

**A STUDY OF THE INSTRUCTIONAL EFFECTIVENESS OF
HOUGHTON MIFFLIN HARCOURT'S
SCIENCEFUSION © 2012**

Report Number 380 A

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



Table of Contents

Project Background.....	2
Research Questions.....	2
Design of the Study.....	3
Instructional Approach under Study.....	4
Description of the Research Sample.....	5
Description of the Assessments.....	5
Data Analyses.....	6
Treatment Group Comparison.....	6
Subgroups by Pretest Performance—Pretest/Posttest Analyses.....	7
Conclusions.....	9

A STUDY OF THE INSTRUCTIONAL EFFECTIVENESS OF HOUGHTON MIFFLIN HARCOURT'S *SCIENCEFUSION* © 2012

This report describes an instructional efficacy study that was conducted to determine the impact of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 on students' knowledge and skills in science.

Project Background

The importance of science skills and knowledge to the future success of our students and to our nation has never been greater. The science programs that young students are using must optimally support them in developing the science skills required for success in high school, college, and their professional lives. As a nation, the United States has not competed well on international comparisons when secondary school science achievement is assessed. National leaders in science education have argued that we cannot wait until middle and secondary school for science education to get underway. If our students are to compete effectively, effective science instruction must begin in the elementary school and continue in the middle school.

Because of the importance of determining the effectiveness of instructional programs, Houghton Mifflin Harcourt contracted with the Educational Research Institute of America (ERIA) to study the effectiveness of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012. Houghton Mifflin Harcourt sought to determine the instructional effectiveness of the program in teaching science concepts and skills to students at the middle school level. This report presents the findings from an effectiveness study of a grade 7 unit from the program.

Research Questions

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

Is Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 effective in improving students' knowledge and skills in science?

Is Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 effective in improving the science knowledge and skills of those students who score at different levels on the pretest?

Design of the Study

The study of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 was conducted at grades 2, 4, and 7. This report includes findings only at the middle school level—grade 7. A companion report describes the results for grade 2 and 4.

For this study, a single unit from the national edition of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 was used. The teacher participating in the study used the *SCIENCEFUSION* materials as her primary program for science instruction with five different science classes over a period of approximately two weeks. This teacher had not used the program prior to her involvement in the study.

The units selected for tryout included a focus on developing students' knowledge and skills around a particular scientific concept. The grade 7 unit focused on the interactions of various organisms in different community types. The community types included predatory/prey communities, symbiotic communities, and competitive communities.

One teacher with five different classes in New York State was included in the study. Upon completion of the study, the teacher filled out a questionnaire that asked about her use of the program during the study, in order to determine the fidelity with which she used the program materials.

The teacher administered the pretest during the first week of January 2010 and administered the posttest in the last week of January 2010. All tests and questionnaires were returned to ERIA by the first week of February 2010.

Instructional Approach under Study

Following is a description of the program provided by the publisher:

Houghton Mifflin Harcourt's *SCIENCEFUSION* ©2012 includes print, digital, and hands-on science project materials and activities for students in grades K through 8. The hands-on inquiry activities include both inquiry flip charts and virtual labs. The program is designed to meet the core standards in science.

The students' edition is a consumable work text. The work text engages students in writing on almost every page. The students' edition is designed to develop students' reading and writing skills.

The program includes science projects designed to be used by groups of students or in science centers. Easy, average, and challenging activities for each project are also included.

Digital lessons provide interactive activities, simulations, and videos. The digital lessons can be used with individual students for use in a computer lab or library setting. As well, the digital lessons can be projected on a digital whiteboard.

Assessments include lesson quizzes, benchmark tests and unit performance assessments. The teacher manual is supported with additional ideas for teaching through an online resource, www.thinkcentral.com.

Description of the Research Sample

There were 5 tryout classes in the study. The classes were all from one teacher in New York State. The teacher who taught the 5 classes had more 10 years experience teaching middle school science and had been teaching various science classes for more than 15 years. She felt very comfortable teaching science and was honored for her excellent teaching of science. During the approximately two weeks of use of the program, the teacher reported using the materials for 30 to 40 minutes each day.

The data provided in Table 1 provide a demographic summary of the school. The table does not provide specific data for the classes included. It does, however, provide a general description of the school and, thereby, an estimate of the make-up of the five classes that comprised the sample.

The table below shows that the school enrollment was 650 students. None of the students were enrolled in free/reduced lunch programs and the minority enrollment in the school was 5%.

