A Study on the Effects of Houghton Mifflin Harcourt's *Science Fusion*: Year 2 Comprehensive Final Report

Prepared by:

Miriam Resendez, Senior Researcher

Dr. Mariam Azin, President

December 2013

PRES Associates, Inc.

Planning, Research & Evaluation Services

For inquiries, please contact PRES Associates at:

info@presassociates.com

(307) 733-3255

Executive Summary

Developing 21st century skills will be critical for students as they enter the workforce of tomorrow. They must learn to think critically, analyze complex situations and employ higher order thinking skills so that they'll be competitive in a global economy. The National Science Teacher Association (NSTA) has identified the central ways quality science education supports the development of these critical 21st century skills; this includes integrating deep content knowledge with inquiry based techniques that allow students to investigate in real-world settings and develop their problem solving and critical thinking skills (2011). However, research shows that students are not being adequately prepared which is leading to fewer U.S. students pursuing undergraduate degrees in science – the result is a shrinking pool of qualified and trained STEM workers (Xie and Killewald, 2009).

In order to more fully prepare students' with the skills they need to become successful in higher level science courses, as well as their futures, Houghton Mifflin Harcourt (HMH) Publishers has developed a new middle grades science program – Science Fusion (2011). Based on the conceptual framework "Enduring Understanding," this core middle school science curriculum incorporates prior research on effective science instruction and seeks to improve upon how science is being taught in classrooms by: 1) embedding inquiry-based teaching strategies into science instruction to promote higher-order thinking skills and student engagement; 2) designing a science curriculum that addresses the latest state and national science standards; and 3) promoting real-world connections so that students have ample opportunity to apply what they learn.

It is important that programs such as Science Fusion be looked at carefully to determine the extent to which they help students attain important science skills. Planning, Research, and Evaluation Services (PRES) Associates, Inc. conducted a two-year study designed to examine the effectiveness of the 2011 Science Fusion program in helping middle school students improve their science skills and understanding. This 2-year national randomized control trial (RCT), which commenced in the Fall of 2011, was conducted in the 6th-7th grades during the 2011-2012 school year and followed these students into the 7th and 8th grades during the 2012-2013 school year. This report presents the final results for the two year study.

A total of four schools participated in the first year of the study, and three schools participated in the second year of the study. The final sample in Year 2 consisted of 576 students (263 control; 313 treatment) in 27 classes (13 control and 14 treatment). Teachers/classes were randomly assigned to conditions prior to the 2011-12 school year (either use of the Science Fusion program or continued use of the science curricula currently available at the school).

Major findings, organized by the key evaluation questions, include:

Does science ability improve as a result of participation in the Science Fusion program?

Results showed significant growth in science performance over the course of both school years as measured by the national, standardized ITBS Science test and a developed science assessment aligned to the content covered during the school year as well as national and state standards. Science Fusion students grew by 14 percentiles on the ITBS Science test over the course of the two year study. In addition, while significant growth was observed during each study year on the Developed Science Test, growth during Year 2 was larger (14 points) than Year 1 (11 points). When tests for each ITBS content area were examined separately, Science Fusion students' showed significant improvement in Life Science, Scientific Inquiry and Earth Science. Furthermore, marginally significant growth was observed in Physical Science.

In addition, Science Fusion students experienced significant learning gains as measured by the science vocabulary and science application/reasoning items of the Developed Science Test, with larger gains observed during Year 2 as compared to Year 1. Taken together, these findings suggest a cumulative gain of the Science Fusion program on science performance, with higher levels of growth observed following two years of exposure to the program as compared to one year of exposure.

Do changes in science performance among Science Fusion students vary by different types of students (e.g., grade, gender, science level, economically disadvantaged status) and levels of implementation?

In contrast to Year 1 findings, results showed that Science Fusion students in all subpopulations examined showed significant learning gains on the Developed Science Test and ITBS, with one exception. High level students showed a significant decline on the Developed Science Test. In sum, during Year 2, males and females, 7th and 8th graders, students receiving free/reduced lunch those not, and students at various ability levels demonstrated significant learning gains in science.

Analysis by implementation fidelity showed that students whose teachers used the Science Fusion program with moderate fidelity showed the lowest gains as compared to teachers using the program with low and high fidelity. This was consistent across both the ITBS and Developed Science Test. This differs from Year 1 findings in which no significant relationship was observed between implementation fidelity and performance gains.

Does using Science Fusion result in increased student achievement as compared to other types of science programs?

Analyses were conducted on two samples: 1) all students participating in the second year of the study, including new students enrolled in participating study classes, and 2) students who participated in both study years and remained in the same study condition throughout. Results showed significant differences between students who used the Science Fusion program and students using other science programs as measured by the Developed Science Test (DST), after controlling for pretest differences. Specifically, Science Fusion students participating in both study years and those participating in Year 2 of the study showed significantly higher performance levels on the DST overall score as compared to students using other science programs. In addition, Science Fusion students outperformed control students on the fill-in-the-blank items designed to measure science vocabulary over both study years. As well, among students who participated in Year 2 of the study, marginally significant differences were observed on DST items measuring science application and reasoning. Given the lack of significant differences observed in Year 1 (only performance on the DST vocabulary items was significant), results suggest stronger effects of the Science Fusion program following two years of study participation.

