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'S *SCIENCE FUSION* ©

Report Number 397

Table of Contents

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

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

Introduction

Increased globalization threatens the economic prosperity and strategic leadership position that the United States has enjoyed since World War II; Americans now compete for jobs against increasingly higher-skilled but much lower-paid workers who can provide labor and service from a great distance, particularly from growing nations such as China and India—therefore the need for the strongest education in science and technology for the next generation is dire (Committee on Prospering in the Global Economy of the 21st Century, 2007). And yet the most recent results in elementary and secondary level science achievement on the National Assessment of Educational Progress (NAEP) indicate that only 34% of fourth-graders, 30% of eighth-graders, and 21% of twelfth-graders are considered proficient and that gaps in achievement between genders, ethnic groups, and family income levels, persist (National Center for Education Statistics, 2011). Worldwide, on the Program for International Student Assessment (PISA), an assessment that every three years measures reading literacy, mathematics literacy, and science literacy of 15 year-olds in dozens of countries around the globe, the United States scored lower than 18 other nations.

The outlook however is less grim when considered in light of this: unlike previous education reformers, such as those of the 195os and 1960s Sputnik-era, who were challenged to decide what to teach and how, we today understand more about how people learn, and how science instruction in particular can be improved for all learners (Duschl, Shouse, and Schweingruber, 2008).

We know, for example, that inquiry-based instruction in science is key. The National Science Education Standards (NRC 1996), the National Research Council (NRC 1996, 2005, 2007), and the National Science Foundation (NSF 2000) all concur that science educators must support students' natural, interactive inquiries. Indeed, the National Research Council, in a 2007 publication entitled, *Taking Science to School: Learning and Teaching Science in Grades K-8*, claimed students who are proficient in science are those who are able to: know, use, and interpret scientific explanations of the natural world; generate and evaluate scientific evidence and explanations; understand the nature and development of scientific knowledge; and participate productively in scientific practices and discourses. This same document further calls for a scaffolded approach to science that includes an effective metacognitive component as these instructional techniques yield increased conceptual understandings.

Houghton Mifflin Harcourt's *SCIENCE FUSION* ©2012 was developed on these and other research-based pedagogical principles for science teaching and learning—and with the increasingly globalized future in mind.

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 *SCIENCE FUSION* © 2012. Houghton Mifflin

Harcourt sought to determine the instructional effectiveness of the program in teaching science concepts and skills to students in elementary and middle school. This report presents the findings from a tryout of several chapters of the program with students at grade 2, grade 4, and grade 7.

Research Questions

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

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

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

Design of the Study

The study of Houghton Mifflin Harcourt's *SCIENCE FUSION* © 2012 was conducted at grades 2, 4, and 7.

For this study, several units from the national version of Houghton Mifflin Harcourt's *SCIENCE FUSION* © 2012 was used at each grade level for instruction. The teachers participating in the study used the *SCIENCE FUSION* materials as their primary program for science instruction over a period of approximately 18 weeks. None of the participating teachers had used the program prior to their involvement in the study.

Each of the units selected for tryout included a focus on developing students' knowledge and skills around a particular scientific concept. In addition, each unit provided focused instruction on the vocabulary related to the topic. Each unit concluded with a hands-on data collection and analysis activity.

Three different schools in three states were included in the study. One of the schools included both grade 2 and grade 4 classes, a second school included only grade 2 classes, and a third school included only grade 7 classes.

Upon completion of their participation in the study, teachers were asked to complete 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. In addition, the survey asked the teachers to evaluate the overall program as well as to evaluate specific program components.

All teachers administered the pretest during the first week of September 2010 and administered the posttest in the third week of January 2011. All tests and questionnaires were returned to ERIA by the first week of February 2011.

Instructional Approach under Study

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

Houghton Mifflin Harcourt's SCIENCE FUSION ©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

The study included 5 grade 2 classes and teachers, 4 grade 4 classes and teachers, 4 grade 7 classes taught by the same teacher. Table 1 provides a demographic summary of the schools. The tables do not provide specific data for the classes included. They do, however, provide a general description of each of the schools and, thereby, an estimate of the make-up of the classes that comprised the sample.

The table below shows that the average school enrollment was 675 students. An average of 31% percent of the students was enrolled in free/reduced lunch programs and the minority enrollment average in the schools was 28%.

