# A STUDY OF THE INSTRUCTIONAL EFFECTIVENESS OF
## *Houghton Mifflin Harcourt*
## *Florida Journeys  Common Core* ©2014
Report Number 435
July 2012

**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

# Contents

## ABSTRACT

*The importance of effective reading of literature and informational text is crucial if a student is to become an effective reader of all types of reading materials. Thoughtful reading is fundamental to becoming a thoughtful and effective citizen. To help elementary school students develop better reading skills and strategies, Houghton Mifflin Harcourt is publishing,* **Houghton Mifflin Harcourt Florida Journeys Common Core ©2014***, a reading program for literature and informational reading, thinking, and analysis.*

*In order to evaluate the program's effectiveness, Houghton Mifflin Harcourt contracted with the Educational Research Institute of America to conduct a field study to test the effectiveness of one unit of the program. The study was conducted at the end of the 2011/2012 academic year with classes at grades 2, 3, and 4.*

*A test was designed to assess students' literature and informational reading skills and strategies. The* **Houghton Mifflin Harcourt Florida Journeys Common Core ©2014** *program had not been previously used in the schools by any classes.*

*The results showed that the* **Houghton Mifflin Harcourt Florida Journeys Common Core ©2014** *classes made statistically significant gains over the course of the four to six week tryout. The results also showed the* **Houghton Mifflin Harcourt Florida Journeys Common Core ©2014** *program proved equally effective with both higher and lower pretest scoring students.*

## Overview of the Study

This report describes a field study conducted to determine the impact of the *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014*   program for elementary school students.  The program addresses the Common Core State Standards in English Language Arts. The program includes both literary and informational texts that progressively develop and apply students' reading skills and strategies.

For this study a single unit was selected for the field study at grades 2, 3, and 4. The unit chosen for study at each grade level was Unit 4. An entire Unit in the program covers about 6 weeks of instruction and due to the time of year, it was unlikely that an entire unit could be taught in the period of time available, particularly with the additional time needed for pretesting and post-testing. Therefore, teachers were asked to be sure to cover the first two lessons of the five lessons in the unit. They were also told to continue teaching the other lessons if they had time to complete them.

## Project Background

The following focus for the program as put forth by the publisher highlights the importance of a research/best practices based program:

> *Houghton Mifflin Harcourt Publishing is committed to partnering with districts to align teaching and learning programs with Common Core State Standards.* ***Houghton Mifflin Harcourt Florida Journeys Common Core ©2014*** *offers implementation options that blend print and digital instructional materials with professional development to ensure a smooth transition to the Common Core. This program includes both literature and informational selections that align with the new Common Core State Standards.*

> *The Common Core State Standards reflect an unprecedented emphasis on the close and careful reading of both literary and informational text.* ***Houghton Mifflin Harcourt Florida Journeys Common Core ©2014*** *provides extensive informational text in each unit to help the reader reach a deeper level of comprehension.*

> *In addition to reading literary and informational texts, the program emphasizes language skills. By addressing all strands of the Common Core State Standards,* ***Houghton Mifflin Harcourt Florida Journeys Common Core ©2014*** *prepares students for every aspect of communication in the 21st Century.*

## Research Questions

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

1. Is *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014 effective* in improving the reading and language skills of elementary school students?
2. Is *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014* effective in improving the reading and language skills of lower performing as well as higher performing elementary school students?

## Design of the Study

The program's efficacy was evaluated using a pretest/posttest design. The study took place during the second semester of the 2011/2012 academic year as an initial field test of the new *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014* program. The grades chose for the test were grades 2, 3, and 4. The research study spanned about 6 weeks which included the time for pretesting and posttesting.

Before program instruction, students were administered a comprehensive test designed to cover the content of the unit included in the study. A similar posttest was used at the end of the study.

### Study Participants

The study was conducted in three different schools in three different states. The program effectiveness data reported here is based on a sample which included the following number of teachers and schools:

Grade 2: 10 teachers in three different schools in three different states
Grade 3: 9 teachers in three different schools in three different states
Grade 4: 7 teachers in three different schools in three different states

All of the teachers reported that they had teaching experience of more than 5 years.

### Timeline and Program Use

All of the teachers used the program for between about 6 weeks at the end of the second semester of the 2011-2012 academic school years. Pretests were administered at the end of April and posttests were administered in some sites as late as mid-June. This was the first time the teachers had used the program and most were unfamiliar with the program prior to the tryout. For the test period, teachers reported using *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014* as their primary reading/language arts program.

## 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 each of the classes that participated in the study. However, the data does provide a general description of each of the schools and, thereby, an estimate of the make-up of the classes included in the study.

