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 *Florida Science Fusion*©2012

Report Number 422

Table of Contents

Abstract

As a means of improving science education to make students in the U.S. more competitive with their international peers, Houghton Mifflin Harcourt has published **Houghton Mifflin Harcourt's *Florida Science Fusion*© 2012**.

In order to evaluate the program's effectiveness, *Houghton Mifflin Harcourt* contracted with the *Educational Research Institute of America* (ERIA) to conduct a full academic year study; this study was conducted during the 2011/2012 academic year and examined student performance at grades 3, 4, and 5. This study consisted of two groups, a treatment group that used ***Science Fusion*** and a Control group, that used a different science program. In total, 31 ***Science Fusion*** and 13 Control classrooms participated in the study.

At each grade level, tests were designed to assess students' understanding of science knowledge and analysis skills. Students were administrated two tests, a pretest in the Fall of 2011 and a posttest in the Spring of 2012. Test analysis indicated that the assessments had sound psychometric properties for making claims regarding program effectiveness.

Statistical analysis indicated that students in both the ***Science Fusion*** group and the Control group made statistically significant gains from pretest to posttests at all grade levels. Students using ***Science Fusion*** had higher scores on both pretests and posttests and this was partially due to the composition of the student body, as there were more minority and disadvantaged students in the control group than the ***Science Fusion*** group. Despite this discrepancy, grade 4 and grade 5 students using ***Science Fusion*** made significantly greater gains from pretest to posttest than Control group students; the rate of improvement at grade 3 was similar for the two groups.

The results of this study indicate that **Houghton Mifflin Harcourt's *Florida Science Fusion*© 2012** is an effective science program leading to student improvement in science skills and strategies.

A STUDY OF THE INSTRUCTIONAL EFFECTIVENESS OF HOUGHTON MIFFLIN HARCOURT'S *Florida 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).

The outlook however is less grim when considered in light of this: unlike previous education reformers, such as those of the 1950s 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 Florida *Science Fusion* © 2012. Houghton Mifflin Harcourt sought to determine the instructional effectiveness of the program in teaching science concepts and skills to students in the elementary school. This report presents the findings from a full-year examination of the program with students at grade 3, grade 4, and grade 5.

Research Questions

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

1. Is **Houghton Mifflin Harcourt's Florida *Science Fusion*©2012** effective in improving students' knowledge and skills in science?

2. Is **Houghton Mifflin Harcourt's Florida *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 3, 4, and 5.

For this study, **Houghton Mifflin Harcourt's Florida *Science Fusion*©2012** materials were used at each grade level as the primary science program. None of the participating teachers had used the program prior to their involvement in the study. Pretesting was conducted in early September 2011 and post-testing was conducted in mid-May 2012. Teachers administered the assessments with directions and answer documents sent to them. All tests were returned to ERIA for scoring and analyses.

The study included two schools using the **Science Fusion p**rogram and two schools which served as Control Groups. The classes in the Control schools were teaching science using a science program which had been in use in the schools for several years.

Upon completion of 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. The survey indicated that the grade 3 teachers used the program for approximately 4 days per week and from 30-35 minutes per class. Grade 4 teachers used the program for approximately 4 days per week and 25 to 30 minutes per class. Grade 5 teachers used the program for about 4 ½ days per week and for over 35 minutes per class period.

Instructional Approach under Study

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

Houghton Mifflin Harcourt's Florida 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 write-in textbook that 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 two ***Science Fusion*** schools and two Control schools at each grade level. The ***Science Fusion*** schools included 10 grade 2 classes, 9 grade 4 classes, and 12 grade 5 classes. The Control schools included 3 grade 2 classes, 5 grade 4 classes, and 5 grade 5 classes. There were a total of 31 Science Fusion classes and 13 Control classes.

Table 1 provides a demographic summary of the ***Science Fusion*** schools and Control 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.