On the ITBS Science test, although a similar trend was evident with Science Fusion students showing higher test scores than control students after controlling for pretest differences, no significant differences were observed. In summary, results from the second year of the study indicate that Science Fusion students outperformed control students as measured by the test designed to measure specific content areas that were covered over the course of the school year. It is also noteworthy that the effect sizes were moderate, with a range of .48 to .64. Indeed, all effect sizes obtained exceeded the threshold for educational significance (.25) which means that these findings are meaningful in terms of impacting a students' educational experience.

Do effects of Science Fusion on student science performance vary as a function of different student characteristics and control programs?

Results among subgroups of Year 2 participants (i.e., grade, gender, free/reduced lunch, and science level) showed that there were no significant subgroup effects. This means that there was no difference between treatment and control students within subgroups. It should be noted that the lack of significant differences may be due to the limited number of students within subgroups. For instance, in Year 1 when the sample size was larger, results showed that low-performing students who used the Science Fusion program demonstrated accelerated learning gains compared to control students on the ITBS and White students who used Science Fusion had higher test scores at post-testing on the Developed Science Test than White control students, after controlling for pretest differences.

Does participation in Science Fusion result in other positive outcomes (e.g., positive attitudes towards science, etc.)?

Affective positive outcomes were reported by both student and teacher users of the Science Fusion program. Science Fusion students reported more positive attitudes than control students relating to their science ability, enjoyment of science, and beliefs in the importance of science. More than control students and teachers, Science Fusion students and teachers agreed that the program positively impacted students' academic skills, especially problem-solving skills, scientific inquiry, and science-related math, reading, and writing ability. Science Fusion teachers also reported that their students were more interested and engaged in learning science than control teachers.

Teachers and students agreed that the Science Fusion program helped students make connections to the real-world more than the control program. The Science Fusion program was also reported to better prepare students to do well in high school, do well on science tests, and do well in future science courses than the control program.

Anecdotal information revealed that Science Fusion teachers were more prepared to give quality lessons, engage students, and provide differentiated instruction because of the Science Fusion program. Compared to control teachers, Science Fusion teachers reported that their science program helped them to minimize lesson preparation time and provided them with good ideas for activities.

What do users of the Science Fusion program think about the programs?

Users of the Science Fusion program were generally very positive about the program and all of its components. Teachers and students reported that the Science Fusion program was better than the previous science program and that they would like to use it again. Teachers gave high ratings to the program design and ease of use, commenting that it is an overall excellent system. Teachers were especially positive about the Teacher's Edition, student online textbook, and videos. Teachers and students agreed that the write-in student worktext is a valuable learning tool. The provided labs were well received by both students and teachers, with teachers noting the usefulness of the virtual labs. Teachers also reported that overall, using the Science Fusion technology was exciting and engaging for students and a good teaching experience for them.

In sum, results from this two-year RCT show that students who use the Science Fusion program perform significantly better than students using other science programs as measured by an assessment designed to measure specific content taught over the course of the school year. Such positive treatment effects were observed in multiple areas (vocabulary and scientific reasoning) and findings suggest a stronger effect following two years of usage of the Science Fusion program. This is to be expected given that it takes time for teachers and students to become accustomed to using a specific program and for effects to be realized. To conclude, the Science Fusion program has a positive impact on student science performance relative to other science programs.

Table of Contents

Table of Figures and Tables

Tables

Figures

Project Background

> ***"One of the things that I've been focused on as President is how we create an all-hands-on-deck approach to science, technology, engineering, and math… We need to make this a priority to train an army of new teachers in these subject areas, and to make sure that all of us as a country are lifting up these subjects for the respect that they deserve." - President Barack Obama, Third Annual White House Science Fair, April 2013***

It becomes increasingly clear that in order for today's students to be adequately prepared to succeed in the 21st century, where ongoing technological advancement, scientific innovation, increased globalization, shifting workforce demands, and the pressures of a global economy are the standard, they must acquire the related skills. The US will not be able to compete in a global economy if today's students do not acquire the skills that enable them to succeed in a 21st century workforce. The Partnership for 21st Skills (2009) and the National Research Council (2010) have recently identified ways that 21st century skills and science education intersect. Such linkages are rooted in the inquiry, process knowledge, experimental design, and scientific reasoning skills that transcend both areas. The National Science Teacher Association (NSTA) has also identified a number of ways quality science education supports the development of 21st century skills; this includes integrating deep content knowledge with inquiry based techniques that allow students to investigate in real-world settings and develop their problem solving and critical thinking skills (2011). Arne Duncan, U.S. Secretary of Education, recently asserted: "In the 21st century, scientific knowledge and inquiry are directly linked to our long-term economic and national security" (2011).

If students hope to achieve success in the larger world and ultimately contribute to a global economy, science proficiency is a requirement. STEM (Science, Technology, Engineering and Math) fields remain a critical area in education as well as in job creation, especially in this time of economic recovery. As so aptly stated in a 2013 report by the President's Council of Advisors on Science and Technology, "Economic projections point to a need for approximately 1 million more STEM professionals than the U.S. will produce at the current rate over the next decade if the country is to retain its historical preeminence in science and technology. To meet this goal, the United States will need to increase the number of students who receive undergraduate STEM degrees by about 34% annually over current rates." As well, according to a recent U.S. Department of Commerce report, STEM careers are expected to grow more rapidly than other occupations in the next 10 years (Langdon et al., 2011). Given these projections it becomes clear that our nation's students must be proficient in all STEM areas.

