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 *ScienceFusion* © 2012

Report Number 380

Table of Contents

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.

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 in elementary and middle school. This report presents the findings from a control group/experimental group effectiveness study of a grade 2 and a grade 4 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 are enrolled in schools at varied socio-economic levels?
>
> 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 only grades at the elementary level—grades 2 and 4. A companion report for middle school describes the results for grade 7.

For this study, a single unit from the national field test edition of Houghton Mifflin Harcourt's *SCIENCEFUSION* © 2012 was used at each grade level for instruction with the experimental group students. The teachers participating in the experimental group of the study used the *SCIENCEFUSION* materials as their primary program for science instruction over a period of approximately two weeks. None of the participating teachers had used the program prior to their involvement in the study. Control group students continued to use their regular programs of study for science instruction.

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. At grade 2, the topic was animal and plant life cycles. At grade 4, the topic was physical and chemical changes of matter.

Five different schools in two states were included in the grade 2 sample of the study. Seven experimental classes, each taught by a different teacher, and four control classes, each also taught by a different teacher, participated. Three different schools in one state were included in the grade 4 sample of the study. Six experimental classes, each taught by a different teacher, and three control classes, each also taught by a different teacher, participated.

Experimental and control teachers were recruited from the same schools. This was done to attempt to control socio-economic and other differences between the experimental and control classes.

Upon completion of their participation in the study, teachers filled out a questionnaire that asked them about their use of the program during the study, in order to determine the fidelity with which they used the program materials. 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 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 11 grade 2 control and experimental classes and nine grade 4 control and experimental classes in the study. The 20 different classes were all from schools in Ohio and New York. The data provided in Tables 1 and 2 provide a demographic summary of the schools included at grades 2 and 4. 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 for the grade 2 classes the average school enrollment was 650 students. An average of sixty percent of the students was enrolled in free/reduced lunch programs and the minority enrollment average in the schools was 58%.

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

Location	*Grades*	*Students Enrolled*	*% Students Free/Reduced Lunch Programs*	*% Minority*	*% Special Education Needs*
Urban Fringe Large City	K to 5	333	33%	32%	11%
Urban Fringe Large City	K to 5	358	13%	18%	10%
Urban Fringe Large City	K to 6	534	85%	83%	0
Urban Fringe Large City	K to 12	1550	72%	59%	3%
Urban Fringe Large City	K to 6	476	99%	100%	0
Average		**650**	**60%**	**58%**	**5%**

Table 2 shows that for the grade 4 classes the average school enrollment was 457 students. An average of eighty-six percent of the students was enrolled in free/reduced lunch programs and the minority enrollment average in the schools was 94%.

Table 2
Demographic Characteristics of Grade 4 Schools Included in the Study

Location	*Grades*	*Students Enrolled*	*% Students Free/Reduced Lunch Programs*	*% Minority*	*% Special Education Needs*
Urban Fringe Large City	K to 6	534	85%	83%	0
Urban Fringe Large City	K to 5	361	75%	100%	5%
Urban Fringe Large City	K to 6	476	99%	100%	4%
Average		**457**	**86%**	**94%**	**3%**

Description of the Assessments

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

The grade 2 test included 32 three-option multiple choice test items assessing students' knowledge and understanding of the life cycle of plants and animals. Items included picture identification of life cycles and understanding of the basic steps in collecting and analyzing data.

Table 3 provides the test statistics for the grade 2 posttest. The reliability of the posttest shows that the test was reliable for making instructional decisions regarding student growth.

Table 3
Grade 2 Posttest Reliability Statistics

	Experimental	*Control*
Number of Test Items	32	32
Maximum Score	32	30
Minimum Score	8	7
Average Score	25.4	21.2
Percent Correct	79.3	67.4
Reliability*	.84	.80

**Kuder-Richardson 20*

The grade 4 test included 32 four-option multiple choice test items assessing students' knowledge of physical and chemical changes. Items included understanding and recognizing physical and chemical changes and identifying how such changes are important to humans.

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

Table 4
Grade 4 Posttest Reliability Statistics

	Experimental	*Control*
Number of Test Items	32	32
Maximum Score	29	22
Minimum Score	6	5
Average Score	19.6	11.0
Percent Correct	61.3	34.3
Reliability*	.83	.64

**Kuder-Richardson 20*

Data Analyses

The results for the grade 2 and grade 4 students were analyzed independently. Two primary analyses were conducted for each grade:

1. A comparison of the experimental and control groups' pretest and posttest scores sought to determine if they differed significantly from each other.

An analysis of variance (ANOVA) was performed in order to determine if the experimental and the control groups' pretest and posttest total test scores differed significantly. In addition, a comparison was made of the percentages of percentage of students in the experimental and control groups scoring at low, middle, and high levels on the pretest and posttest.

2. A comparison of the experimental group's pretest and posttest scores sought to determine if students demonstrated significant growth from pretest to posttest.

A Paired Comparison *t*-test was used to compare the pretest and posttest scores of the experimental group. Subgroup analyses for the experimental group based on pretest performance and on socio-economic status of the schools were also conducted using Paired Comparison *t*-tests.

Grade 2 Results

Control Group/Experimental Group Comparison

Researchers at ERIA conducted an Analysis of Variance (ANOVA) to determine if the differences in the scores for the control group and the experimental group were significant. A total of 74 control group students and 213 experimental group students were included in these analyses. The total test included 32 items, each worth one point. Pretest and posttest percent correct scores were analyzed. The .05 level of significance was used as the level at which differences would be considered statistically significant.