Table 1
Demographic Characteristics of Grade 2 Schools Included in the Study

<i>Location</i>	<i>Grades</i>	<i>Students Enrolled</i>	<i>% Students Free/Reduced Lunch Programs</i>	<i>% Minority</i>	<i>% Special Education Needs</i>
Urban Fringe Large City	K to8	894	0%	5%	0%

Description of the Assessments

The outcome measures used for the study were developed by researchers at ERIA. Test items on the pretest were scrambled for the posttest. The test was developed to match the instruction in and the learning outcomes of the unit being taught.

The test included 32 four-option multiple choice test items assessing students' knowledge and understanding of the interactions of various organisms in community life. Items included an understanding of the steps in collecting and analyzing data.

Table 4 provides the test statistics for the posttest. The reliability of the posttest shows that the test was reliable for making instructional decisions regarding student growth.

Table 2
Posttest Reliability Statistics

	<i>Experimental</i>
Number of Test Items	32
Maximum Score	30
Minimum Score	19
Average Score	27.2
Percent Correct	85%
Reliability*	.82

*Kuder-Richardson 20

Data Analyses

The results for the students were analyzed using two different procedures:

1. A Paired Comparison *t*-test of the group's pretest and posttest scores sought to determine if students demonstrated significant growth from pretest to posttest.
2. Paired Comparison *t*-tests were also used to determine subgroup growth based on pretest performance was also conducted.

Treatment Group Comparison

Researchers at ERIA conducted a Paired Comparison *t*-test determine if the differences in the pretest and posttest raw scores and percent correct scores were significantly different. A total of 90 students were included in this analysis. The total test included 32 items, each worth one point. The .05 level of significance was used as the level at which differences would be considered statistically significant.

In addition to the *t*-tests analyses were computed for each of the 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:

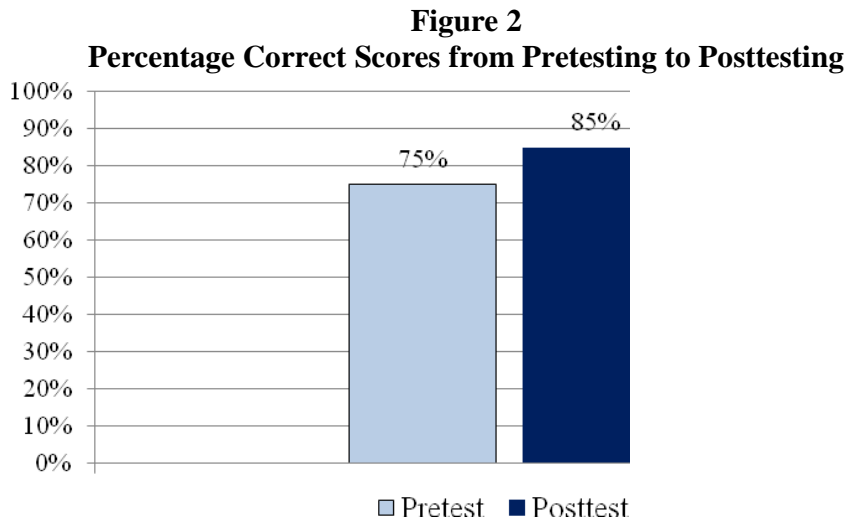
- .2 = small effect
- .5 = medium effect
- .8 = large effect

Table 3 presents the results of this Paired Comparison *t*-test. The average percent correct score increased from 75% on the pretest to 85% on the posttest. The difference was statistically significant at the $\leq .0001$ level, indicating that such a change would have occurred by chance less than once out of 10,000 repetitions. The effect size was large.

Table 3
Paired Comparison *t*-test Results
for Pretest/Posttest Comparison of the
Total Test Mean Percent Correct Scores a

<i>Results</i>	<i>Group</i>	<i>Number Students</i>	<i>Mean% Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
Pretest	Total	90	75.0%	9.2%	12.907	$\leq .0001$	1.26
Posttest	Total	90	85.0%	6.2%			

Figure 1 shows the pretest-to-posttest changes from pretesting to posttesting. There was a 10 percentage point increase taking the average posttest mean percent score to 85%.



Subgroups by Pretest Performance—Pretest/Posttest Analyses

To determine the gains by students scoring at different levels on the pretest, the total group of experimental students was ranked from lowest to highest based on pretest scores. These 90 students were then divided into two equal groups of 45 students. The scores of the low pretest group ranged from 46% to 78% correct and the high group scores ranged from 78% to 91% correct.