Table 1 shows that the percentage of minority students as well as the percentage of students on free/reduced lunch programs in the Control schools was much larger than in the ***Science Fusion*** schools. Nineteen percent of the students were identified as minority students in the Science Fusion schools while 99% were identified as minority in the Control schools. In addition, 13 percent of the students in the Science Fusion schools were enrolled in free/reduced lunch programs while 88% were enrolled in those programs in the Control schools.

Table 1
Demographic Characteristics of the Schools Included in the Study

Location	*Grades*	*Students Enrolled*	% Minority	% Free/ Reduced Lunch	% Special Needs	% ELL
Science Fusion Schools						
Rural	PK to 5	752	26%	16%	20%	4%
Suburban	PK to 5	708	12%	9%	13%	1%
Averages		**730**	**19%**	**13%**	**17%**	**3%**
Control Schools						
Suburban	PK to 5	370	99%	89%	11%	22%
Suburban	PK to 5	361	98%	86%	24%	2%
Averages		**366**	**99%**	**88%**	**18%**	**12%**

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 assess the students' skills and knowledge in science.

Tables 2 and 3 provide the test statistics for the pretests and posttests. The reliability of the posttest shows that the test was reliable for making instructional decisions regarding student growth. The lower reliability of the pretests particularly at grades 4 and 5 in the Control Classes indicates that students were making some 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
Reliability Estimates, Mean Scores and Reliability
for the Science Fusion Assessments at Grades 3, 4, and 5
Control Classes

Test	Standard Deviation	Reliability*	SEM**
Grade 3 Pretest	5.70	.76	2.8
Grade 3 Post-Test	4.95	.75	2.5
Grade 4 Pretest	4.47	.58	2.9
Grade 4 Post-Test	5.47	.73	2.8
Grade 5 Pretest	4.68	.63	2.8
Grade 5 Post-Test	6.78	.82	2.9

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

Table 3
Reliability Estimates, Mean Scores and Reliability
for the Science Fusion Assessments at Grades 3, 4, and 5
Science Fusion Classes

Test	Standard Deviation	Reliability*	SEM**
Grade 3 Pretest	5.99	.80	2.7
Grade 3 Post-Test	4.95	.75	2.5
Grade 4 Pretest	5.33	.72	2.8
Grade 4 Post-Test	5.77	.77	2.8
Grade 5 Pretest	5.82	.77	2.8
Grade 5 Post-Test	6.85	.87	2.5

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

Data Analyses

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

Three primary analyses were conducted for each grade:

1. A Repeated Measures Analysis of Variance (ANOVA) was used to determine if the two groups increased their scores from pretest to post-test, whether there was a significant difference between each group at pretesting and post-testing, and whether either group showed a greater increase from pretesting to post-testing.
2. A Paired Comparison *t*-test was performed for each grade to determine the specific increase, if any, from pretesting to post-testing for the Science Fusion Group and the Control Group.
3. A Paired Comparison *t*-test of the pretest lower scoring and higher scoring students was used to determine if each of these two groups demonstrated significant growth from pretest to posttest.

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

Grade 3 Results

Total Group Comparison

Researchers at ERIA conducted a Repeated Measures Analysis of Variance to determine three different effects as follows:

- ***Time*** indicates if the two groups increased their scores from pretest to post-tests.
- ***Time x Group*** indicates whether one group showed a higher increase than the other group.
- ***Group*** indicates whether there is a significant difference between the two groups at pretesting and post-testing.