In addition to the ANOVAs, effect-size 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 5 indicates that the comparison of percent correct scores on the pretest for the experimental group and the control group resulted in non-significant differences.

Table 5
ANOVA Results Comparing the Test Percent Correct Scores for the Control Group and Experimental Group on the Pretest

Score	*Group*	*Number Students*	*Mean % Score*	*SD*	**F Test**	*Significance*	*Effect Size*
Percent	Control	74	67.6%	16.2%	.098	Non-Significant	--
Percent	Experimental	213	66.9%	15.7%			

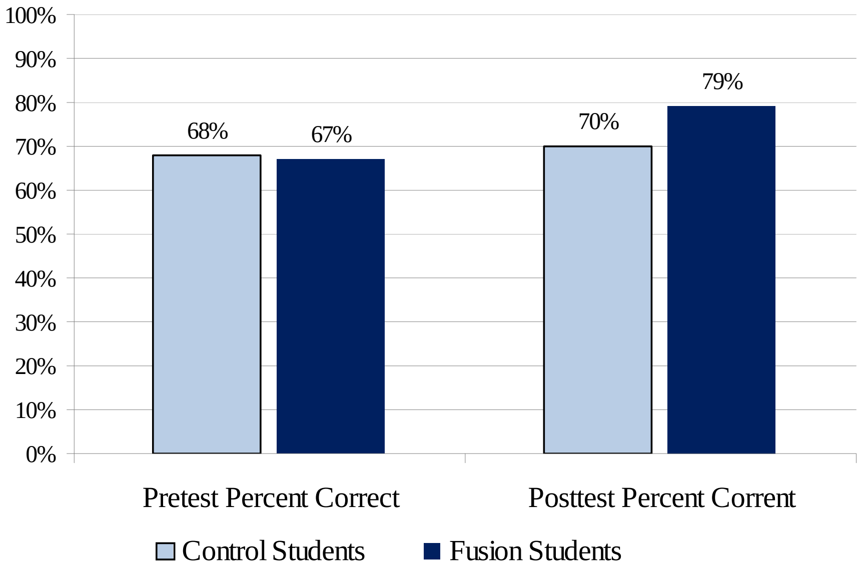
Table 6 indicates that the comparison of the percent correct scores on the posttest for the experimental group and the control group were statistically significant at the <.0001 level. This level indicates a difference that would occur by chance fewer than once out of 10,000 repetitions. In addition, the effect size was medium.

Table 6
ANOVA Results Comparing the Test Percent Correct Scores for the Control Group and Experimental Group on the Posttest

Score	*Group*	*Number Students*	*Mean % Score*	*SD*	**F Test**	*Significance*	*Effect Size*
Percent	Control	74	70.1%	14.4%	19.346	<.0001	.61
Percent	Experimental	213	79.3%	15.8%			

Figure 1 shows the correct percentage scores for the control and experimental groups from pretesting to posttesting. The tryout students increased their scores from 67% correct at pretesting to 79% correct at postesting. The control students increased their scores from 68% correct at pretesting to 70% correct at posttesting.

Figure 1
Percentage Correct Scores for Tryout and Control Groups From Pretesting to Posttesting



Experimental Group Pretest/Posttest Analysis

A Paired Comparison *t*-test was conducted to analyze the significance of the change from pretest to posttest in the average percent correct scores for the experimental group. A total of 213 students comprised the sample; eliminated were those who had only a pretest or posttest score.

Table 7 presents the results of this Paired Comparison *t*-test. The average percent correct score increased from 66.9% on the pretest to 79.3% on the posttest. The difference was statistically significant at the .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 7
Paired Comparison *t*-test Results for Pretest/Posttest Comparison of the Total Test Mean Percent Correct Scores for the Experimental Group

Results	***Group***	***Number Students***	***Mean% Score***	***SD***	***t-test***	***Significance***	***Effect Size***
Pretest	Total	213	66.9%	15.7%	12.259	<.0001	.80
Posttest	Total	213	79.3%	15.8%			

Subgroups by Socio-Economic Status—Pretest/Posttest Analyses

Classrooms from five schools were included in the grade 2 sample. These schools had differing percentages of students enrolled in free or reduced lunch programs. The percentages were 13%, 33%, 72%, 85% and 99%. While the school percentages do not necessarily indicate the percentage of students in the experimental population on free or reduced lunch programs, one can assume that the classroom numbers will be similar. Based on those figures, the total group was divided into two groups. The students in the experimental group from the schools with 72%, 85%, and 99% of the total school population on free or reduced lunch programs were considered lower socio-economic status, while those students from the schools with 13% and 33% of the students enrolled in free or reduced lunch programs were considered higher socio-economic status. These categorizations do not correspond to high and low socio-economic status among the general population of grade 2 students; the categorizations apply to these schools only. However, these analyses do provide an opportunity to determine if significant differences in the pretest/posttest comparisons can be found between subgroups based on socio-economic status.