Despite the importance of scientific knowledge and skills, results from national assessments show that students continue to lag behind proficiency standards. According to recent National Assessment of Educational Progress (NAEP) results, only 2% of U.S. students in 2011 had attained advanced levels of science achievement by twelfth grade. Although the nation's eighth graders scored higher on the NAEP science in 2011 than in 2009, still only 32% scored at or above a level of proficient, and over one-third of our nation's eighth graders scored below a basic level. Cross-cultural comparisons, such as the 2012 Program for International Student Assessment (PISA), also indicate that American students are no longer in the top performers in science compared to other nations and that since 2009 science students in countries like Poland, Ireland and the Czech Republic outperform American students, while students in Lithuania, Spain, and Italy are now on par with American students.

In addition to these troubling statistics, other research suggests that our youth are falling behind due to a lack of interest in science. For example, research shows that students' interest in science tends to decline as they move from elementary grades to middle school and high school (Rani, 2006). As well, students who strongly exhibit attributes representing positive engagement in science tend to have significantly higher average National Assessment of Education Progress (NAEP) science scores (A NAEP Data Analysis Report, 2013). Researchers have also recently noted a decline in students pursuing undergraduate degrees in science, which results in fewer U.S. students with the necessary training to enter the workforce in STEM careers (Xie and Killewald, 2009). In sum, educators are challenged to capture and retain student excitement for science as students educationally advance, and are left with questions about how to improve students' academic achievements in science (Froschauer, 2006).

To help address the gap in students' science skills, Houghton Mifflin Harcourt (HMH) Publishers developed a new science program, Science Fusion. This program is designed to provide students with multiple ways to learn and apply science concepts and inquiry skills through curricula created specifically for a digital, print, and hands-on environment. For students, the Student Edition print materials promote active learning through constant interaction with the text. The new Write-in Student Edition challenges students to ask questions, think critically, make informed decisions and provides students with ownership opportunities, all of which create enduring understandings of science. The Write-in Student Edition also serves as a personal record of knowledge and a study guide for end-of-year exams.

For teachers, Science Fusion offers two powerful tools—a comprehensive Teacher Edition and the Teacher Digital Management Center. In the Teacher Edition, each lesson has a wealth of teaching support including activities, probing questions, misconception alerts, differentiated instruction, and vocabulary support. The lesson overviews give teachers the prerequisite knowledge to effectively plan lessons and include detailed learning goals, supporting concepts for each key topic, classroom organization information and estimated times for each activity. The Teacher Digital Management Center is a one-stop resource to plan, teach, assess, and track student progress. It serves as a quick interface for digital lesson and virtual lab access and allows teachers to preview, schedule and assign resources.

Given the increase in careers that rely heavily on STEM skills (Terrell, 2007), it is essential that programs designed to improve upon scientific skills be closely examined so as to inform the broader educational community. Planning, Research, and Evaluation Services (PRES Associates) Inc.[1], conducted a two-year randomized control trial (RCT) designed to examine the effectiveness of the Houghton Mifflin

[1] PRES Associates, Inc. is an external, independent, educational research firm with over 20 years of experience in applied educational research and evaluation.

Harcourt Science Fusion program in helping middle school students improve their scientific skills and understanding. The 2-year randomized control trial (RCT) on Science Fusion, which commenced in the Fall of 2011, was conducted in the 6th-7th grades during the 2011-12 school year and followed these students into the 7th-8th grades during the 2012-13 school year. What follows is a report that presents summative findings from the two year RCT.

Project Overview

The overarching purpose of this study was to rigorously evaluate the effectiveness of the 2011 HMH Science Fusion program in helping middle school students attain scientific knowledge and skills. Specifically, this study was designed to address the following research questions:

- Does science ability improve as a result of participation in the Science Fusion program?
- Do changes in science performance among Science Fusion students vary by different types of students (e.g., grade, gender, science level, economically disadvantaged status) and levels of implementation?
- Does using Science Fusion result in increased student achievement as compared to other types of science programs?
- Do effects of Science Fusion on student science performance vary as a function of different student characteristics?
- Does participation in Science Fusion result in other positive outcomes (e.g., positive attitudes towards science and so forth)?
- What do users of the Science Fusion program think about the programs? What aspects of the programs do they find most useful? Least useful? What, if any, suggestions for program improvement do they have?

This report presents descriptive information and results of the two year RCT. Specifically, the remainder of this report includes: 1) a description of the design and methodology; 2) sample and site information, including descriptions of Science Fusion implementation; 3) results of the two year evaluation; and 4) conclusions. In addition, Appendix A contains detailed statistical results of all baseline, attrition and assessment analyses conducted, including the analytical goals and framework employed.

Methodology

Designed to address all standards and criteria described in the What Works Clearinghouse (WWC) Study Review Standards (2008) and the Joint Committee on Standards for Educational Evaluation's Program Evaluation Standards (1994), the research design consists of a two-year randomized control trial, with random assignment of teacher/classes to a treatment (i.e., use of Science Fusion) or control group[2]. Other important design and methodological features include:

[2] Teacher/class level of random assignment was conducted for several reasons. From a research design perspective, it is desirable to conduct random assignment at the lowest level possible given both the nature of the intervention and the practical realities of the settings the research is being conducted in. In addition, using the lowest level of random assignment possible is a design strategy used to eliminate competing explanations for any observed differences and to enhance the ability of the study to make causal inferences.

