.61
Posttest	ELL Only	25	65.8%	15.0%			
Experimental Non-ELL Students							
Pretest	Non-Ell	108	69.7%	13.1%	6.570	<.0001	.44
Posttest	Non-Ell	108	75.8%	14.4%			

Figures 7 and 8 present the percentages of ELL and non-ELL students in the grade 1 experimental group scoring below 50%, from 50% to 84%, and 85% or higher on the pretests and posttests.

Figure 7 shows that for the ELL students the percentage of students scoring at the lowest levels declined from pretest to posttest and that the percentage of students scoring at the middle and high levels increased from pretest to posttest.

Figure 7
Percentage of ELL Students in the Experimental Group Scoring at Various Levels on the Pretest and Posttest
Grade 5

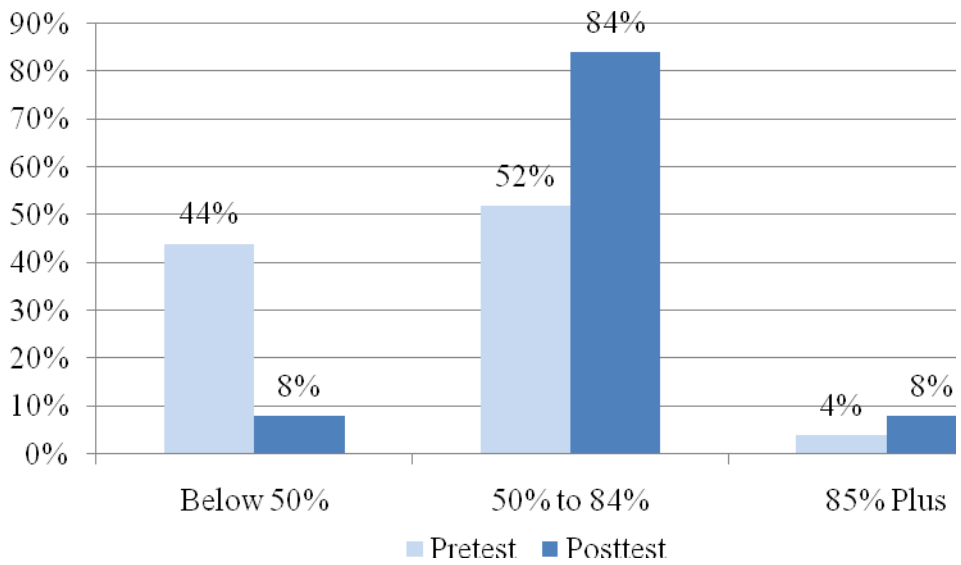
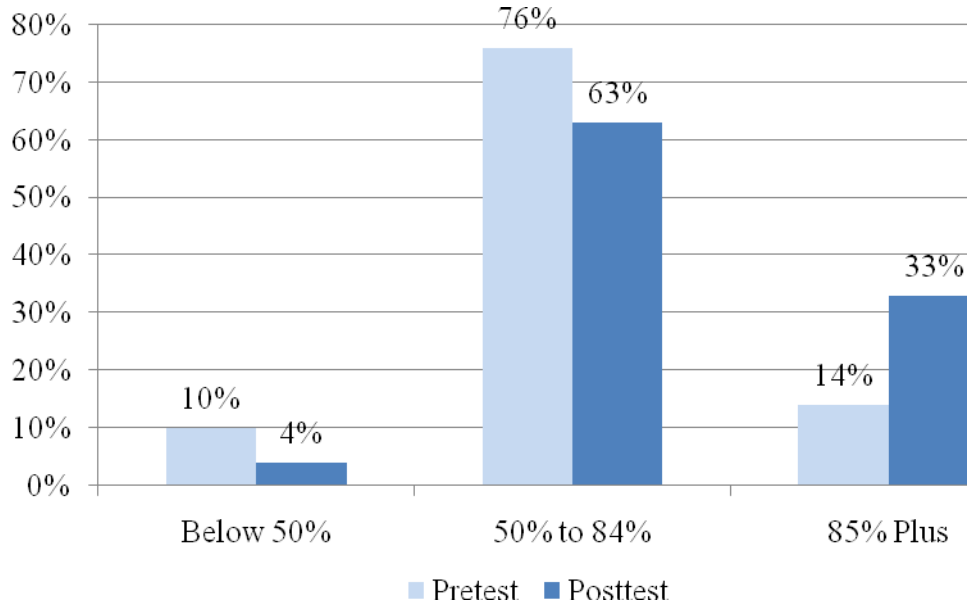


Figure 8 shows that for the non-ELL students the percentage of students scoring at the lowest levels was cut in half and that the percentage of students scoring at the highest level increased by 19% from pretest to posttest.