Table 1
Demographic Characteristics of the Schools Included in the Study

Location	*Grades*	*Students Enrolled*	*% Students Free/Reduced Lunch Programs*	*% Minority*	*% Special Education Needs*
Large Central City*	PK to 5	605	69%	64%	12%
Urban Fringe Mid-Size City**	PK to 5	262	6%	9%	22%
Urban Fringe Large City***	5 to 8	1157	17%	10%	6%
Average		**675**	**31%**	**28%**	**13%**

**Included both grade 2 and grade 4 classes*
***Included only grade 2 classes*
****Included only grade 7 classes*

Description of the Assessments

The outcome measures used for the study were developed by researchers at ERIA. A different assessment was developed for each grade level. Each test was developed to match the instruction in, and the learning outcomes of, the units being taught.

The grade 2 test included 30 three-option multiple choice test items assessing students' knowledge and understanding of scientific measurement, scientific method, rocks and soil, meteorology and weather, and properties of matter.

Table 2 provides the test statistics for the grade 2 pretest and posttest. The reliability of the posttest shows that the test was reliable for making instructional decisions regarding student growth. The low reliability of the pretest indicates that students were making many guesses on the pretests. The sharp increase in the reliability of the posttest when compared with the pretest shows the effect of instruction and thus the decrease in guessing answers.

Table 2
Grade 2 Pretest/Posttest Reliability Statistics

	Pretest	***Posttest***
Number of Test Items	30	30
Maximum Score	22	29
Minimum Score	7	6
Average Score	14.3	18.2
Percent Correct	47.6	60.6
Reliability*	.44	.82

**Kuder-Richardson 20*

The grade 4 test included 30 four-option multiple choice test items assessing students' knowledge of investigations and the scientific method, astronomy and space exploration, geology, magnetic principles, and recognizing physical and chemical changes and identifying how such changes are important to humans.

Table 3 provides the test statistics for the grade 4 pretest and posttest. The reliability of the posttest, while at the lower end of acceptability, shows that the test was reliable for making instructional decisions regarding student growth. The relatively lower reliability of the pretest indicates that students were making many guesses on the pretests. The increase in the reliability of the posttest when compared with the pretest shows the effect of instruction and thus the decrease in guessing answers.

Table 3
Grade 4 Pretest/Posttest Reliability Statistics

	Pretest	***Posttest***
Number of Test Items	30	30
Maximum Score	18	19
Minimum Score	0	4
Average Score	9.5	11.1
Percent Correct	31.6	36.9
Reliability*	.65	.54

**Kuder-Richardson 20*

The grade 7 test included 35 four-option multiple choice test items assessing students' knowledge of the scientific method and investigation, scientific theories, conducting

experiments, earth science, environmental science, climate change, pollution, and principles of sound and sound waves.

Table 4 provides the test statistics for the grade 7 pretest and posttest. The reliability of the posttest shows that the test was reliable for making instructional decisions regarding student growth. The relatively lower reliability of the pretest indicates that students were making guesses on the pretests. The increase in the reliability of the posttest when compared with the pretest shows the effect of instruction and thus the decrease in guessing answers.

Table 4
Grade 7 Pretest/Posttest Reliability Statistics

	Pretest	***Posttest***
Number of Test Items	35	35
Maximum Score	24	26
Minimum Score	3	6
Average Score	12.1	14.6
Percent Correct	34.6	41.8
Reliability*	.61	.71

**Kuder-Richardson 20*

Data Analyses

The results for each of the three grades were analyzed independently. All raw scores were first converted to standard scores using a mean of 300 and a standard deviation of 50. This conversion provided a more normal distribution of test scores.

Two primary analyses were conducted for each grade:

1. A comparison of the pretest and posttest scores sought to determine if they differed significantly.

A Paired Comparison *t*-test was performed in order to determine if the students pretest standard scores increased statistically significantly from pretest to posttest.

2. A comparison of the pretest lower scoring and higher scoring students sought to determine if each of these two groups demonstrated significant growth from pretest to posttest.

A Paired Comparison *t*-test was used to compare the pretest and posttest standard scores of both the higher and lower scoring subgroups.

In addition to the t-test, effect-size analyses were computed using Cohen's *d* statistic. 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

In order to provide a more visual interpretation of test score changes from pretesting to posttesting, figures were produced to compare changes for the total group at each grade as well as changes for the lower and higher scoring pretest groups.

A final section of the data analysis provides a summary of the teacher fidelity and program evaluation questionnaires.