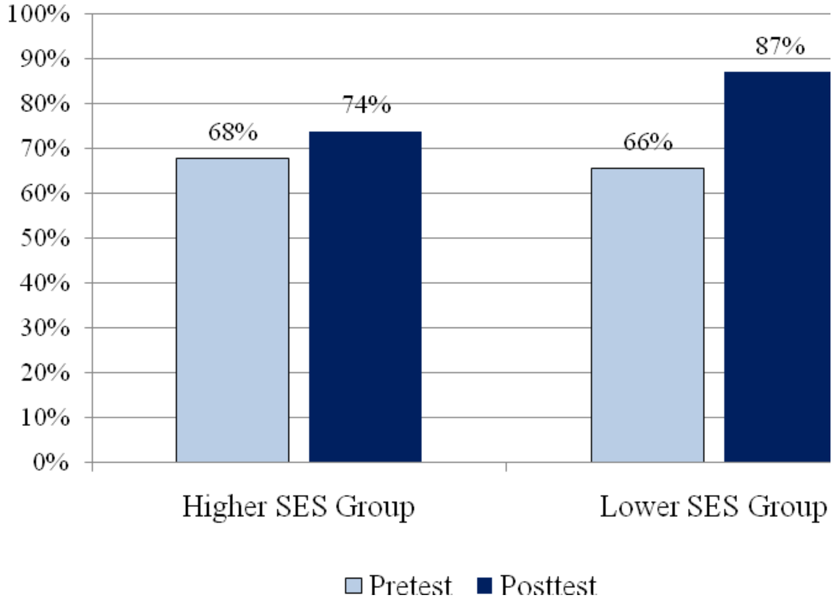
Table 8 presents the results of the paired comparison *t*-test performed for each of the two socio-economic groups to determine if the difference between the pretest and posttest total test percent correct scores was significant. The average percent correct score increased for both SES groups although the largest gain was achieved by the low SES group—from 65.5% on the pretest to 87.4% on the posttest. Both of the increases were statistically significant at the <.0001 level. 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 higher SES group was small while the effect size for the lower SES group was large.

Table 8
Experimental Group
Paired Comparison *t*-test Results
for Pretest/Posttest Comparison of the Total Test Mean Percent Correct Scores
for Subgroups Based on Socio-Economic Status

Test Form	*Number Students*	*Mean % Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Higher Socio-Economic Group						
Pretest	127	67.9%	17.7%	4.986	<.0001	.40
Posttest	127	73.9%	16.2%			
Lower Socio-Economic Group						
Pretest	86	65.5%	12.0%	18.541	<.0001	1.90
Posttest	86	87.4%	11.1%			

Figure 2 shows the increase from pretesting to posttesting for the higher SES group was 6 percentage points. The increase for the lower SES group was 21 percentage points.

Figure 2
Percentage of Higher SES Group Students
And Lower SES Group Students
Percentage Correct Scores from Pretesting to Posttesting



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 213 students were then divided into three equal groups of 71 students. The scores of the low pretest group ranged from 3% to 63% correct, the middle group scores ranged from 63% to 75% correct, and the high group scores ranged from 75% to 97% correct.

Table 9 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 about the same for the low and middle pretest groups. The high pretest group's pretest-to-posttest gain was not as large. Almost certainly, high pretest scores limited these students' opportunity for growth from pretest to posttest. A review of the data for the higher scoring group indicates that 27 of the 71 students in the group had scores of 95% correct or higher and 13 of the 71 students had scores of 100% correct.
The increases were statistically significant at the <.0001 level for the low and middle groups. 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 <.001 level. This level of significance indicates that such a change would have occurred by chance less than once out of 1,000 repetitions.

The effect size for the low and middle scoring groups was large and for the high scoring group the effect size was small.

Table 9
Experimental Group
Paired Comparison *t*-test Results
for Pretest/Posttest Comparison of the Total Test Mean Percent Correct Scores
for Subgroups Based on Pretest Performance

Results	*Number Students*	*Mean % Score*	*SD*	t-*test*	*Significance*	*Effect Size*
Low Pretest Group						
Pretest	71	49.3%	11.1%	10.176	<.0001	1.26
Posttest	71	69.2%	18.0%			
Middle Pretest Group						
Pretest	71	68.7%	3.6%	8.622	<.0001	.83
Posttest	71	81.7%	13.0%			
High Pretest Group						
Pretest	71	82.7%	6.3%	3.554	<.001	.28
Posttest	71	87. 1%	9.5%			

Figures 3, 4, and 5 show the pretest-to-posttest changes among these subgroups based on pretest scores. The figures show the percentages scoring below 60% correct, from 60% to 80% correct, and above 80% correct. Figure 3 shows that for the low pretest group the percentage of students scoring at the lowest level declined by 54% while the percentage scoring at the middle level increased by 19% and at the highest level by 35%.