States included one each in the Northeast, Midwest, and Southwest United States.

**Table 1**
**Demographic Characteristics**
**Of the Schools Included in the Study**

| Location | Grades | Enrollment | % Minority | % Students Free/Reduced Lunch Programs | % Special Education |
|---|---|---|---|---|---|
| Suburban | K to 5 | 430 | 56% | 23% | 6% |
| Suburban | K to 8 | 692 | 27% | 9% | 10% |
| Suburban | K to 5 | 409 | 9% | 49% | 9% |
| **AVERAGES** | | **510** | **31%** | **27%** | **8%** |

　Educational Research Institute of America

## Description of the Assessment

The pretests and posttest used in the study were developed to assess students' reading skills and strategies. Based on the instructional objectives listed in the teacher's editions for each unit, a test was developed for each grade level included in the study. The assessments focused on the first two lessons in unit four at each grade level. Each assessment was made up of multiple-choice test items and focused on assessing students' understanding of a series of literary and informational selections.

The reliability and standard error of measurement of each assessment are reported in Table 2. The reliabilities of the assessments average .88 which is quite high and provide results which can be used to compare student achievement from pretesting to post-testing.

**Table 2**
**Reliability Estimates, Mean Scores and Reliability**
**for the Florida Journeys Unit 4 Assessment at Grades 2, 3, and 4**

| Test | Posttest Results | | |
|---|---|---|---|
| | Standard Deviation | Reliability* | SEM** |
| Grade 2 Pretest | 6.84 | .90 | 2.16 |
| Grade 2 Post-Test | 6.17 | .89 | 2.05 |
| Grade 3 Pretest | 6.56 | .86 | 2.45 |
| Grade 3 Post-Test | 6.56 | .87 | 2.37 |
| Grade 4 Pretest | 7.08 | .88 | 2.45 |
| Grade 4 Post-Test | 7.42 | .90 | 2.35 |

*Reliability computed using the Kuder-Richardson 20 formula.*
*** SEM stands for Standard Error of Measurement.*

## Data Analyses

Data analyses and descriptive statistics were computed for the standard scores from the **Houghton Mifflin Harcourt Florida Journeys Common Core ©2014** assessments used at each grade. Standard scores were developed in order 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.

The <.05 level of significance was used as the level at which increases would be considered statistically significant for all of the statistical tests.

The following statistical analyses were conducted to compare students' pretest scores to posttest scores at grades 2, 3, and 4:

- A paired comparison *t*-test was used to compare the pretest mean standard scores with the posttest mean standard scores for all students.

- The students at each grade level 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 lower performers and higher performers.

Descriptive statistics were also used to compare pretest and posttest standard test scores at all three grades included in the study.

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

Educational Research Institute of America

# Grade Two Data and Analyses

## Total Group Analysis

Researchers at ERIA conducted a paired comparison *t*-test to determine if the difference from pretest standard scores to posttest standard scores was statistically significant. For this analysis, researchers were able to match the pretest and posttest scores for 267 students. Students who did not take both the pretest and the posttest were not included.

Table 3 shows that the average standard score on the pretest was 291, and the average standard score on the posttest was 309. The increase was statistically significant (≤.0001). The effect size was small.

**Table 3**
**Grade 2 Paired Comparison *t*-test Results**
**Pretest/Posttest Comparison of Standards Scores**

| Test | Number Students | Mean Standard Score | SD | *t*-test | Significance | Effect Size |
|------|-----------------|---------------------|------|----------|--------------|-------------|
| Pretest | 267 | 291.4 | 53.2 | 6.674 | ≤.0001 | .35 |
| Posttest | 267 | 308.6 | 45.0 | | | |

Figure 1 provides a pretest-to-posttest comparison of the percentage of students who scored less than 50% correct, from 50% to 84% correct and 85% or higher on both the pretests and post-tests. The percentage of lower and medium scoring students decreased while the percentages of the medium and high scoring students increased.