Grade 2 Results

Total Group Comparison

Researchers at ERIA conducted a Paired Comparison *t*-Test to determine if the differences in the pretest standard scores were significantly different from the posttest standard scores for 59 grade 2 students. The .05 level of significance was used as the level at which differences would be considered statistically significant.

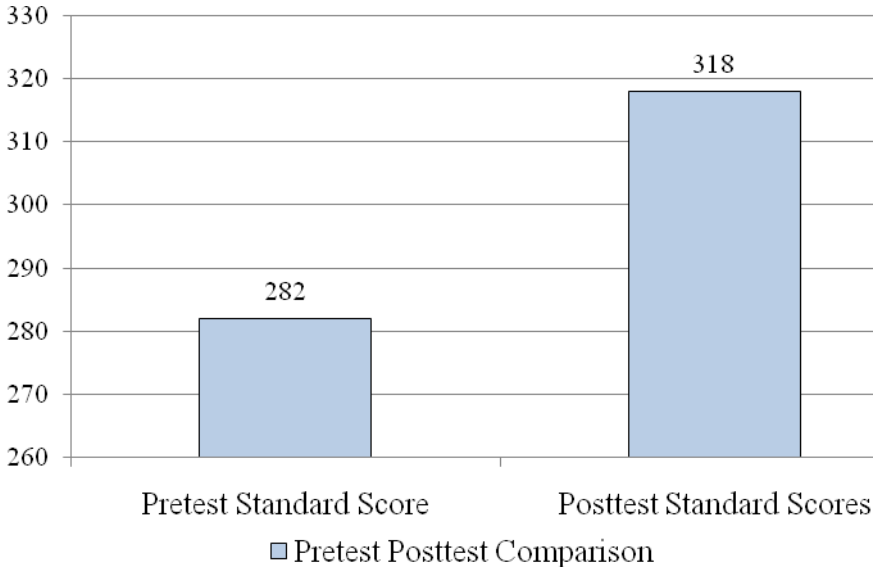
Table 5 indicates that the comparison of standard scores for the pretest and posttest comparison were statistically significant (<.0001) indicating a difference that would have occurred by chance less than once out of 10,000 repetitions. The effect size was large.

Table 5
Paired Comparison t-Test Results Comparing Standard Score Pretest and Posttest Results for Grade 2 Students

Students	*Mean Standard Score*	*SD*	*t*-Test	*Significance*	*Effect Size*
59	282	36	5.244	<.0001	.80
59	318	55			

Figures 1 provides a graphic display of the standard score comparisons for the grade 2 students from pretest standard scores to posttest standard scores. The grade 2 students increased their standard scores by 36 standard score points which is just over 70% of a standard deviation.

Figure 1
Average Standard Scores from Pretesting to Posttesting
Grade 2 Students



Subgroups by Pretest Performance—Pretest/Posttest Comparison

To determine the gains by students scoring at different levels on the pretest, the total group of grade 2 students was ranked from lowest to highest based on pretest raw scores. These 59 students were then divided into two groups of 29 students for the lower scoring group and 30 students for the higher scoring group. The standard scores of the lower scoring pretest group ranged from 199 to 282 and the high group scores ranged from 282 to 355.

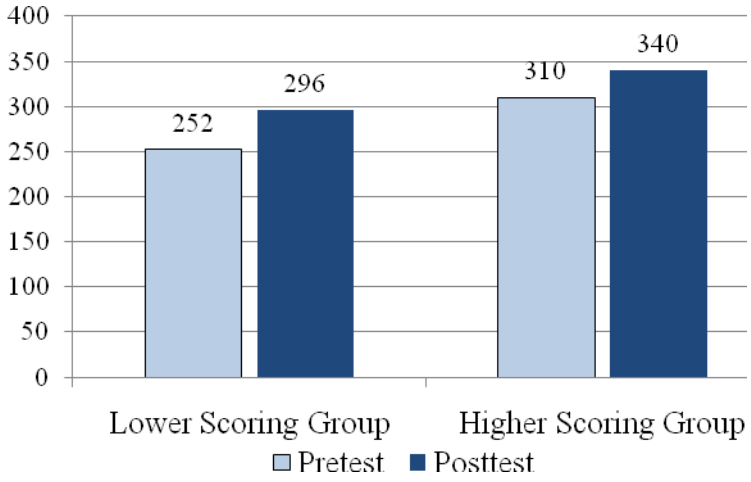
Table 6 presents the results of the Paired Comparison *t*-test performed for each of the subgroups which were based on pretest performance. The average standard score increased to a larger extent for the lower scoring students than for the higher scoring students. The increase was statistically significant at the <.0001 level for the low group. This level of significance indicates that such a change would have occurred by chance less than once out of 10,000 repetitions. For the high scoring group the increase was statistically significant at the <.004 level. This level of significance indicates that such a change would have occurred by chance less than four times out of 1,000 repetitions. The effect size for the lower scoring group was large and for the higher scoring group the effect size was medium.