Figure 3
Percentage of Low Pretest Group Students
Scoring Below 60% Correct, From 60% to 80%, and Above 80% Correct
on the Pretest and Posttest

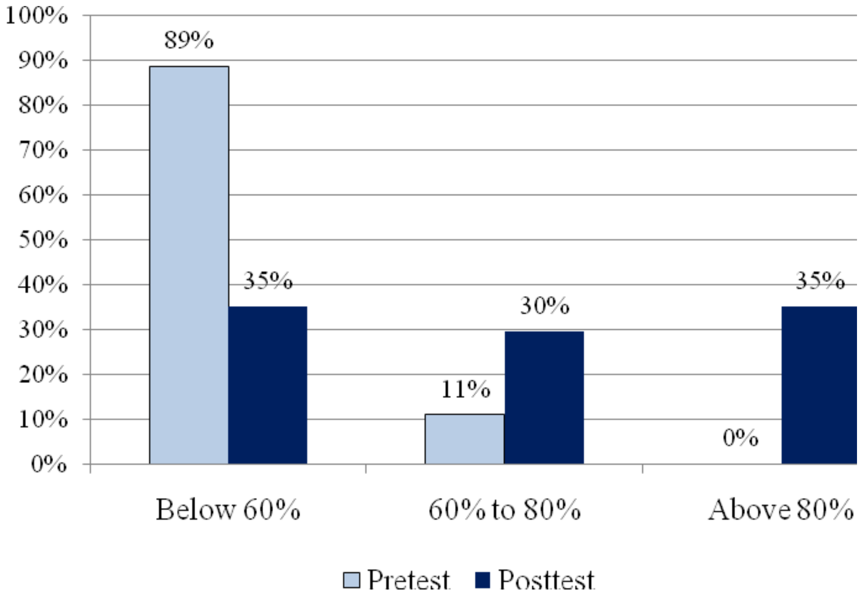


Figure 4 shows that for the middle pretest group the percentage of students scoring at the lowest level increased by 9% from pretest to posttest. The percentage of those scoring at the middle level decreased by 72%. The percentage scoring at the highest level increased by 63%.

Figure 4
Percentage of Middle Pretest Group Students
Scoring Below 60% Correct, From 60% to 80%, and Above 80% Correct
on the Pretest and Posttest

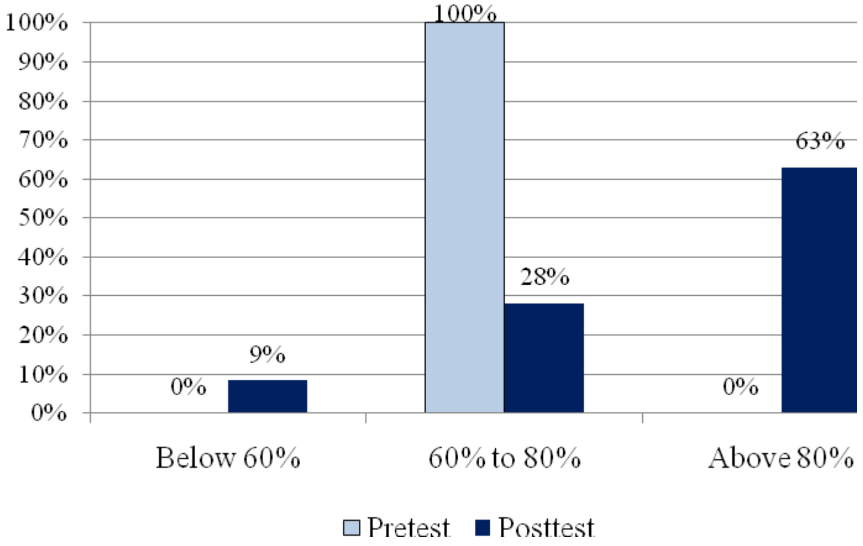
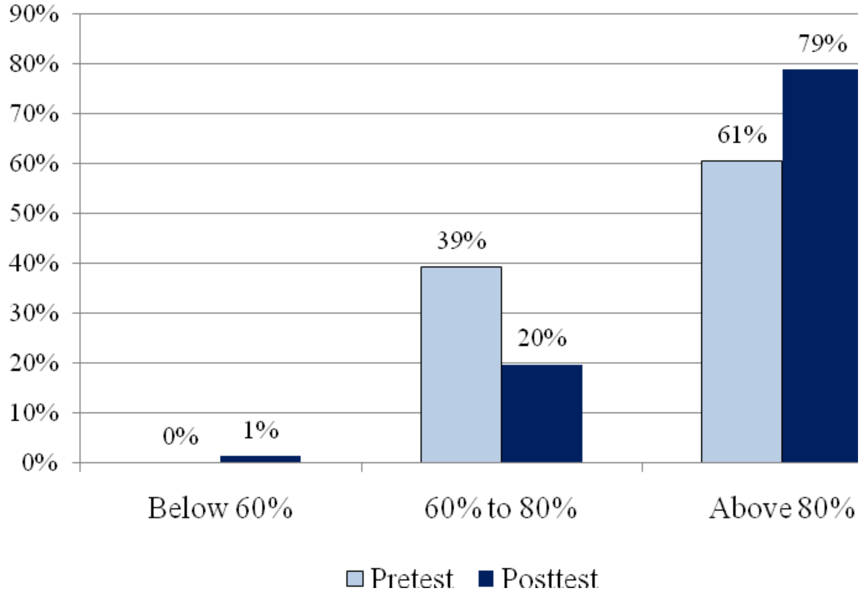


Figure 5 shows that for the high pretest group the percentage of students scoring at the lowest level increased by 1%. The percentage scoring at the middle level decreased by 19%. The percentage of students scoring at the highest level increased by 18%.