Figure 8
Percentage of Non-ELL Students Scoring at Various Levels on the Pretest and Posttest
Grade 5



Experimental Group Pretest/Posttest Comparisons for High and Low Pretest Scorers

Table 14 provides a comparison of the pretest and posttest scores for the high and low pretest scorers. A paired comparison *t*-test was used to determine whether the pretest to posttest gains of the students who scored the lowest on the pretests were as large as those who scored the highest on the pretests. The students were ranked from lowest to highest based on their pretest scores. The lowest 66 students were considered the low pretest group and the highest 67 students were considered the high pretest group. The pretest to posttest gains made by both the low and the high pretest student groups were significant at the .0001 level. This level of significance indicates that such a difference would have occurred by chance less than once out of 10,000 repetitions. The effect sizes for both groups were large.

Table 14
Paired Comparison *t*-test Results Comparing the Lower and Higher Scoring Pretest Groups' Pretest and Posttest Percent Correct Scores on the Total Test
Grade 5

Test	*Group*	*Number of Students*	*Mean Percent Correct Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Low Pretest Students							
Pretest	Total Experimental	66	54.8%	10.0%	6.143	<.0001	.82
Posttest	Total Experimental	66	64.3%	13.0%			
High Pretest Students							
Pretest	Total Experimental	67	78.7%	7.3%	4.362	<.0001	1.20
Posttest	Total Experimental	67	83.2%	10.2%			

Figures 9 and 10 compare the percentage of grade 5 students in the lower and in the higher scoring pretest groups scoring below 50%, from 50% to 84%, and 85% or higher on the pretests and posttests.

Figure 9 shows that for the lower scoring pretest group, the percentage of students scoring at the lowest levels declined by more than 20% from pretest to posttest. The percentage of lower scoring pretest students scoring at the middle and higher levels both increased.

Figure 9
Percentage of Low Pretest Scoring Students Scoring at Various Levels on the Pretest and Posttest
Grade 5

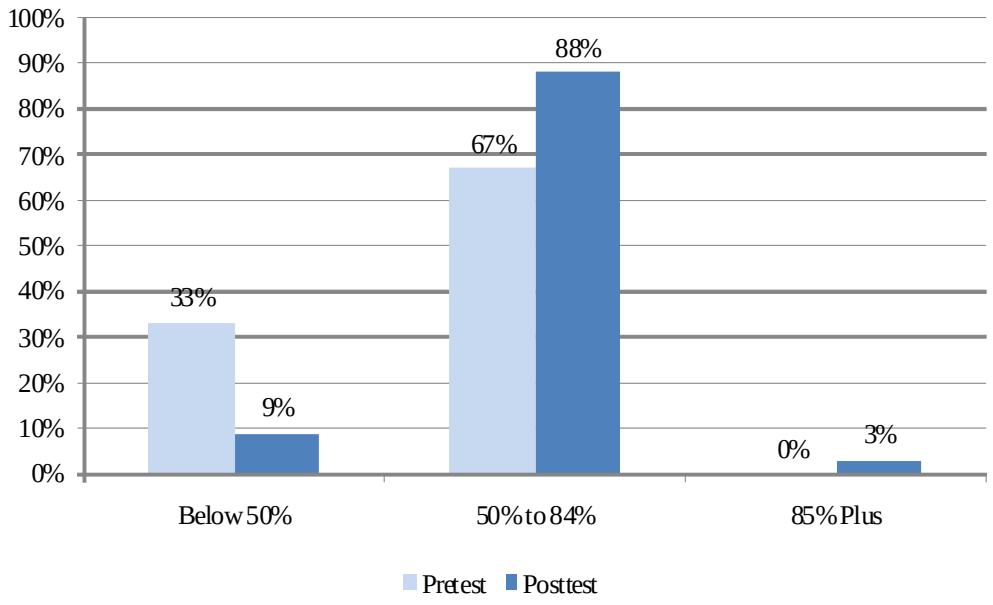
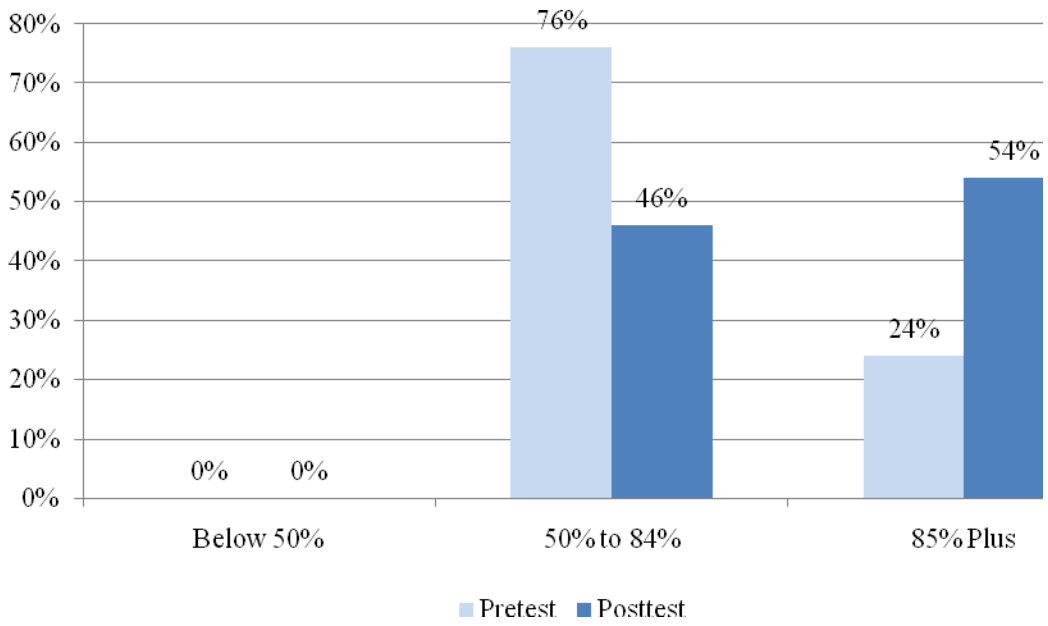