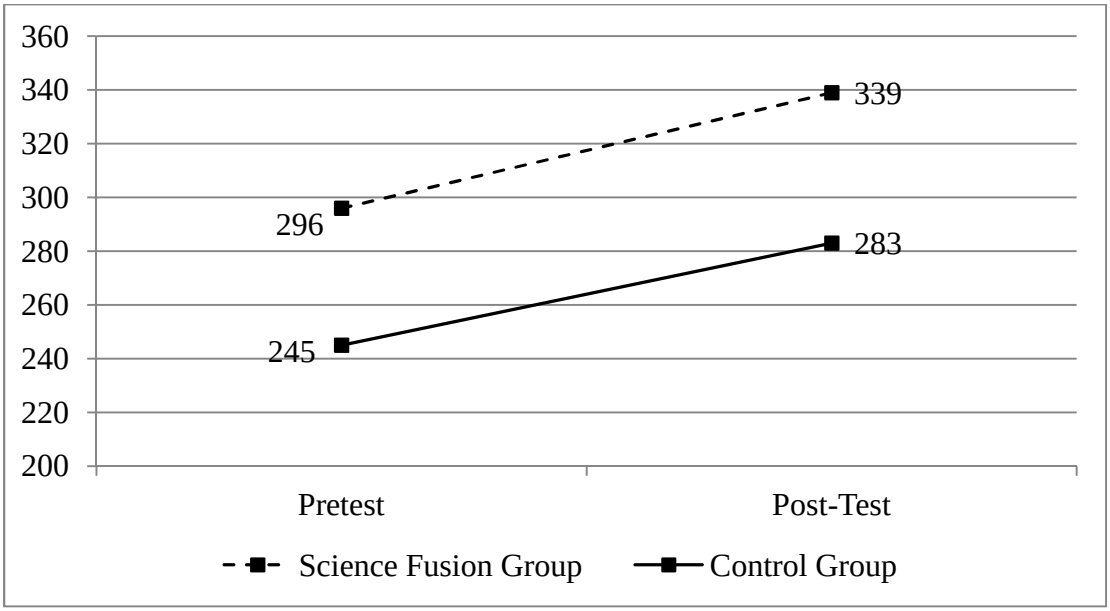
Table 4 shows that there was a significant effect for ***Time*** which indicates that both groups increased their scores from pretesting to post-testing. The ***Time x Group*** factor was non-significant indicating that neither the control or ***SCIENCE FUISON*** group made greater gains than the other. ***Group*** was significant and indicates that the groups were significantly different at both pretesting and post-testing.

Table 4
Repeated Measures Analysis of Variance Comparing Pretest and Post-Standard Scores for Control and Science Fusion Grade 3 Students

Source	df	Mean Square	F Test	Significance
Time	1	142982.726	251.710	≤.0001
Time x Group	1	591.650	1.042	≤.309
Group	1	250703.173	107.406	≤.0001

Figure 1 shows the results from Table 4 in graphic format. It is easy to see the ***time factor*** as both groups increased their scores significantly from pretesting to post-testing. The ***Group factor*** is also obvious in that there is a statistically significant difference between the two groups at pretesting and post-testing. However, Table 4 also shows that the **Time x Group** factor is non-significant and indicates no difference in the amount of increase from pretesting to post-testing. Both groups increased in a parallel fashion.

Figure 1
Grade 3 Standard Score Increases for the *Science Fusion* Group and the Control Group from Pretesting to Post-Testing



Despite the lack of significant difference in the growth of each group, a Paired Comparison *t*-test was computed to determine if there were any differences in gains made by each group from pretesting to post-testing. The figures in Table 5 support the findings from the Repeated Measures ANOVA in that both groups increased statistically significantly (≤.0001) and there was only a small difference between the two groups. The effect size for each group was large.

Table 5
Paired Comparison *t*-Tests Control and *Science Fusion*

Control Group (N=65)	***Mean***	***Standard Deviation***	***t Test***	***Significance***	***Effect Size***
Pretest	245.0	40.5	9.187	≤.0001	.94
Post-Test	282.9	39.8			
Science Fusion Group (N=134)					
Pretest	296.0	40.8	14.665	≤.0001	1.16
Post-Test	339.0	32.9			

High Pretest and Low Pretest Student Comparison

An analysis was next conducted to determine if both the ***Science Fusion*** Group and the Control Group increased the scores of the lower and higher pretest scoring groups. Table 6 shows that both the low pretest group and the high pretest group from the ***Science Fusion*** Group made statistically significant gains (≤.0001). The effect size was large for both groups.