Figure 5
Percentage of High Pretest Group Students
Scoring Below 60% Correct, From 60% to 80%, and Above 80% Correct
on the Pretest and Posttest



Grade 4 Results

Control Group/Experimental Group Comparison

Researchers at ERIA conducted an Analysis of Variance (ANOVA) to determine if the differences in the scores for the control group and the experimental group were significant. A total of 48 control group students and 175 experimental group students were included in these analyses. The total test included 32 items, each worth one point. Pretest and posttest percent correct scores were analyzed. The .05 level of significance was used as the level at which differences would be considered statistically significant. In addition to the ANOVAs, effect-size 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 10 indicates that the comparison of percent correct scores for the experimental group and the control group resulted in non-significant differences.

Table 10
ANOVA Results Comparing the Test Percent Correct Scores for the Control Group and Experimental Group on the Pretests

Score	*Group*	*Number Students*	*Mean % Score*	*SD*	**F Test**	*Significance*	*Effect Size*
Percent	Control	48	30.9%	10.5	.201	Non-Significant	--
Percent	Experimental	175	30.3%	8.84			

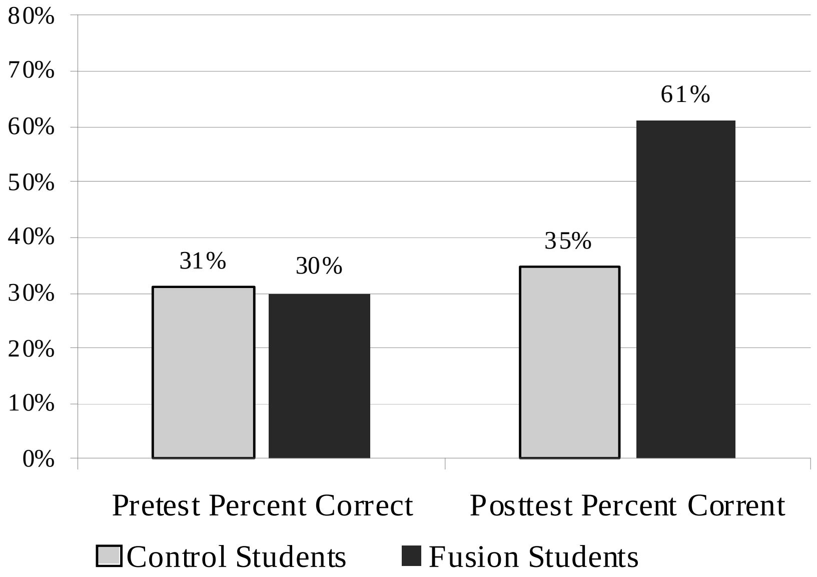
Table 11 indicates that the comparison of the percent correct scores for the experimental group and the control group were statistically significant at the <.0001 level indicating a difference that would occur by chance fewer than once out of 10,000 repetitions. In addition, the effect size was large.

Table 11
ANOVA Results Comparing the Test Percentage Correct Scores for the Control Group and Experimental Group on the Posttests

Score	*Group*	*Number Students*	*Mean % Score*	*SD*	**F Test**	*Significance*	*Effect Size*
Percent	Control	48	34.6%	13.13	95.103	<.0001	1.70
Percent	Experimental	175	61.3%	17.7			

Figure 6 shows the correct percentage scores for the control and experimental groups from pretesting to posttesting. The tryout students increased their scores from 30% correct at pretesting to 61% correct at postesting. The control students increased their scores from 31% correct at pretesting to 35% correct at posttesting.

Figure 6
Percentage Correct Scores for Tryout and Control Groups From Pretesting to Posttesting



Experimental Group Pretest/Posttest Analysis

A Paired Comparison *t*-test was conducted to analyze the significance of the change from pretest to posttest in the average percent correct scores for the experimental group. A total of 175 students comprised the sample; eliminated were those who had only a pretest or posttest score.

Table 12 presents the results of this Paired Comparison *t*-test. The average percent correct score increased from 30.3% on the pretest to 61.3% on the posttest. The difference was statistically significant at the .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 12
Paired Comparison *t*-test Results for Pretest/Posttest Comparison of the Total Test Mean Percent Correct Scores for the Total Experimental Group

Results	*Group*	*Number Students*	*Mean % Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Pretest	Total	175	30.3%	8.8%	23.869	<.0001	2.22
Posttest	Total	175	61.3%	17.7%			

Subgroups by Socio-Economic Status—Pretest/Posttest Analyses

Classrooms from three schools were included in the grade 4 sample. These schools had differing percentages of students enrolled in free or reduced lunch programs. The percentages were 75%, 85%, and 99%. While the school percentages do not necessarily indicate the percentage of students in the experimental population on free or reduced lunch programs, one can assume that the classroom numbers will be similar. Based on those figures, the total group was divided into three groups. The students in the experimental group from the school with 99% of the total school population on free or reduced lunch programs were considered lower socio-economic status, those students from the school with 85% of the students enrolled in free-reduced lunch programs were considered middle socio-economic status, and those students from school with 75% of the students on free-reduced lunch programs were considered higher socio-economic status. These categorizations do not correspond to high, middle, and low socio-economic status among the general population of grade 4 students; the categorizations apply to these three schools only. However, these analyses do provide an opportunity to determine if significant differences in the pretest/posttest comparisons can be found between subgroups based on socio-economic status.