- The study was conducted in the 6-7th grades during the 2011-12 school year (n=947 students), and extended to the 7-8th grades during the 2012-13 school year[3] (n=576 students). Such a study design allowed researchers to more fully examine the impact of the Science Fusion program over the span of two years. This is important because it takes time for teachers and students to become familiar with a new curriculum and for any effects to manifest.
- A total of 47 classes ($n_{treatment}$ = 27; $n_{control}$ = 20) participated in the first year of the study. As a result of the loss of one large school in the second year of the study[4], the sample size was reduced to 27 classes ($n_{treatment}$ = 14; $n_{control}$ = 13) in the second year of the study.
- Clear site selection criteria were established along with accompanying rationale.
- Extensive background data was collected on instructional activities and materials used in classrooms so as to describe the context in which science instruction took place.
- The threat of differential attrition was addressed via: 1) the initial site selection process[5]; 2) random assignment among teachers/classes within schools to help ensure that attrition was relatively constant across both treatment and control groups; and 3) the characteristics of students who left were statistically compared between treatment and control groups.
- Implementation guidelines and monitoring procedures[6] were embedded to ensure the fidelity of treatment implementation. Furthermore, monitoring mechanisms were put into place to address potential threats to validity such as contamination (i.e., students not assigned to use Science Fusion who end up using Science Fusion) and attrition (i.e., students dropping out). These included: a) site visits; and b) teacher monthly activity logs.
- Assessments measuring concepts in Life, Earth, and Physical Science as well as Science & Technology at the middle school level were developed based on released items from existing international, national and state science exams. In addition, the norm-referenced ITBS Science test was used. The assessments consisted of both multiple-choice and open-response test items that were aligned to content that is typical in middle school science courses.
- The study employed pre/post measures of, among other things: (1) student performance; (2) student attitudes regarding science; and (3) teacher characteristics, attitudes towards student learning, and perceptions of the Science Fusion program.
- Student assessments, surveys, and classroom observation forms are valid and reliable as shown by technical documentation and statistical analyses performed.

[3] The study was conducted at these grade levels in order to evaluate all Science Fusion modules. Since there is variation in terms of when science concepts are taught, having the study encompass all middle school grades allowed researchers to more fully evaluate this program.

[4] As a result of major restructuring of schools within the district, School D withdrew from study participation in 2012-13.

[5] Sites that historically had more than 20% student attrition were not used in the study.

[6] Training provided and implementation guidelines reflect how the Science Fusion program should typically be used in schools.

Table 1. Study Timeline of Activities for Year 1 and 2 of the RCT

2011-12	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.-Feb.	Mar.	April	May	June
Training and Program Implementation Begins	♦	♦								
Follow Up Training Occurred	*Varied for each site*									
Assessments and Surveys Administered		♦	♦						♦	♦
Site Observations			♦	♦	♦				♦	
Teacher Logs*		♦	♦	♦	♦	♦	♦	♦	♦	♦
2012-13	**Aug.**	**Sept.**	**Oct.**	**Nov.**	**Dec.**	**Jan.-Feb.**	**Mar.**	**April**	**May**	**June**
Program Implementation Begins	♦	♦								
Assessments and Surveys Administered		♦ (All took the Developed Science Test and only new students took the ITBS and Surveys							♦	♦
Site Observations				♦				♦	♦	
Teacher Logs*		♦	♦	♦	♦	♦	♦	♦	♦	♦

*Note that teachers completed monthly teacher logs that monitor instructional activities and the use of program and other resources.

- The study employed the use of statistical controls as well as random assignment to establish initial group equivalence[7]. Analyses of assessment data were primarily conducted via multilevel models to take into account clustering and baseline differences.

Table 1 displays the timeline for the important study activities during the RCT. More detailed information on these activities, as well as measures used, is provided in the following section.

[7] Random assignment helps to create group equivalence. However, it must be noted that with small sample sizes random assignment in and of itself does not assure initial group equivalence (Lipsey, 1990).

Measures

This section reviews the outcome and assessment measures that were administered, including descriptions of the items, and available reliability and validity information.

Student Assessments: In order to enhance the sensitivity of the RCT to detect any effects associated with the Science Fusion program, two assessments were used: (1) ITBS Science test; and (2) a custom developed science test. Following a thorough literature review of existing standardized, published assessments to identify tests that were valid, reliable, sensitive, and aligned to national science standards, it was determined that there were no readily available science assessments that fully captured the range of scientific knowledge and skills that students can potentially gain in middle school science classrooms. Assessments available typically consisted of state science exams that were aligned to specific state science standards, and/or did not give students adequate opportunities to explain their reasoning and to illustrate their analytical thinking process. As such, in addition to the ITBS, a supplemental assessment was developed that included fill-in the blank and constructed response test items.