Table 6
***Science Fusion* Group Paired Sample *t*-Tests**
for Low and High Pretest Scoring Groups

Low Pretest Group (N=67)	Mean	Standard Deviation	t Test	Significance	Effect Size
Pretest	264.1	32.5	13.297	≤.0001	1.74
Post-Test	322.8	34.4			
High Pretest (N=67)					
Pretest	327.8	15.9	9.782	≤.0001	1.49
Post-Test	355.2	21.5			

Table 7 shows that both the low pretest group and the high pretest group from the Science Fusion Group made statistically significant gains (≤.0001). The effect size was large for both groups.

Table 7
Control Group Paired Sample *t*-Tests
for Low and High Pretest Scoring Groups

Low Pretest Group (N=32)	Mean	Standard Deviation	t Test	Significance	Effect Size
Pretest	213.0	19.1	7.243	≤.0001	1.59
Post-Test	257.2	34.3			
High Pretest (N=33)					
Pretest	276.2	29.9	5.842	≤.0001	1.11
Post-Test	307.8	27.0			

Grade 4 Results

Total Group Comparison

Researchers at ERIA conducted a Repeated Measures Analysis of Variance to determine three different effects as follows:

- ***Time*** indicates if the two groups increased their scores from pretest to post-tests.
- ***Time x Group*** indicates whether one group showed a higher increase than the other group.
- ***Group*** indicates whether there is a significant difference between the two groups at pretesting and post-testing.

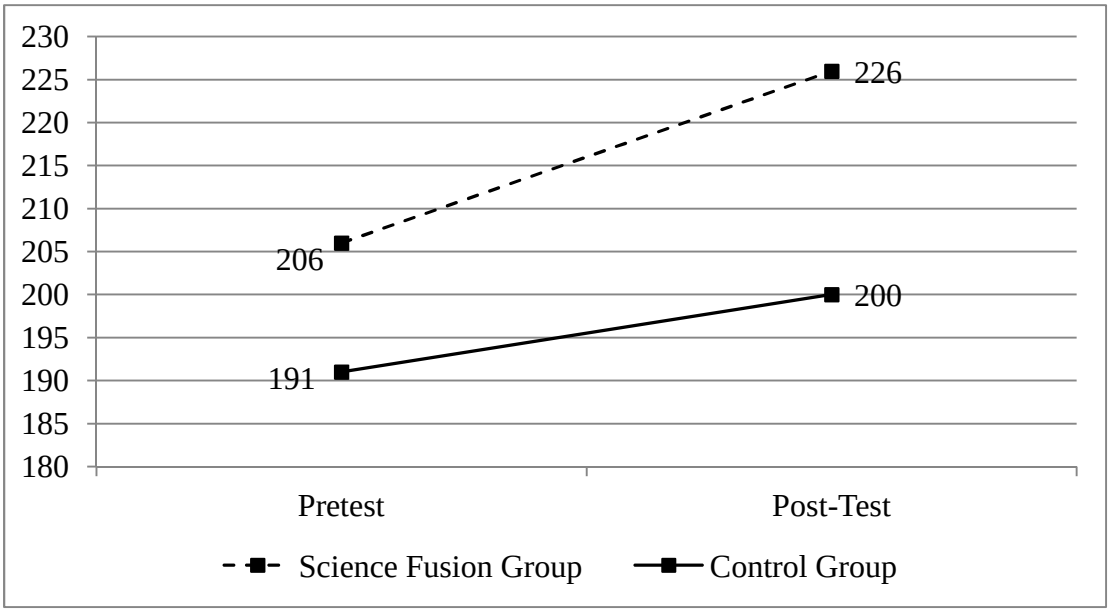
Table 8 shows that there was a significant effect for ***Time*** which indicates that both groups increased their scores from pretesting to post-testing. The ***Time x Group*** factor was significant indicating that either the control or ***Science Fusion*** group made greater gains than the other. ***Group*** was also significant and indicates that the groups were significantly different at both pretesting and post-testing.