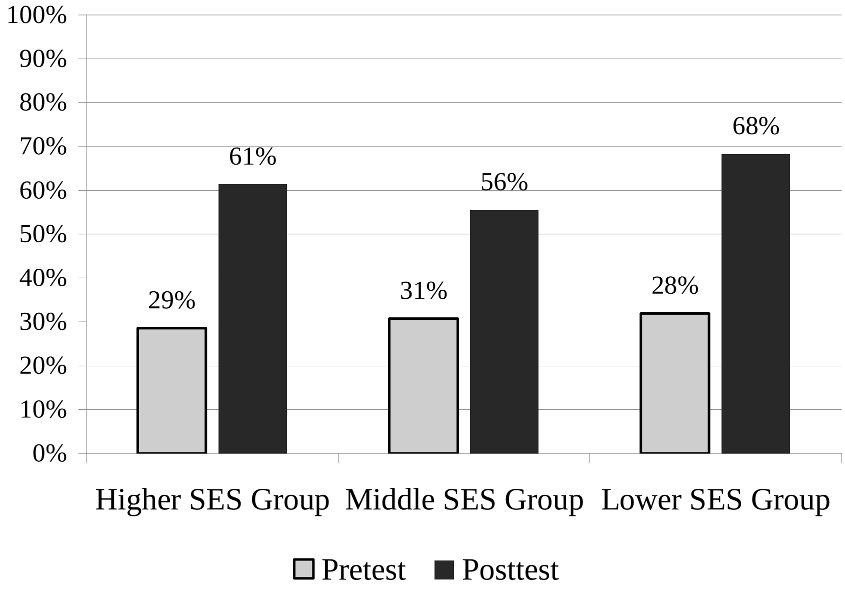
Table 13 presents the results of the Paired Comparison *t*-test performed for each of the three socio-economic groups to determine if the difference between the pretest and posttest total test percent correct scores was significant. The average percent correct score increased about the same for all three groups although the largest gain was achieved by the lowest SES group—from 31.9% on the pretest to 68.3% on the posttest. All of the increases were statistically significant at the .0001 level. 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 all three groups was large.

Table 13
Paired Comparison *t*-test Results Comparing the Experimental Group's Pretest and Posttest Total Test Mean Percent Correct Scores

Results	*Number Students*	*Mean % Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Higher Socio-Economic Group						
Pretest	70	28.6%	9.2	14.332	<.0001	2.30
Posttest	70	61.4%	19.0			
Middle Socio-Economic Group						
Pretest	56	30.8%	8.6	11.636	<.0001	1.91
Posttest	56	55.5%	16.0			
Lower Socio-Economic Group						
Pretest	49	31.9%	8.3	18.403	<.0001	2.99
Posttest	49	68.3%	15.1			

Figure 7 shows the increase from pretesting to posttesting for the higher, middle, and lower SES group. The increase for the higher SES group was 32 percentage points; for the middle SES group the increase was 25 percentage points; and for the lower SES group the increase was 36 percentage points.

Figure 7
Percentage of Higher SES Group Students
And Lower SES Group Students
Percentage Correct Scores from Pretesting to Posttesting



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 175 students were then divided into three approximately equal groups of 58, 58, and 59 students. The scores of the lowest pretest group ranged from 6% to 25% correct, the middle group scores ranged from 25% to 34% correct, and the high group scores ranged from 34% to 59% correct.

Table 14 presents the results of the paired comparison *t*-test performed for each of the three pretest groups. The average percent correct score increased about the same for all three groups although the largest gain was achieved by the lowest pretest group from 20.8% on the pretest to 55.7% on the posttest. All of the increases were statistically significant at the .0001 level. 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 all three groups was large.

Table 14
Paired Comparison *t*-test Results Comparing the Experimental Group's Pretest and Posttest Total Test Mean Percent Correct Scores

Results	*Number Students*	*Mean % Score*	*SD*	*t-test*	*Significance*	*Effect Size*
Lower Pretest Group						
Pretest	58	20.8%	4.7	15.116	<.0001	2.70
Posttest	58	55.7%	17.7			
Middle Pretest Group						
Pretest	58	30.3%	2.8	13.278	<.0001	2.47
Posttest	58	60.0%	16.6			
Higher Pretest Group						
Pretest	59	39.5%	5.3	13.223	<.0001	2.31
Posttest	59	68.0%	16.8			

Figures 8, 9, and 10 show the changes in student performance from pretest to posttest. The figures show the percentage of students in each of the three pretest groups scoring below 40% correct, from 40% to 60% correct, and above 60% correct. Figure 8 shows that for the lower pretest group the percentage of students scoring at the lowest level declined by 79% while the percentages scoring at the middle level increased by 31% and at the highest level by 48%.

Figure 8
Percentage of Lower Pretest Group Students Scoring Below 40% Correct, From 40% to 60%, and Above 60% Correct on the Pretest and Posttest