- *Developed Science Assessments*: Prior to the study, information was obtained from participating schools on the science topic areas that would be covered during the school year for each grade level. Because coverage of science concepts varied across schools and across grade levels, an item bank was first created that covered typical middle school science concepts in Life, Earth and Physical Science, and Science & Technology. Items were then drawn from the item bank in order to customize assessments for each grade level and school; both treatment and control classes within the same grade level and school took the same version of the test. The assessments were worth 50 points and contained 30 multiple choice items, 10 fill in the blank items and 5 short answer items (each worth 2 points). The vast majority of items were drawn from released state science assessments, TIMSS, and NAEP, although in a very few instances custom-developed items were embedded to measure content taught.

- *ITBS:* The ITBS Form C Science test was also administered so that information on student performance could be obtained using a national standardized science test. The ITBS Form C exam is a norm referenced achievement test developed by Riverside Publishing, which was standardized in 2005 using a nationally representative sample. The Science test assesses students' knowledge of scientific principles and information as well as the methods and processes of scientific inquiry, in accordance with the recommendations of The American Association for the Advancement of Science (AAAS) and the National Science Teachers Association (NSTA). At Levels 9 through 14 are scenarios presenting real-life science investigations with questions emphasizing the thought processes used in designing and conducting research and in analyzing data. The four major content areas covered in the Science test are:

    - Scientific inquiry — Methods of science; analysis and interpretation of data
    - Life science — Structures and life cycles of living things; environmental interactions
    - Earth and space science — Earth's composition and structure and its changes; the universe
    - Physical science — Forces and motion; energy; properties of and changes in matter

Students were administered the science portion of the ITBS Level 12, 13, and 14 tests for grades 6, 7 and 8, respectively. The ITBS has demonstrated reliabilities ranging from .88 to .94 in the Fall.

In addition to content specific test scores in the areas of Life, Earth, Physical, and Scientific Inquiry, an overall scale score was calculated based on data from all ITBS test items taken. In order to obtain more specific information on the areas impacted by the Science Fusion program, fill-in-the-blank items (primarily measuring science facts and vocabulary) and open-response items (primarily measuring scientific reasoning skills and application of science concepts) from the Developed Science test were analyzed separately. For all analyses of custom assessments, percent correct was the metric used. For analyses involving overall performance on the ITBS Science test, the scale score was used. For analyses of the ITBS content areas, percent correct was the metric used.

Student Survey: In an effort to examine other potential areas that may be influenced by the Science Fusion program, a student survey was developed primarily to measure:

- Perceived science ability *(e.g. I'm good at science)*
- Enjoyment of science *(e.g. I look forward to my science class)*
- Perceived relevance/usefulness of science *(e.g. Science is a worthwhile, necessary subject)*
- Science- and school-related effort and aspirations *(e.g. I study hard for science tests)*

The survey also included items on parental knowledge and support, classroom experiences and, in the Spring survey, satisfaction with their science program. These scales were included in order to obtain measures of the impact of the Science Fusion program on affective student outcomes and to measure potential variables that may serve as covariates as needed (e.g., parental support). While some items were created by PRES Associates, others were derived from additional measures with published reliability and validity[8]. Internal consistency of the scales measuring attitudinal constructs range from .68 to .88. High scores represent a very positive attitude or strong agreement (scales are from 1 to 5).

Teacher Survey: Information was collected via surveys from all participating teachers. In addition to obtaining teacher background and demographic information, the survey was developed to measure:

- Current and past classroom and instructional practices
- Science-related preparation and knowledge
- Teacher knowledge of effective teaching practices (including those specific to science instruction)
- Organizational factors/context
- Attitudes about student learning and effective science instruction
- Attitudes about science curriculum

These measures were obtained to examine affective outcomes as well as to gather background information (e.g., years of experience, education, etc.). Some items were obtained from existing scales, while others were developed for the study[9]. Internal consistency of the scales measuring attitudinal constructs range from .79 to .92.

---

[8] Portions of this survey were adapted from the: *2003 TIMSS Student Questionnaire-8th Grade*; O'Neill and Abedi (1996) *Reliability and Validity of a State Metacognitive Inventory* (Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST)); and the *Fennema-Sherman Math Attitude Scale*.

[9] Items in this survey were developed by PRES Associates and modified from the *Trends in International Mathematics and Science Study (TIMSS) 2003 Teacher Questionnaire Science Grade 8* (Washington, DC: National Center For Education Statistics) and the *2000 National Survey of Science and Mathematics Education Science Questionnaire* (Rockville, MD: Westat).

High scores represent a very positive attitude or strong agreement (scales are from 1 to 5).

Classroom Observations: A classroom observation form was developed to guide observations. This form was largely based on existing protocols that have been used across the nation[10]. Modifications were made to reflect content and practices typical of middle school science classes, as well as to examine implementation of key components of the Science Fusion program. Researchers conducting site visits and using classroom observation forms were trained extensively until a high level of agreement (.90 and above) was demonstrated among observers on the various quantitative and qualitative items.

Procedures

To ensure that all treatment teachers participating in the study had sufficient knowledge and skills to successfully implement Science Fusion, teachers were provided with both implementation guidelines and Science Fusion specific training prior to implementation. In addition, monitoring procedures (via monthly instructional logs completed by teachers and classroom observations and interviews) were instituted to measure the extent to which teachers were implementing a similar instructional model as outlined by the Science Fusion program implementation guidelines.