Table 6
Paired Comparison t-Test Results Comparing Standard Score Pretest and Posttest Results for Grade 2 Students

Students	***Mean Standard Score***	***SD***	***t*-Test**	***Significance***	***Effect Size***
Lower Scoring Students					
29	252	22	4.296	<.0001	1.17
29	296	48			
Higher Scoring Students					
30	310	22	3.105	<.004	.74
30	340	54			

Figure 2 shows the pretest-to-posttest standard score for the two subgroups. The figures show that the lower scoring group increased their average standard score by 48 points and the higher scoring group increased their average standard score by 30 points.

Figure 2
Average Standard Scores from Pretesting to Posttesting
For Grade 2 Students Scoring Lower and Higher on the Pretests


400
350
300
250
200
150
100
50
0
252
296
310
340
Lower Scoring Group
Higher Scoring Group
Pretest
Posttest

Grade 4 Results

Total Group Comparison

Researchers at ERIA conducted a Paired Comparison *t*-Test to determine if the differences in the pretest standard scores were significantly different from the posttest standard scores for 54 grade 4 students. The .05 level of significance was used as the level at which differences would be considered statistically significant.

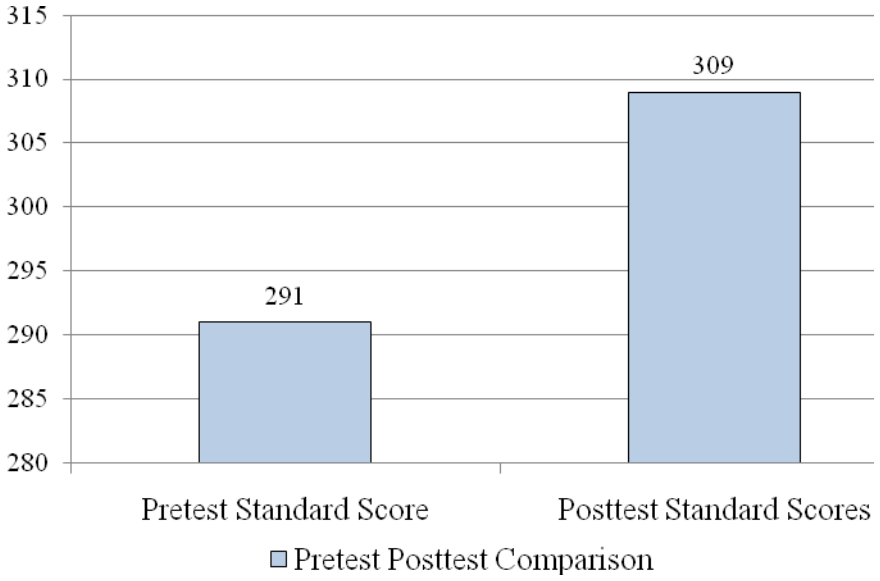
Table 7 indicates that the comparison of standard scores for the pretest and posttest comparison were statistically significant (<.03) indicating a difference that would have occurred by chance less than three times out of 100 repetitions. The effect size was small.

Table 7
Paired Comparison t-Test Results Comparing Standard Score Pretest and Posttest Results for Grade 4 Students

Students	*Mean Standard Score*	*SD*	*t*-Test	*Significance*	*Effect Size*
54	291	51	2.264	<.03	.34
54	309	48			

Figure 3 provides a graphic display of the standard score comparisons for the grade 4 students from pretest standard scores to posttest standard scores. The grade 4 students increased their standard scores by 18 standard score points which is just over a third of a standard deviation.