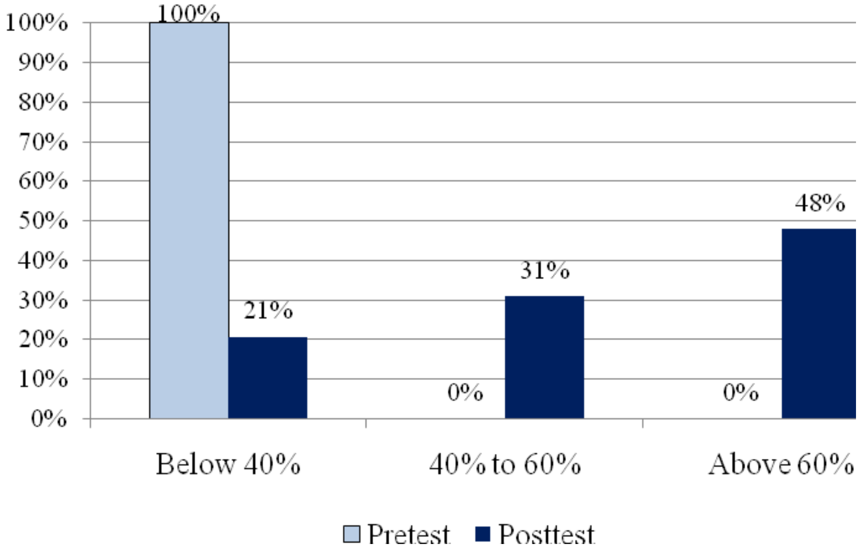


Figure 9 shows that for the middle pretest group the percentage of students scoring at the lowest level declined by 88% from pretest to posttest. The percentages of students scoring at the middle and highest levels increased, by 31% and 57% respectively.

Figure 9
Percentage of Middle Pretest Group Students Scoring Below 40% Correct, From 40% to 60%, and Above 60% Correct on the Pretest and Posttest

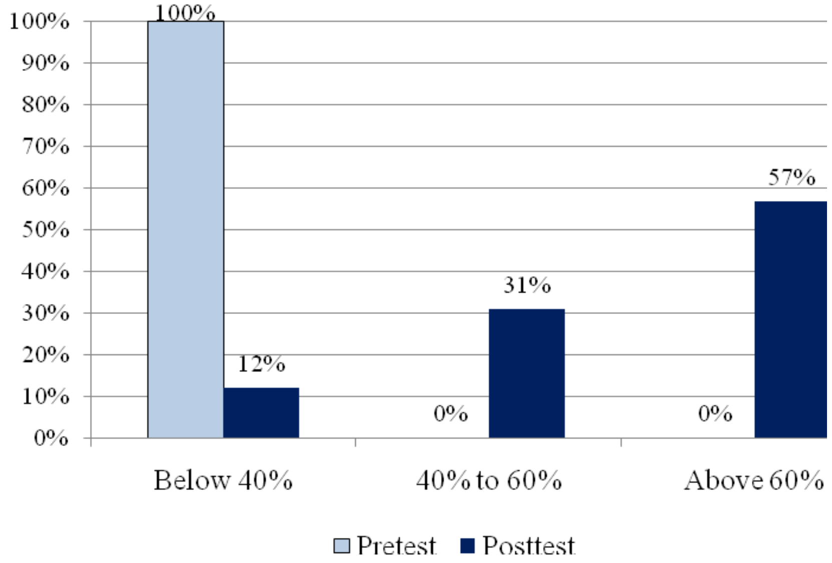
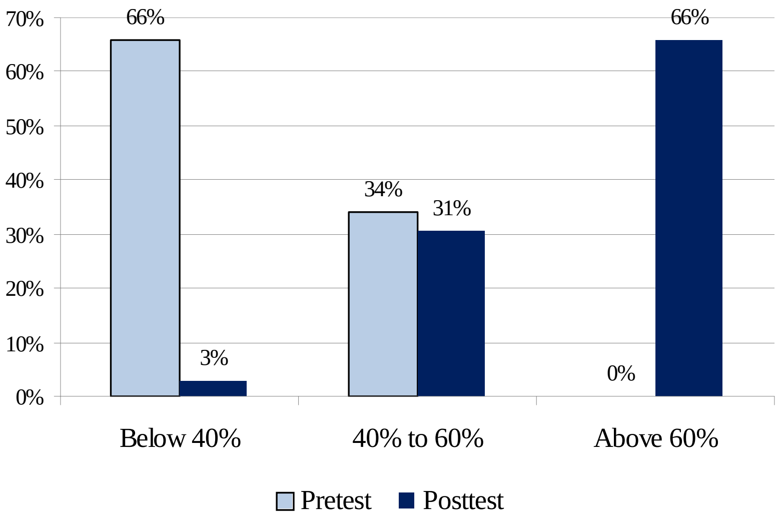


Figure 10 shows that for the higher pretest group the percentage of students scoring at the lowest level declined by 63% and the percentage scoring at the middle level decreased by 3%. The percentage of students scoring at the highest level increased by 66% from pretest to posttest.

Figure 10
Percentage of Higher Pretest Group Students Scoring Below 40% Correct, From 40% to 60%, and Above 60% Correct on the Pretest and Posttest



Teachers' Fidelity of Use and Program Evaluations

Table 15 summarizes the results of the survey of the teachers' fidelity of use of the program. The numbers of responses differ across the various questions because each teacher did not respond to each of the questions.

The teachers' experience ranged from two years or less to 10 years or more teaching at the current grade levels. All of the teachers reported feeling comfortable or mostly comfortable in teaching science. Class sizes ranged from 11 to 15 students to more than 25 students in a class. All of the teachers used the program for the two weeks for which the tryout was scheduled and the length of time each day ranged from 30 to 40 minutes 50 or more minutes each day.