Table 8
Repeated Measures Analysis of Variance Comparing Pretest and Post-Standard Scores for Control and *Science Fusion* Grade 4 Students

Source	df	Mean Square	F Test	Significance
Time	1	18535.244	97.066	≤.0001
Time x Group	1	2599.742	13.614	≤.0001
Group	1	36994.065	61.584	≤.0001

Figure 2 shows the results in Table 8 in graphic format. It is easy to see the ***time factor*** as both groups increased their scores significantly from pretesting to post-testing. The ***Group*** factor is also obvious in that there is a statistically significant difference between the two groups at pretesting and post-testing. Figure 4 also shows that the ***Time x Group*** factor indicates a strong difference in the amount of increase from pretesting to post-testing. The ***Science Fusion*** Group increased to a greater extent than did the Control Group.

Figure 2
Grade 4 Standard Score Increases for the *Science Fusion* Group and the Control Group from Pretesting to Post-Testing



A Paired Comparison *t*-test was computed to determine if the differences in gains made by each group from pretesting to post-testing. The figures in Table 9 support the findings from the Repeated Measures ANOVA in that both groups increased statistically significantly (≤.0001) but there is a difference between the two groups. The effect size for the Control Group was small while the effect size for the ***Science Fusion*** Group was large.

Table 9
Paired Comparison *t*-Tests Control and *Science Fusion*

Control Group (N=62)	***Mean***	***Standard Deviation***	***t Test***	***Significance***	***Effect Size***
Pretest	191.4	17.2	3.402	≤.001	.48
Post-Test	200.5	20.5			
Science Fusion Group (N=155)					
Pretest	206.4	19.5	13.057	≤.0001	.98
Post-Test	226.3	21.0			

High Pretest and Low Pretest Student Comparison

An analysis was next conducted to determine if both the ***Science Fusion*** Group and the Control Group increased the scores of the lower and higher pretest scoring groups. Table 10 shows that both the low pretest group and the high pretest group from the Science Fusion Group made statistically significant gains (≤.0001). The effect size was large for both groups.

Table 10
***Science Fusion* Group Paired Sample *t*-Tests**
for Low and High Pretest Scoring Groups

Low Pretest Group (N=78)	Mean	Standard Deviation	t Test	Significance	Effect Size
Pretest	190.8	11.4	12.343	≤.0001	1.69
Post-Test	217.9	19.6			
High Pretest (N=77)					
Pretest	222.3	11.5	7.111	≤.0001	.80
Post-Test	234.9	18.9			

Table 11 shows that both the low pretest group and the high pretest group from the Science Fusion Group made statistically significant gains (≤.0001). The effect size was large for low pretest group and small for the high pretest group.

Table 11
Control Group Paired Sample *t*-Tests
for Low and High Pretest Scoring Groups

Low Pretest Group (N=31)	Mean	Standard Deviation	t Test	Significance	Effect Size
Pretest	178.4	9.0	3.76	≤.001	.93
Post-Test	194.0	21.7			
High Pretest (N=31)					
Pretest	204.4	13.2	.844	Non-Significant	.16
Post-Test	206.9	17.3			

Grade 5 Results

Total Group Comparison

Researchers at ERIA conducted a Repeated Measures Analysis of Variance to determine three different effects as follows:

- ***Time*** indicates if the two groups increased their scores from pretest to post-tests.
- ***Time x Group*** indicates whether one group showed a higher increase than the other group.
- ***Group*** indicates whether there is a significant difference between the two groups at pretesting and post-testing.