**Figure 1**
**Percentage of Grade 2 Students Scoring Low, Medium, and High**
**On the Pretests and Post-Tests**
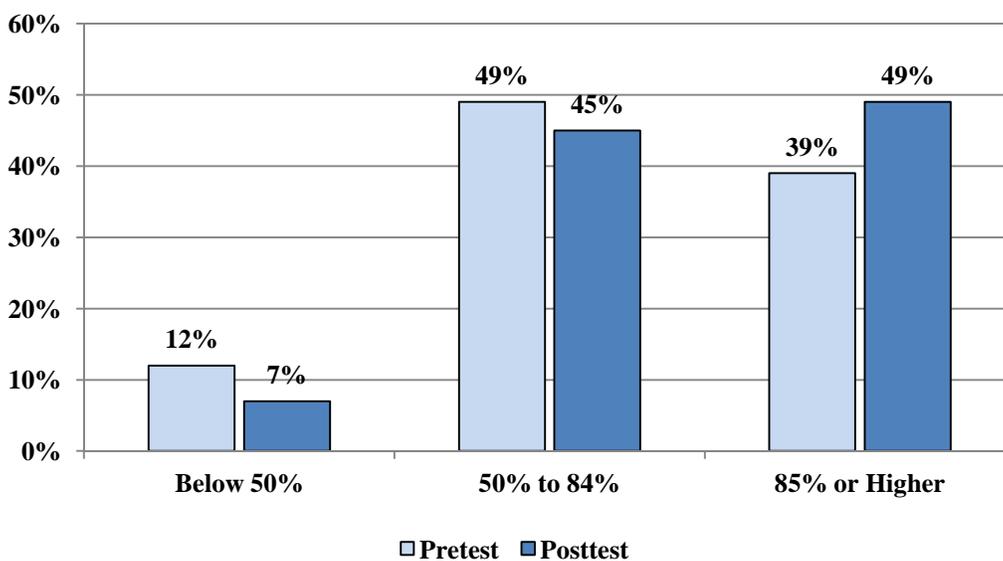
## Higher and Lower Scoring Students

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 on the basis of their pretest standard scores. The group of 267 students was divided into two groups of 133 and 134 students. The first group included those students who scored lower on the pretest with a mean of 251 with scores ranging from 130 to 308. The higher scoring group scored an average standard score on the pretest of 332 with scores ranging from 308 to 362.

Table 4 presents the analysis of student growth for the lower and higher pretest scoring students. Pretest-to-posttest comparisons, using paired-wise t-tests, indicated that the low pretest scoring group's average scores increased statistically significantly (≤.0001) and the effect size was medium. However, the higher pretest scoring group did not have a significant increase in test scores.

A major part of the reason the higher posttest scoring group did not increase from pretesting to post-testing is that the pretest scores of the higher scoring group were already quite high so there was a partial ceiling effect.[1] A total of 52 higher scoring pretest students had pretest scores of 90% or higher and 6 had scores of 100%. On the post-tests the higher scoring group included 116 students with scores of 90% or higher and 13 of these students had scores of 100%. They showed some growth but the fact is that there was not much higher they could go.

**Table 4**
**Grade 2 Paired Comparison *t*-test Results for Pretest/Posttest Standard Scores**
**For the High- and Low-Scoring Pretest Groups**

| Test | Test Form | Number of Students | Mean Standard Score | SD | t-test | Significance | Effect Size |
|------|-----------|--------------------|---------------------|-----|--------|--------------|-------------|
| **Lower Scoring Group** | | | | | | | |
| Total | Pretest | 133 | 250.9 | 46.8 | 8.222 | ≤.0001 | .72 |
| Total | Posttest | 133 | 285.3 | 49.8 | | | |
| **Higher Scoring Group** | | | | | | | |
| Total | Pretest | 134 | 331.7 | 14.8 | 0 | Non-Significant | -- |
| Total | Posttest | 134 | 331.7 | 22.6 | | | |

---

[1] *An undesirable measurement outcome occurring when the dependent measure puts an artificially lower ceiling on how high a participant may score.*

## Grade Three Data and Analyses

### Total Group Analysis

Researchers at ERIA conducted a paired comparison *t*-test to determine if the difference from pretest standard scores to posttest standard scores was statistically significant. For this analysis, researchers were able to match the pretest and posttest scores for 131 students. Students who did not take both the pretest and the posttest were not included.

Table 5 shows that the average standard score on the pretest was 293, and the average standard score on the posttest was 307. The increase was statistically significant ($\leq$.0001). The effect size was small.

**Table 5**
**Grade 3 Paired Comparison *t*-test Results**
**Pretest/Posttest Comparison of Standards Scores**

| *Test* | *Number Students* | *Mean Standard Score* | *SD* | *t*-test | *Significance* | *Effect Size* |
|---|---|---|---|---|---|---|
| Pretest | 131 | 292.9 | 51.0 | 7.415 | $\leq$.0001 | .28 |
| Posttest | 131 | 307.1 | 48.1 | | | |

Figure 2 provides a pretest-to-posttest comparison of the percentage of students who scored less than 50% correct, from 50% to 84% correct and 85% or higher on both the pretests and post-tests. The percentage of lower and medium scoring students decreased 14% while the percentage of the high scoring students increased 14%.