The following section presents the procedures used to assist teachers in implementing the Science Fusion program, the monitoring procedures used by evaluators to determine treatment fidelity, methods used to obtain program feedback, and the test administration and scoring procedures employed.

Training

The training model for the Science Fusion study was designed to provide teachers with the necessary background and practical experiences to begin implementing the program with fidelity from the start of the 2011-2012 school year.

Teachers met with a Houghton Mifflin Harcourt professional trainer for approximately 5-6 hours at the start of the 2011-2012 school year. During the training, the trainer walked through all the print and digital components of the Science Fusion program and clearly indicated which components were considered "core" and which components were considered optional. A strong emphasis was placed on how the various pieces could be implemented in the classroom. The HMH professional trainer also helped teachers log in to their digital accounts and emphasized how the digital materials could be integrated into each lesson.

In addition to the initial in-depth training, follow-up sessions were conducted at two of the four sites. The follow-up training sessions were somewhat less formal than the initial training and allowed opportunities for teachers to ask questions and receive additional training on program components that felt they needed additional training. The follow up training for School C was a 3 hour in person training that focused on lab activities and the digital components. The follow up training for School B was a 3 hour webinar that occurred towards the end of the school year also focused on the lab activities. As well this training served as a refresher for the integration of the online

[10] The Classroom Observation Form was derived from the following protocols: Horizon Research's *Local Systematic Change Professional Development Classroom Observation Protocol*, and the *Texas Collaborative for Excellence in Teacher Preparation Classroom Observation Protocol.*

components. Schools A and D declined to receive any follow up training as they felt it was not necessary.

In year two of the study only School C requested an additional follow up training. This follow up training session took place via WebEx and lasted for approximately 2-3 hours. This training was to provide initial training for a new teacher and to address questions about program components and implementation of the remaining second year teacher. As well the training also allowed teachers additional instruction on using the Science Fusion program digital components. The remaining schools were provided with ongoing support from the publisher during year 2 of the study. Table 2 shows training received by each site during years 1 and 2 of the study.

Table 2. 2012-13 Training Sessions by Site

	Training Session1: Initial	Training Session2: Follow-up	Year 2 Follow-Up Training:
Site A OH	9/21	None	None
Site B OH	9/1	5/21	None
Site C DC	8/17	12/6	12/6
Site D RI – Year 1	9/7	None	

Another item of note is that the focus of these trainings was not on general science professional development but rather on the Science Fusion program (both print and digital materials), implementation of the essential components, and how the program could best be used to effectively help students learn science.

Implementation Guidelines

Science Fusion teachers were provided with detailed implementation guidelines at the onset of the study in order to ensure they had a concise understanding of the essential program components and design basis of the Science Fusion program. Teachers that were new to the study in year 2 were provided with the implementation guidelines at the beginning of the school year. Implementation guidelines were based on key program components and pedagogy. The guidelines were developed by PRES Associates with final input and revisions from HMH. These offered detailed direction on how the program should be used in the classroom, as well as what parts of the program were considered key (and required), versus what program elements were considered optional. The key components of the program include:

<u>Unit Activities</u>

- The Big Idea
- Advance Planning
- Summative Assessment (Unit Tests, Unit Review, Practice Tests – 1 per unit)

<u>Lesson Activities</u>

- Essential Questions
- Engage Your Brain
- Active Reading
- Lesson Activities & Discussion
- Labs/Demos (can select from Quick Lab, Daily Demo, Exploration Lab, and Virtual Lab)
- Visualize It!
- Think Outside the Box
- Do the Math
- Predict/Infer/Identify
- Lesson Vocabulary
- Formative Assessment (Strategies, Lessons Reviews, Quizzes – 1 per lesson)

For a full description of these key components, please see Appendix C.

Program Monitoring

Teacher Logs. Online teacher logs were used so that program implementation could be monitored on a real-time basis and to identify any issues or local events that had the potential to influence study results. Teachers were instructed to complete these on a monthly basis from September through May/June. The primary purpose of the teacher logs was to monitor program implementation and fidelity among Science Fusion classes. Researchers also collected monthly logs from control classes so instructional activities and content covered could be noted and also to monitor the extent to which any contamination may have occurred. Such background information provided researchers with a detailed data source on what was occurring in treatment and control classrooms with respect to science instruction and practices. It also allowed researchers to identify areas of overlap in terms of content taught and instructional activities. The extent to which there are similarities and differences between classrooms can have an impact on observed differences between treatment and control classes and effect sizes thus, it is important to take these factors into consideration when interpreting study results. Information obtained via these logs included changes in student rosters, typical classroom activities, use of other print resources and related exercises (including homework and independent practice), the degree to which technology was used and in what ways, use of labs and time spent on them, and coverage of science topics and content, and for treatment classes, use of key Science Fusion program components, both print and digital.

Results showed that teachers had, on average, an 87% completion rate during year 1 and 71% during year 2. The ranges were 11% to 100%[11]. Teachers were contacted after failure to complete teacher logs each month. In cases of noncompliance, the school liaison was asked to consult with the teacher to see if there was anything that could be done to assist the teacher in completing the logs and for the most part this was an effective practice with majority of teachers missing only one log on average. For teachers that did not have 100% completion rates, a more extensive implementation checklist was completed in the Spring to ensure that information on implementation was available from all teachers.