Table 15
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	2	2	3		
Grade 4		2	1	2	
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	1	1	1	4	
Grade 4		2	2	1	
How would you classify your level of comfort with science content?					
	Comfortable	*Mostly Comfortable*	*Uncomfortable*		
Grade 2	6	1			
Grade 4	4	1			
Use of the Houghton Mifflin Harcourt *SCIENCEFUSION* 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		2		4	1
Grade 4			4		1
How many days did you use the materials?					
	Fewer than 5	*6 to 7*	*8 to 9*	*10 to 11*	*12 or more*
Grade 2				7	
Grade 4				3	2
How many minutes per day did you/your students use the program for science instruction?					
	Fewer than 20	*20 to 30*	*30 to 40*	*40 to 50*	*50 or more*
Grade 2			5	2	
Grade 4			1	3	1
How many minutes per day did you/your students use the digital content for science instruction?					
	Fewer than 10	*10 to 20*	*20 to 30*	*30 to 40*	*More than 40*
Grade 2		3	2		
Grade 4		3	1		

Teacher Comments

In addition to the ratings, teachers were asked to provide any comments they might like to make about the program. The following comments were provided by the teachers in the study.

The materials worked well, and students enjoyed the unit.

Student book was great. Excellent photos! Ease of reading for students. The books kept them engaged.

Easy to read, great pictures! The students were always engaged in the text.

Children loved the website. Easy for them to do by themselves! Loved the authentic photographs and pictures!

Great lessons! It was easy for the kids to navigate on their own.

The digital lessons were great. The kids loved this also, very beneficial to follow up and reinforce lessons taught.

The digital lessons were easy to access, interactive, good visuals, and the class wanted more!

The digital lesson was good—more would have been great. I was unable to get to the link from every computer.

The teacher edition was very helpful. It provided many extra ideas and resources.

I have no additional suggestions for improvement. I felt the materials were sufficient to develop understanding amongst virtually all of my students.

I liked how the student editions allowed them to include answers directly on the text pages.

Teacher Evaluations of Program Materials

Table 16 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.

Teachers rated the program as being *Very Successful* or *Somewhat Successful* except for one teacher who rated the program as *Somewhat Unsuccessful* in supporting below-level students' learning of the targeted science knowledge and skills. The grade 2 teachers rated each of the program components as being *Excellent* or *Good* while the grade 4 teachers rated the program components as *Excellent, Good* or *Fair*.

Table 16
Teachers' Evaluations of Program Materials

Program Evaluation					
In general, how successful was the information presented in supporting your on-level students in learning the targeted knowledge and skills?					
	Very Successful	*Somewhat Successful*	*Somewhat Unsuccessful*	*Very Unsuccessful*	*N/A*
Grade 2	7				
Grade 4	1	3			
In general, how successful was the information presented in supporting your below-level students in learning the targeted knowledge and skills?					
	Very Successful	*Somewhat Successful*	*Somewhat Unsuccessful*	*Very Successful*	*N/A*
Grade 2	1	5	1		
Grade 4	1	4			
Program Components Evaluation					
Please rate the program components in how well they support learning and instruction.					
	Excellent	*Good*	*Fair*	*Poor*	*N/A*
Field-Tested Teacher Edition Pages (Overall)					
Grade 2	2	5			
Grade 4	1	1	2		
Benchmark Alignment					
Grade 2	2	5			
Grade 4	1		3		1
Call-Out Information/Teacher Guidance					
Grade 2	2	5			
Grade 4	2	2	1		
Field-Tested Student Edition Pages (Overall)					
Grade 2	2	5			
Grade 4	1	1	1		
Key Terms (Highlighted)					
Grade 2	2	5			
Grade 4	2	2	1		
Follow-Up Questions					
Grade 2	2	5			
Grade 4	2	2	1		
Visual Aids (Pictures/Charts/Graphs)					
Grade 2	4	2	1		
Grade 4	1	4			
Applying Concepts (Activities/Investigations)					
Grade 2	2	4	1		
Grade 4	1	2	1		
Standards Test Practice and Questions					
Grade 2	2	5			
Grade 4	1	2	2		
Digital Lessons					
Grade 2	4	1			
Grade 4	1	1	1		1

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 at grades 2 and 4.

As can be seen in Table 17, for students in both grades 2 and 4, significant pretest to posttest gains were made for the total experimental group of students using the program, and for each of the subgroups based on SES and pretest performance. Furthermore, the effect sizes for nine of the 11 subgroup (by SES and by pretest performance) comparisons were large.

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 17
Summary of Significance of Control/Experimental ANOVA and Paired Comparison *t*-tests and Effect Sizes for Pretest/Posttest Gains on the Total Test Grade 2 and Grade 4

	Grade 2 Students		***Grade 4 Students***	
	Significance	**Effect Size**	**Significance**	**Effect Size**
Experimental/Control	<.0001	Large	<.0001	Large
Pre/Post Experimental	<.0001	Large	<.0001	Large
Subgroups Based on SES				
Higher SES Group	<.0001	Small	<.0001	Large
Middle SES Group	Not Included at Grade 2		<.0001	Large
Lower SES Group	<.0001	Large	<.0001	Large
Subgroups Based on Pretest Performance				
Higher Pretest Group	<.001	Small	<.0001	Large
Middle Pretest Group	<.0001	Large	<.0001	Large
Lower Pretest Group	<.0001	Large	<.0001	Large

The conclusion, based on a highly 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 the study who received instruction using a tryout unit of the program increased statistically significantly. These results are particularly significant considering the very short duration of the study (two weeks of program use) and the fact that the teachers had never used the program before.