Table 4 presents the results of the Paired Comparison *t*-test performed for each of the subgroups based on pretest performance. The average percent correct score increased to a greater extent for the low pretest group than for the high pretest group. Almost certainly, high pretest scores limited the high pretest group students' opportunity for growth from pretest to posttest.

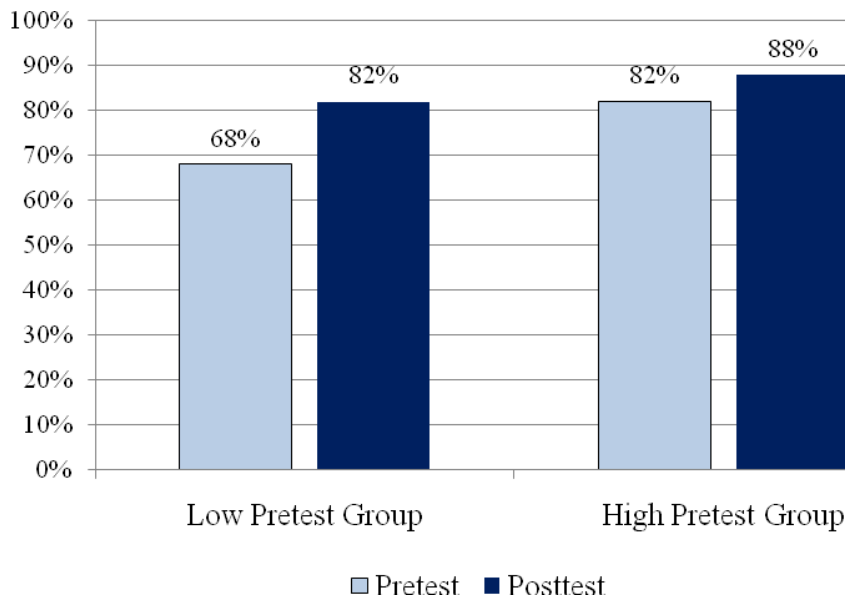
The increases were statistically significant at the $<.0001$ level for both groups. This level of significance indicates that such a change would have occurred by chance less than once out of 10,000 repetitions. This level of significance indicates that such a change would have occurred by chance less than once out of 1,000 repetitions. The effect sizes for the both groups were large.

Table 4
Paired Comparison *t*-test Results
for Pretest/Posttest Comparison of the Total Test Mean Percent Correct Scores
for Subgroups Based on Pretest Performance

<i>Results</i>	<i>Number Students</i>	<i>Mean % Score</i>	<i>SD</i>	<i>t-test</i>	<i>Significance</i>	<i>Effect Size</i>
<i>Low Pretest Group</i>						
Pretest	45	68.0%	7.7%	12.481	<.0001	1.90
Posttest	45	81.9%	6.5%			
<i>High Pretest Group</i>						
Pretest	45	82.0%	3.4%	8.619	<.0001	1.59
Posttest	45	88.2%	4.1%			

Figure 2 shows the pretest-to-posttest changes among these subgroups based on pretest scores. The low pretest group increased 14 percentage points from pretesting to posttesting, and the high pretest group increased 6 percentage points from pretesting to posttesting.

Figure 2
Percentage of Low and High Pretest Group Students
Percentage Correct Scores on the Pretest and Posttest



Conclusions

This study sought to determine the effect of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 program on students' knowledge and skills in science. For this study, a single unit from the national field test edition of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 was used with students in five different classes at grade 7. The same teacher, an experience and expert science teacher, taught all 5 classes.

The results indicate that for the 90 students in the five classes, statistically significant pretest to posttest gains were made for the total group of students using the program and the effect size was large. Furthermore, the effect sizes for the two subgroups based on pretest performance were also statistically significant and the effect sizes were large.

The conclusion, based on a reliable test designed to measure growth on science skills and knowledge related to a single unit of instruction, is that use of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 significantly increases students' knowledge and skills in science. The scores of students in this study who received instruction using a tryout unit of the program increased statistically significantly. When coupled with a separate study for grades 2 and 4 which showed similar significant results, the results provide evidence that the program is effective across the grades. These results for all three grades are particularly significant considering the very short duration of the study (two weeks of program use) and the fact that the teacher had never used the program before.