Figure 10 shows that for the higher scoring pretest group, the percentage of students scoring at the highest level increased by 30% from pretest to posttest.

Figure 10
Percentage of High Pretest Scoring Students Scoring at Various Levels on the Pretest and Posttest
Grade 5



Conclusions

This study sought to determine if *HOUGHTON MIFFLIN HARCOURT JOURNEYS used* in 19 different schools by 22 different teachers' increased reading achievement for students in grades 1 and 5. In addition, the study analyzed whether ELL students also made significant gains.

Control Group/Experimental Group Posttest Analyses

Table 15 provides a summary of the ANOVA results comparing the control groups to the experimental groups. At both grade levels, the experimental total groups scored significantly higher on the posttests than did the control groups. In addition, at both grades the ELL students' posttest average percent correct scores exceeded the control group students' posttest average percent correct scores.

Table 15
Experimental Group/Control Group Summary of Gains
Grade 1 and Grade 5

	Mean	Statistical Significance	Effect Size
Grade 1			
Experimental Group - Total	77.4	<.0001	Medium
Control Group Students – Total	67.9		
Grade 5			
Experimental Group - Total	73.9	<.0001	Medium
Control Group Students – Total	62.7		

Pretest/Posttest Analyses

As can be seen in Table 16, for the experimental group students at both grade 1 and grade 5, significant pretest to posttest gains were made for the total test scores. As well, the ELL and Non-ELL students made significant gains as did the high and low pretest groups. The only exception to the statistically significant gains was for the high pretest students at grade 1. A reason for this non-significance, even though the scores did increase, was that the high pretest students had already scored at a high level on the pretests and there was little room for them to score higher on the posttests.

Table 16
Experimental Group Pretest to Posttest Summary of Gains
Grade 1 and Grade 5

	Statistical Significance	*Effect Size*
Grade 1		
All Experimental Students	<.0001	Medium
ELL Students	<.005	Medium
Non-ELL Students	<.0001	Medium
High Pretest Students	Not-Significant	--
Low Pretest Students	<.0001	Large
Grade 5		
All Experimental Students	<.0001	Medium
ELL Students	<.001	Medium
Non-ELL Students	<.0001	Small
High Pretest Students	<.0001	Large
Low Pretest Students	<.0001	Large

The conclusion based on a reliable test designed to measure growth on the reading strategies and skills taught in a single unit of instruction to students who received instruction using *HOUGHTON MIFFLIN HARCOURT JOURNEYS, is* that the program significantly increases students' reading skills and strategy use. The scores of students in the study who received instruction using *HOUGHTON MIFFLIN HARCOURT JOURNEYS, increased* statistically significantly. The results for the ELL students were equally impressive. These students as a group scored higher than the control group students and showed statistically significant growth from pretesting to postesting.

These results were all the more remarkable and significant considering the very short duration of the study and the fact that the teachers had never used the program before.