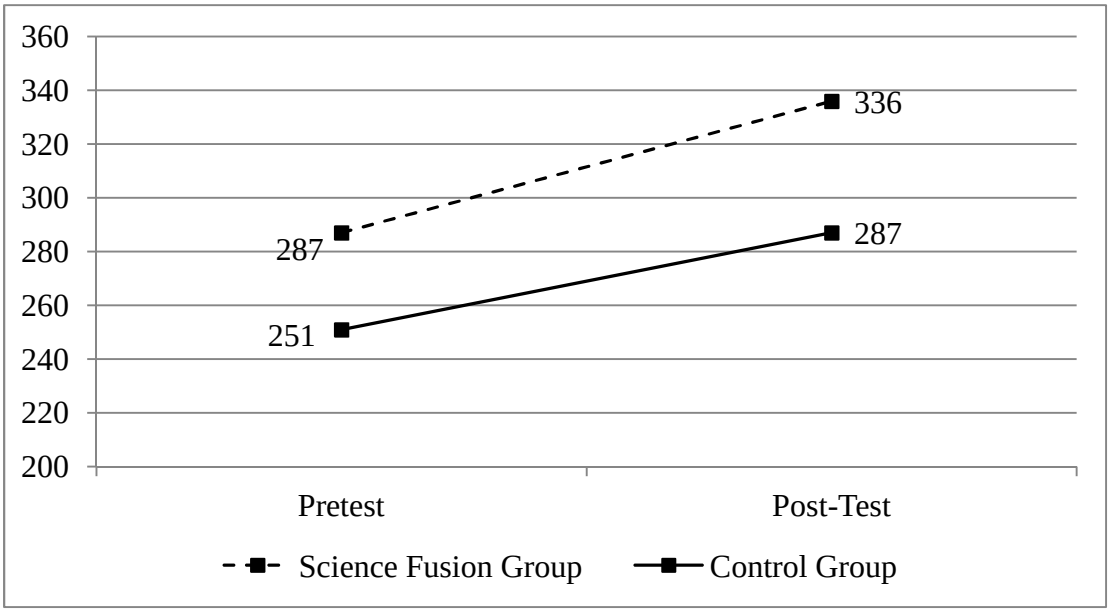
Table 12 shows that there was a significant effect for ***Time*** which indicates that both groups increased their scores from pretesting to post-testing. The ***Time x Group*** factor was significant indicating that either the control or ***Science Fusion*** group made greater gains than the other. ***Group*** was significant and indicates that the groups were significantly different at both pretesting and post-testing.

Table 12
Repeated Measures Analysis of Variance Comparing Pretest and Post-Standard Scores for Control and Science Fusion Grade 5 Students

Source	df	Mean Square	F Test	Significance
Time	1	6504587.385	10573.069	≤.0001
Time x Group	1	29521.847	47.987	≤.0001
Group	1	29468.956	45.746	≤.0001

Figure 3 shows the results in Table 12 in graphic format. It is easy to see the ***time factor*** as both groups increased their scores significantly from pretesting to post-testing. The ***Group*** factor is also obvious in that there is a statistically significant difference between the two groups at pretesting and post-testing. Figure 3 also shows that the ***Time x Group*** factor indicates a strong difference in the amount of increase from pretesting to post-testing. The ***Science Fusion*** Group increased to a greater extent than did the Control Group.

Figure 7
Grade 5 Standard Score Increases for the Science Fusion Group and the Control Group from Pretesting to Post-Testing



A Paired Comparison *t*-test was computed to determine if the differences in gains made by each group from pretesting to post-testing. The figures in Table 13 support the findings from the Repeated Measures ANOVA in that both groups increased statistically significantly (≤.0001) but there was a difference in the standard score gains between the two groups as the Control Group increased by 36 standard score points and the ***Science Fusion*** group increased by 49 standard score point. The difference is reflected in the larger effect size for the Science Fusion Group although the effect sizes for both groups were large.