Classroom Observation. Classroom observations were conducted for treatment and control classes during the Fall (October-December, 2011 and November 2012) and the Spring (May 2012 and April-May 2013) of each school year. The purpose of these observations was to better understand the instructional approaches and materials used by teachers with their students and to identify differences and similarities between classes taught by teachers that were randomly assigned to treatment or control conditions. Specifically, observations focused on how classroom activities were structured, what and how print and digital materials were used, and characteristics of the class including student engagement, classroom environment and culture, and teacher-student interactions. In addition, teachers were interviewed after the observations to obtain more specific information on the representativeness of the lesson, resources used, ability levels of the students, assessment practices, pacing, independent practices, test preparation strategies and feedback related to the program. The observations also allowed

[11] Calculation based on 9 months in which teachers were asked to report on their activities. In year 2, two teachers had a low completion rate (11% and 22%,). All remaining teachers had over 80%-100% completion.

researchers to examine the extent to which class and teacher level differences could have influenced study results and to examine the threat of possible contamination between treatment and control classes.

Test/Survey Administration and Scoring

Assessments were administered during three to four time periods over the course of the study: (1) Fall 2011; (2) Spring 2012; (3) Fall 2012 (only new students and for the Developed Science Test, all students); and (4) Spring 2013[12]. For the ITBS Science test, student survey, and teacher survey, these were administered in Fall 2011, Spring 2012 (post-test 1), and Spring 2013 (post-test 2). Students new to the school in 2012-13 also took a "pre" ITBS assessment/ survey in Fall 2012. In administering the ITBS, teachers followed the test publisher's standard testing procedures. For the Developed Science assessments, students were administered the test in Fall 2011 (pretest 1), Spring 2012 (post-test 1), Fall 2012 (pretest 2), and Spring 2013 (post-test 2). Pre and post-tests were administered each study year since the content that was planned to be covered over the school year varied from one year to the next, and as a result, students were administered different assessments each year. In contrast, the ITBS is a vertically scaled test and although each level is different from one year to the next, direct comparisons over time can be made.

In both cases, teachers were instructed to contact PRES Associates if they needed additional guidance related to assessment administration. The short answer portion of the Developed Science test was scored by an external university student and former teacher who were blind to group assignment.

Site Selection Criteria

Criteria for developing an initial list of schools to be contacted for possible inclusion in the study included geographical diversity across different states, public schools, and a minimum school size of 600 so that a sufficient number of teachers would be available for purposes of random assignment. Additional criteria for study participation included:

- Schools had to have multiple science teachers at each grade level, or be willing to do class level random assignment if this was not the case;
- Historically low student mobility rates (less than 20%) as a means of helping control for the threat of attrition;
- Willingness/commitment to fully participate in all aspects of the study (e.g., random assignment and data collection).

Other major criteria included: 1) that there be no other major science initiative(s) at the school; and 2) the typical science curricula employed by the school fell under the "comparison" programs which provided a contrast to the Science Fusion program.

[12] Administration dates depended on the school's start and end date. Teachers within each school followed a similar testing schedule. Generally, administration occurred within 1 month after the school year commenced (pretest) and within 1 month prior to the end of the school year (posttest).

Sample Description

Site Characteristics

Four schools participated in the first year of the study, and three schools participated in year 2 of the study. As previously noted, one school (D) dropped out of the study following a district-wide decision to restructure schools which meant teacher layoffs and movement of students to different schools. Among the remaining schools, they were located in rural, suburban, and urban areas and in the states of Ohio and Washington DC. A detailed case study of each of the schools is available in Appendix D.

Table 3 on the following page shows the school-wide characteristics of each of the participating sites. As shown, the three school populations in Year 2 of the study were predominantly African American, and at all sites a substantial proportion of students were classified as economically disadvantaged. Indeed, the loss of School D meant less ethnic diversity. Characteristics specific to the study participants are provided in Table 4.

Table 3. School-Wide Student Demographics

School	School Size	Ethnic Breakdown	% Economically Disadvantaged	% by Gender
Site A ***Ohio*** **Grades 6-8**	724	92.6% Black, not Hispanic 3.2% Multi Racial	83.4%	50% Male 50% Female
Site B ***Ohio*** **Grades 7-8**	603	11.7% White, not Hispanic 81.6% Black, not Hispanic 4.7% Multi Racial	59.8%	50% Male 50% Female
Site C ***D.C.*** **Grades 6-9**	667	19.6% Hispanic 80.4% Black, not Hispanic	43.6%	52% Male 48% Female
YEAR ONE ONLY ***Site D*** ***Rhode Island*** **Grades 6-8**	1336	54% White, not Hispanic 27% Hispanic <1% American Indian 11% Black, not Hispanic 8% Asian/Pacific Islander	72%	51% Male 49% Female
National Population		**White-53.5%** **Hispanic-21.9%** **African Am.-17.6%** **Asian/Pacific Islander-5%** **Native American 1.2%** **Other 0.5%**	**45.4%**	**Male-50.8%** **Female-48.0%**

Data on National Population was obtained from U.S. Department of Commerce, Census Bureau, Current Population Survey (CPS), and U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD). Figures represent distributions across all grade levels and reported for 2009. School data obtained from respective State Department of Education websites. NR=Not Reported

Student Characteristics

The final analytical sample consisted of 576 students who participated in Year 2 of the study (263 control; 313 treatment) in 27 classes (13 control and 14 treatment). In contrast, the first year sample consisted of 947 students (383 control; 564 treatment) in 47 classrooms (20 control; 27 treatment). As previously noted, the decrease in sample size is largely due to the loss of School D.