Figure 3
Average Standard Scores from Pretesting to Posttesting Grade 4 Students



Subgroups by Pretest Performance—Pretest/Posttest Comparison

To determine the gains by students scoring at different levels on the pretest, the total group of grade 4 students was ranked from lowest to highest based on pretest raw scores. These 54 students were then divided into two equal groups of 27 students for the lower and higher scoring

groups. The standard scores of the lower scoring pretest group ranged from 180 to 284 and the higher group scores ranged from 284 to 400.

Table 8 presents the results of the Paired Comparison *t*-test performed for each of the subgroups which were based on pretest performance. The average standard score increased for the lower scoring students and there was a small decrease for the higher scoring group. The increase for the lower scoring group was statistically significant at the <.0001 level for the low group. This level of significance indicates that such a change would have occurred by chance less than once out of 10,000 repetitions.

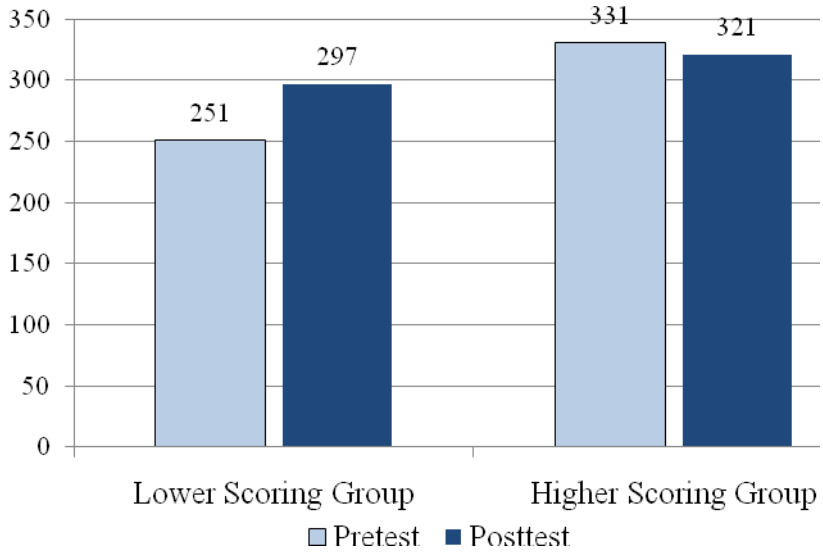
The effect size for the lower scoring group was large and for the higher scoring group there was a small negative effect size.

Table 8
Paired Comparison t-Test Results Comparing Standard Score Pretest and Posttest Results for Grade 4 Students

Students	***Mean Standard Score***	***SD***	***t*-Test**	***Significance***	***Effect Size***
Lower Scoring Students					
27	251	31	4.868	<.0001	1.23
27	297	41			
Higher Scoring Students					
27	331	31	-.991	Non-Significant	.23
27	321	52			

Figure 4 shows the pretest-to-posttest standard score for the two subgroups. The figures show that the lower scoring group increased their average standard score by 46 points and the higher scoring group decreased their average standard score by 10 points.

Figure 4
Average Standard Scores from Pretesting to Posttesting for Grade 4 Students Scoring Lower and Higher on the



Pretests

Grade 7 Results

Total Group Comparison

Researchers at ERIA conducted a Paired Comparison *t*-Test to determine if the differences in the pretest standard scores were significantly different from the posttest standard scores for 54 grade 7 students. The .05 level of significance was used as the level at which differences would be considered statistically significant.

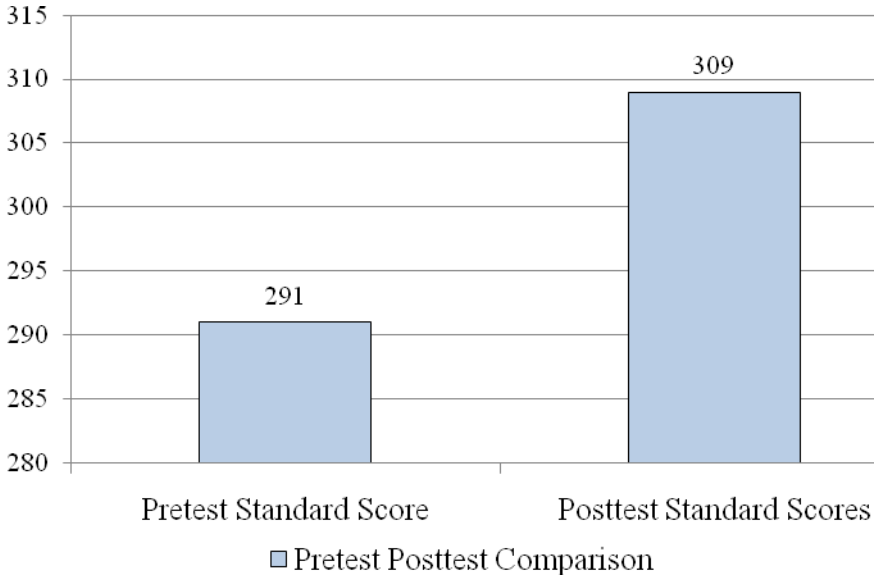
Table 9 indicates that the comparison of standard scores for the pretest and posttest comparison were statistically significant (<.002) indicating a difference that would have occurred by chance less than twice out of 1,000 repetitions. The effect size was small.