**Figure 2**
**Percentage of Grade 3 Students Scoring Low, Medium, and High**
**On the Pretests and Post-Tests**
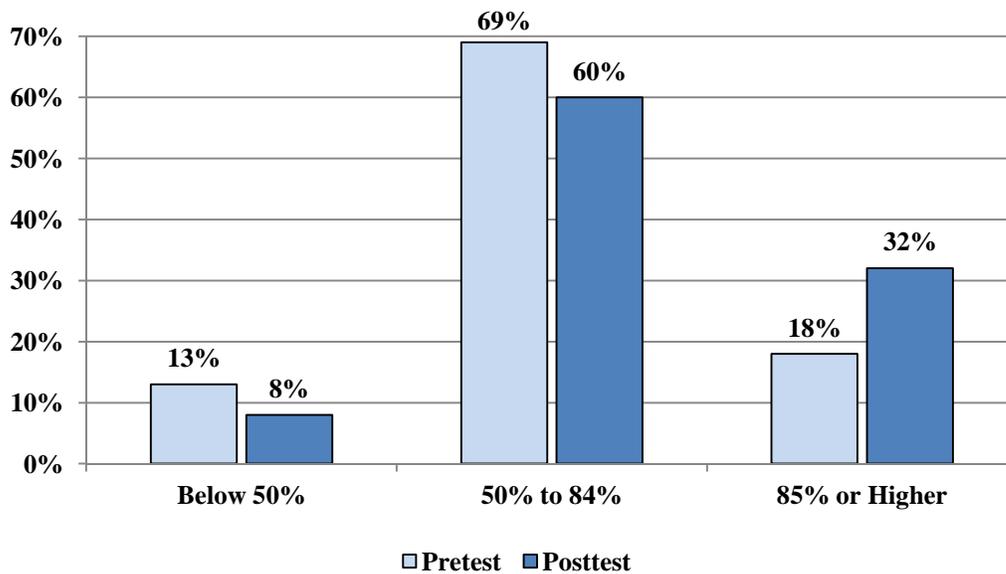
## Higher and Lower Scoring Students

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 on the basis of their pretest standard scores. The group of 131 students was divided into two groups of 65 and 66 students. The first group included those students who scored lower on the pretest with a mean of 254 with scores ranging from 137 to 308. The higher scoring group scored an average standard score on the pretest of 331 with scores ranging from 316 to 371.

Pretest-to-posttest comparisons are shown in Table 6 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 lower and higher scoring group, the average scores increase was statistically significant (≤.0001). The effect size for each group was small.

**Table 6**
**Grade 3 Paired Comparison *t*-test Results for Pretest/Posttest Standard Scores**
**For the High- and Low-Scoring Pretest Groups**

| Test | Test Form | Number of Students | Mean Standard Score | SD | t-test | Significance | Effect Size |
|------|-----------|--------------------|---------------------|-----|--------|--------------|-------------|
| **Lower Scoring Group** | | | | | | | |
| Total | Pretest | 65 | 254.3 | 45.7 | 6.740 | ≤.0001 | .45 |
| Total | Posttest | 65 | 275.2 | 47.7 | | | |
| **Higher Scoring Group** | | | | | | | |
| Total | Pretest | 66 | 330.8 | 14.2 | 3.848 | ≤.0001 | .45 |
| Total | Posttest | 66 | 338.5 | 19.4 | | | |

# Grade Four Data and Analyses

## Total Group Analysis

Researchers at ERIA conducted a paired comparison *t*-test to determine if the difference from pretest standard scores to posttest standard scores was statistically significant. For this analysis, researchers were able to match the pretest and posttest scores for 118 students. Students who did not take both the pretest and the posttest were not included.

Table 7 shows that the average standard score on the pretest was 294, and the average standard score on the posttest was 307. The increase was statistically significant ($\leq$.0001). The effect size was small.

**Table 7**
**Grade 4 Paired Comparison *t*-test Results**
**Pretest/Posttest Comparison of Standards Scores**

| Test | Number Students | Mean Standard Score | SD | *t*-test | Significance | Effect Size |
|------|-----------------|---------------------|------|----------|--------------|-------------|
| Pretest | 131 | 293.5 | 50.0 | 5.807 | $\leq$.0001 | .26 |
| Posttest | 131 | 306.5 | 49.4 | | | |

Figure 3 provides a pretest-to-posttest comparison of the percentage of students who scored less than 50% correct, from 50% to 84% correct and 85% or higher on both the pretests and post-tests. The percentage of lower and medium scoring students decreased 15% while the percentage of the high scoring students increased by 15%.