Table 13
Paired Comparison *t*-Tests Control and Science Fusion

Control Group (N=62)	***Mean***	***Standard Deviation***	***t Test***	***Significance***	***Effect Size***
Pretest	250.7	32.4	7.785	≤.0001	.91
Post-Test	287.0	45.7			
Science Fusion Group (N=172)					
Pretest	286.6	36.5	15.869	≤.0001	1.21
Post-Test	335.8	44.3			

High Pretest and Low Pretest Student Comparison

An analysis was next conducted to determine if both the ***Science Fusion*** Group and the Control Group increased the scores of the lower and higher pretest scoring groups. Table 14 shows that both the low pretest group and the high pretest group from the ***Science Fusion*** Group made statistically significant gains (≤.0001). The effect size was large for both groups.

Table 14
Science Fusion Group Paired Sample *t*-Tests
for Low and High Pretest Scoring Groups

Low Pretest Group (N=86)	Mean	Standard Deviation	t Test	Significance	Effect Size
Pretest	256.5	23.2	14.348	≤.0001	1.79
Post-Test	318.1	42.6			
High Pretest (N=86)					
Pretest	316.8	17.5	9.031	≤.0001	1.22
Post-Test	353.4	38.6			

Table 15 shows that both the low pretest group and the high pretest group from the Control Group made statistically significant gains (≤.0001). The effect size was large for both groups.

Table 15
Control Group Paired Sample *t*-Tests
for Low and High Pretest Scoring Groups

Low Pretest Group (N=31)	Mean	Standard Deviation	t Test	Significance	Effect Size
Pretest	225.0	16.9	5.962	≤.0001	1.52
Post-Test	263.9	31.9			
High Pretest (N=31)					
Pretest	274.7	23.8	5.039	≤.0001	.92
Post-Test	308.7	46.4			

Conclusions

This study sought to determine the effect of **Houghton Mifflin Harcourt's Florida *Science Fusion*© 2012** program on students' knowledge and skills in science. For this year-long study, **Houghton Mifflin Harcourt's Florida *Science Fusion*© 2012** was used with students at grades 3, 4, and 5. ***Science Fusion*** Classes were compared to Control Classes at all three grade levels.

At grade 3, the following results were determined;

Both ***Science Fusion*** classes and Control classes made statistically significant gains from pretesting to post-testing and effect sizes were large.

The ***Science Fusion*** classes were significantly higher than the Control classes at both pretesting and post-testing.

The low scoring and high scoring ***Science Fusion*** classes and the Control classes both made significant gains from pretesting to post-testing and effect sizes were large.

At grade 4, the following results were determined;

Both ***Science Fusion*** classes and Control classes made statistically significant gains from pretesting to post-testing and effect sizes were large. But, the ***Science Fusion*** classes made significantly greater gains from pretesting to post-testing than the Control classes. The effective size was large for the ***Science Fusion*** classes and small for the Control classes.

The low scoring and high scoring ***Science Fusion*** classes both made significant gains from pretesting to post-testing and effect sizes were large. The low scoring Control classes made significant gains with a large effect size; however, the gains for the high scoring group were non-significant and the effect size was negligible.

At grade 5, the following results were determined;

Both ***Science Fusion*** classes and Control classes made statistically significant gains from pretesting to post-testing and effect sizes were large. The Science Fusion classes made significantly greater gains from pretesting to post-testing than the Control classes.

The ***Science Fusion*** classes were significantly higher than the Control classes at both pretesting and post-testing.

The low scoring and high scoring students for both the ***Science Fusion*** classes and the Control classes made significant gains from pretesting to post-testing and effect sizes were large.

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

- ***The Houghton Mifflin Harcourt's Science Fusion©2012 is effective in improving students' knowledge and skills in science at grades 3, 4, & 5.***
- ***Houghton Mifflin Harcourt's Science Fusion©2012 program is effective in improving skills and knowledge in science of lower performing as well as higher performing elementary school students.***

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.