Table 9
Paired Comparison t-Test Results Comparing Standard Score Pretest and Posttest Results for Grade7 Students

Students	*Mean Standard Score*	*SD*	*t*-Test	*Significance*	*Effect Size*
54	291	46	3.198	<.002	.36
54	309	53			

Figure 5 provides a graphic display of the standard score comparisons for the grade 7 students from pretest standard scores to posttest standard scores. The grade 7 students increased their standard scores by 18 standard score points which is about one third of a standard deviation.

Figure 5
Average Standard Scores from Pretesting to Posttesting
Grade 7 Students



Subgroups by Pretest Performance—Pretest/Posttest Comparison

To determine the gains by students scoring at different levels on the pretest, the total group of grade 7 students was ranked from lowest to highest based on pretest raw scores. These 54 students were then divided into two equal groups of 27 students for the lower and higher scoring

groups. The standard scores of the lower scoring pretest group ranged from 182 to 291 and the higher group scores ranged from 291 to 411.

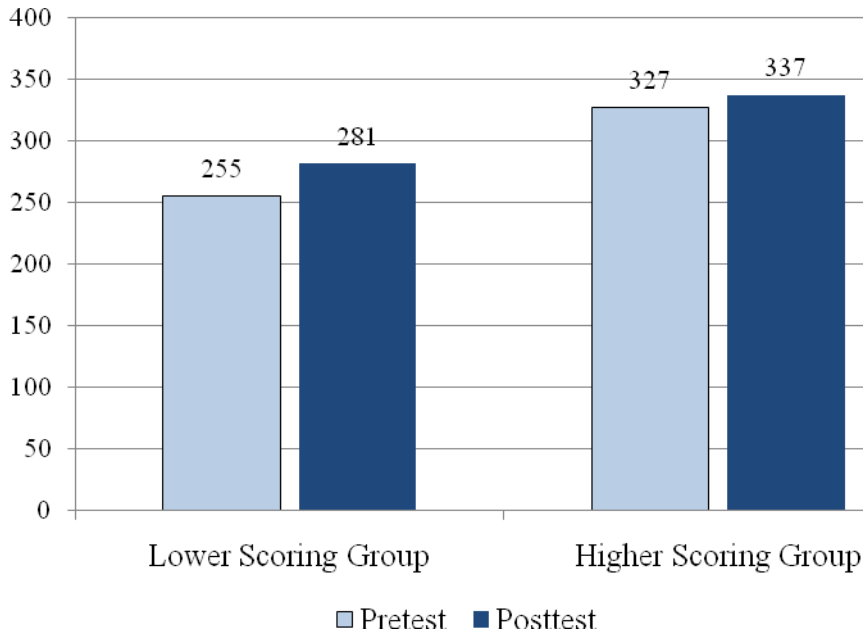
Table 10 presents the results of the Paired Comparison *t*-test performed for each of the subgroups which were based on pretest performance. The average standard score increased to a larger extent for the lower scoring students than for the higher scoring students. For the lower scoring group the increase was statistically significant at the <.005 level. This level of significance indicates that such a change would have occurred by chance less than five times out of 1,000 repetitions. For the higher scoring group the increase was not statistically significant. The effect size for the lower scoring group was medium and for the higher scoring group the effect size was small.

Table 10
Paired Comparison t-Test Results Comparing Standard Score Pretest and Posttest Results for Grade 7 Students

Students	***Mean Standard Score***	***SD***	***t*-Test**	***Significance***	***Effect Size***
Lower Scoring Students					
27	255	26	3.071	<.005	.77
27	281	40			
Higher Scoring Students					
27	327	30	1.385	Non-Significant	.24
27	337	50			

Figure 6 shows the pretest-to-posttest standard score for the two subgroups. The figures show that the lower scoring group increased their average standard score by 26 points and the higher scoring group increased their average standard score by 10 points..