**Figure 3**
**Percentage of Grade 4 Students Scoring Low, Medium, and High**
**On the Pretests and Post-Tests**
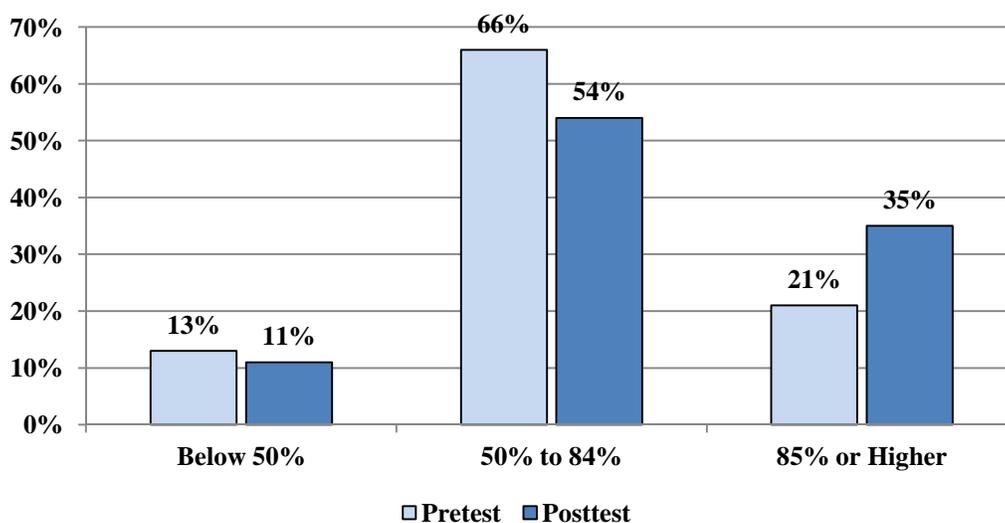
## Higher and Lower Scoring Students

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 on the basis of their pretest standard scores. The group of 118 students was divided into two groups of 59 students. The first group included those students who scored lower on the pretest with a mean of 256 with scores ranging from 162 to 303. The higher scoring group scored an average standard score on the pretest of 333 with scores ranging from 303 to 367.

Pretest-to-posttest comparisons are shown in Table 8 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 lower and higher scoring group, the average scores increase was statistically significant (≤.0001). The effect size was small.

**Table 8**
**Grade 4 Paired Comparison *t*-test Results for Pretest/Posttest Standard Scores For the High- and Low-Scoring Pretest Groups**

| Test | Test Form | Number of Students | Mean Standard Score | SD | t-test | Significance | Effect Size |
|------|-----------|--------------------|--------------------|-----|--------|--------------|-------------|
| **Lower Scoring Group** | | | | | | | |
| Total | Pretest | 59 | 255.9 | 43.2 | 4.773 | ≤.0001 | .39 |
| Total | Posttest | 59 | 274.2 | 49.5 | | | |
| **Higher Scoring Group** | | | | | | | |
| Total | Pretest | 59 | 331.1 | 17.1 | 3.595 | ≤.0001 | .44 |
| Total | Posttest | 59 | 338.9 | 18.6 | | | |

## Conclusions

This study sought to determine the effectiveness of *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014,* an elementary school reading program published by Houghton Mifflin Harcourt. The study was carried out with classes at grades 2, 3, and 4. The teachers 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 Florida Journeys Common Core ©2014 effective in improving the reading and language skills of elementary school students?*

A test which included both literature and language skills and strategies was developed to assess students at the beginning and end of a field test of the program. Statistical analyses of students' scores at three grade levels showed that the students increased their scores statistically significantly on the assessment.

*Question 2: Is Houghton Mifflin Harcourt Florida Journeys Common Core ©2014 effective in improving the reading and language skills of lower performing as well as higher performing elementary school students?*

Statistical analyses of higher and lower pretest scoring students' scores at grades 2, 3, and 4 showed that for the lower pretest scoring students the increase was statistically significant. For the higher pretest scoring students the average scores increased significantly in grades 3 and 4. However, at grade 2 there was not a significant increase in test scores. At all three grades the higher pretest group did not show as great an increase as the lower pretest scoring group. This was due in part because a number of students scored quite high on the pretest and thus there was a ceiling effect on the posttest.

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

- *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014 is* effective in improving the reading and language skills of elementary school students.
- *Houghton Mifflin Harcourt Florida Journeys Common Core ©2014* is effective in improving the reading and language skills of lower performing as well as higher performing elementary school students.