Figure 6
Average Standard Scores from Pretesting to Posttesting Grade 7 Students



Teachers' Fidelity of Use and Program Evaluations

Table 11 summarizes the results of the survey of the teachers' fidelity of use of the program. Four grade 2 teachers and 1 grade 7 teacher completed and returned the teacher surveys. The numbers of responses differ across the various questions because each teacher did not respond to each of the questions.

Table 11
Teacher Fidelity of Use of Program Materials

Teacher Experience					
How long have you been in your current position?					
	2 years or less	*3 to 5 years*	*6 to 10 years*	*10 years or more*	
Grade 2	1	1	2		
Grade 7			1		
How long have you been employed as an educator?					
	2 years or less	*3 to 5 years*	*6 to 10 years*	*10 years or more*	
Grade 2		2	1	1	
Grade 7			1		
Use of the Houghton Mifflin Harcourt *SCIENCE FUSION* Tryout Materials					
How many students participated in the tryout?					
	Fewer than 10	*11 to 15*	*16 to 20*	*21 to 25*	*25 or more*
Grade 2		1	2	1	
Grade 7					1
How many days per week did you use the materials?					
	1	*2*	*3*	*4*	*5*
Grade 2			4		
Grade 7				1	
How many minutes per day did you/your students use the program for science instruction?					
	Fewer than 20	*20 to 30*		*30 to 60*	
Grade 2		4			
Grade 7		1			
How many times during the week did you/your students use the digital path the digital content for science instruction?					
	1-2	*3-4*		*5*	
Grade 2	4				
Grade 7	1				
How many times during the week did you/your students use the leveled readers?					
	1-2	*3-4*		*5*	
Grade 2	3				
Grade 7		1			

Teacher Comments

Which feature or aspect of the program did you feel was the most effective or most valuable? Please explain.

The student editions I found the most successful. The students really enjoyed having their own place to write down their thoughts, ideas, and answers.

I love the inquiry flip chart and the online resources.

Being able to write in the books and tear out the pages seemed to keep the students more engaged and interactive so I feel that the information stuck better.

I think the student edition is valuable because it is interactive, incorporates reading strategies, and it is engaging for students due to the easy to relate to, colored pages.

Which feature or aspect of the program did your students seem to like the best? Please explain.

I think the student editions were also their favorite.

The students love their science journals and participating in the hands on activities.

They loved writing in the books and the hands on activities and the experiments.

The students love the inquiry flip chart and found them easy to follow.

Which feature or aspect of the program did you feel was the most difficult for your students? Please explain.

My classroom did not receive the leveled readers and we don't have enough computers or smartboard to use the digital resources, so I feel it was difficult to give my students additional paths to learning the materials.

Doing the inquiry activities without assistance or being guided through the activity.

The students have a difficult time completing some of the inquiries independently. Also, some of the higher leveled readers were too difficult for my students to read.

Which feature or aspect of the program do you feel was the least effective or least valuable for your students? Please explain.

Some of the units do not align with our standard course of study and some of the things on our standard course of study were not included in this program. Other than that I really enjoyed it.

I felt that all the components were valuable.

Which feature or aspect of the digital path did you feel was the most effective or most valuable for your students? Please explain.

All of it! I also like the idea of being able to get updates without having to order new textbooks.

I enjoyed being able to pull the text book up on the smartboard so I could point to things and have all of the students see it.

The online resources, outside of the online textbook, were the most valuable.

Which feature or aspect of the digital path did you feel was the least effective or least valuable for your students? Please explain.

I'm not sure.

I didn't use the online textbook aspect of the digital path.

Are there any components or learning materials that you feel should be added to the program? Please identify and, if possible, include a brief description.

The only constructive comment I can say is that I wish the flipcharts had the same material on both sides. This would make it easier when you have them working in groups.

The North Carolina standard that was not covered is sound. For schools in NC to use this program it would be helpful if sound was included. I understand that this program was made to align with Florida's standard course of study.

I cannot think of any at this time.

Teacher Evaluations of Program Materials

Table 12 provides the teacher evaluations of the program tryout materials overall as well as of specific features of the program. The numbers of responses differ across the various questions because each teacher did not respond to each of the questions.

Table 12
Teachers' Evaluations of Program Materials

Program Evaluation					
In general, how successful was the information presented in supporting your students in learning the targeted knowledge and skills?					
	Very Successful	*Somewhat Successful*	*Somewhat Unsuccessful*	*Very Unsuccessful*	*N/A*
Grade 2	3	1			
Grade 7		1			
Program Components Evaluation					
Please rate the program components in how well they support learning and instruction.					
	Excellent	*Good*	*Fair*	*Poor*	*N/A*
Student Edition					
Grade 2	3	1			
Grade 7		1			
Inquiry Flipchart					
Grade 2	2	2			
Grade 7		1			
Virtual Labs					
Grade 2	2	1			1
Grade 7		1			
Digital Path					
Grade 2	3				1
Grade 4		1			
Leveled Readers					
Grade 2	3				1
Grade 4		1			

Conclusions

This study sought to determine the effect of Houghton Mifflin Harcourt's *SCIENCE FUSION* © 2012 program on students' knowledge and skills in science. For this study, several units from the national field test edition of Houghton Mifflin Harcourt's *SCIENCE FUSION* © 2012 were used with students at grades 2, 4, and 7.

As can be seen in Table 13, for students at all three grades, significant pretest to posttest gains were made for the total group of students using the program. Also, at all 3 grades the increase from pretest to posttests was statistically significant for each of the groups scoring lower on the pretests. However, for the higher pretest scoring groups, increases were statistically significant only for the grade 2 higher scoring students. For the higher pretest scoring groups at grades 4 and 7, increases were not statistically significant. Effect sizes for 9 comparisons were large for 3 of the comparisons, medium for 2 of the comparisons, and small for 4 of the comparisons.

The teachers' reports regarding fidelity of use and their evaluations of the overall program materials and specific components of the program were all very positive.

Table 13
Summary of Significance of Paired Comparison *t*-tests and Effect Sizes for Pretest/Posttest Gains on the Total Test Grade 2, Grade 4, and Grade 7

	Grade 2		***Grade 4***		***Grade 7***	
	Significance	**Effect Size**	**Significance**	**Effect Size**	**Significance**	**Effect Size**
Total	<.0001	Large	<.03	Small	<.002	Small
Subgroups Based on Pretest Performance						
Higher Pretest	<.004	Medium	Non-Sig.	Small	Non-Sig.	Small
Lower Pretest	<.0001	Large	<.0001	Large	<.005	Medium

The conclusion, based on 3 separate assessments designed to measure growth on science skills and knowledge related to a single unit of instruction at 3 grade levels, is that use of Houghton Mifflin Harcourt's *SCIENCE FUSION* © 2012 significantly increases students' knowledge and skills in science for the total group of students at each grade. The scores of students in the study who received instruction using a tryout unit of the program increased statistically significantly. These results are particularly significant considering the fact that teachers had never used the program prior to the tryout and there was relatively little time for teachers to prepare for teaching this new program.

Bibliography

Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology, National Academy of Sciences, National Academy of Engineering, Institute of Medicine. (2007). Rising above the gathering storm: Energizing and employing American for a brighter economic future. Washington, D.C.: National Academies Press.

Duschl, R. A., Shouse, A. W., and Schweingruber, H.A. (2008). What research says about K-8 science learning and teaching. *Education Digest*, 73(8), 46-50.

Fleischman, H.L., Hopstock, P.J., Pelczar, M.P., and Shelley, B.E. (2010). *Highlights From PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context* (NCES 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics (2011). The Nation's Report Card: Science 2009 (NCES 2011–451). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Research Council. (1996). *National science education standards: Observe, interact, change, learn.* Washington, DC: National Academies Press.

National Research Council. (2005a). *How students learn: Science in the classroom.* Committee on *How People Learn*, A Targeted Report for Teachers, M.S. Donovan & J. D. Bransford, Eds. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2005b). *Mathematical and scientific development in early childhood.* Washington, DC: The National Academies Press.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Committee on Science Learning, Kindergarten through Eighth Grade, R. A. Duschl, H. A. Schweingruber, & A. W. Shouse, Eds. Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Science Foundation. (2000). Foundations: A monograph for professionals in science, mathematics and technology education. Vol. 2. Inquiry: Thoughts, views, and strategies for the K-5 classroom. Arlington, VA: Author.