Table 4 presents the demographic distribution among Year 2 study participants. Note that only students who remained in the study throughout the year are included in this table and in the analyses. The majority of students were African American (91%), with a high proportion of students receiving free/reduced lunch (47.5%).

Table 4. Student Demographics Distributions* - Year 2 Sample

Characteristics		Control (n=263)		Science Fusion (n=313)		Total (n=576)		National
		Count	Percent	Count	Percent	Count	Percent	Percent
Gender *($\chi^2(1)$=0.004, p=.95)*	Male	96	46.2%	116	45.8%	212	46.0%	50.2%
	Female	112	53.8%	137	54.2%	249	54.0%	49.8%
Ethnicity *($\chi^2(4)$ =7.26 p=.12)*	White	5	2.9%	3	1.3%	8	2.0%	55.0%
	Hispanic	3	1.7%	15	6.5%	18	4.4%	21.5%
	African American	163	93.7%	207	89.2%	370	91.1%	17.0%
	Asian	2	1.1%	4	1.7%	6	1.5%	5.0%
	Other	1	0.6%	3	1.3%	4	1.0%	1.2%
Grade *($\chi^2(1)$=0.13, p=.72)*	7th	133	50.6%	163	52.1%	296	51.4%	--
	8th	130	49.4%	150	47.9%	280	48.6%	--
Subpopulations								
($\chi^2(1)$=1.32, p=.25)	Free/Reduced Lunch Status	77	44.3%	116	50.0%	193	47.5%	45.4%
($\chi^2(1)$=2.19, p=.14)	Limited English Proficiency	2	1.1%	8	3.4%	10	2.5%	9.6%
($\chi^2(1)$=8.32, p=.004)	Special Ed Status	22	12.6%	11	4.7%	33	8.1%	13.6%
($\chi^2(2)$=0.40, p=.82)	Low Science Level	128	53.8%	165	56.5%	293	55.3%	--
	Mid Science Level	87	36.6%	100	34.2%	187	35.3%	--
	High Science Level	23	9.7%	27	9.2%	50	9.4%	--

*Counts (and percents) do not include missing information. Ability level was determined by percentile standing on the ITBS pretest. Students scoring at the top 33rd percentile were classified as high, students scoring at the bottom 33rd percentile were classified as low, and students scoring at the middle 66th percentile were classified as mid level.

Preliminary analyses[13] were performed to examine whether baseline differences existed as a function of student demographics. Chi-square analyses on the demographic characteristics noted in Table 4 showed one significant difference, $p>.05$[14]. There were more special education students in the control group than treatment students. However, it should also be noted that the number of special education students was small (22 in control and 11 in treatment). Overall, students were comparable in terms of demographic characteristics.

Differences in baseline science performance were also examined based on analyses of pretest scores. Student level t-test analyses revealed two significant differences on the ITBS and Developed Science Tests, $p<.05$, see Table 5. Treatment students had significantly lower pretest scores than control students on the ITBS Science overall scale score and significantly higher scores on the Developed Science Test-short answer items. Thus, treatment and control students were not equivalent with respect to pretest science performance on these measures.

Table 5. Sample Size, Means, Standard Deviations, and t-test (Student Level) Results for Assessments at Pre-testing

Pretest*	Group	N	Mean	Std. Dev.	*t*	Sig. Level
ITBS Science Test - Overall	Control	236	211.57	30.14	2.513	.012*
	SF	288	204.59	32.80		
ITBS Scientific Inquiry	Control	236	32.74	18.95	1.112	.267
	SF	288	30.92	18.26		
ITBS Life Science	Control	236	32.03	19.11	1.159	.247
	SF	288	30.07	19.31		
ITBS Earth Science	Control	236	29.31	19.53	-.033	.974
	SF	288	29.37	22.64		
ITBS Physical Science	Control	236	38.82	20.41	-.347	.729
	SF	288	39.47	22.35		
Developed Science Test (DST)-Overall	Control	232	30.61	12.62	-.728	.467
	SF	288	31.38	11.46		
DST Vocabulary (fill in the blank items)	Control	232	36.85	24.10	-.803	.422
	SF	288	38.54	23.61		
DST Science Application (constructed-response items)	Control	232	6.77	10.42	-3.237	.001*
	SF	288	9.90	11.37		

*$p<.05$

[13] All details regarding analyses on baseline differences and attrition analyses are provided in Technical Appendix A.

[14] "Significant" means that we can be 95% or more confident that the observed differences are real. If the significance level is less than or equal to .05, then the differences are considered statistically significant. If this value is greater than .05, this means that any observed differences are not statistically significant and may be interpreted as inconclusive. However, at times this may be referred to as "marginally significant." In this case, the criterion is more liberal and means that we can be 90% or more confident that the observed differences are real.

Differences on other student characteristics were also examined. Results showed no significant differences between treatment and control students in perceived parental support, mother's educational background, father's educational background, amount of English spoken at home, participation in extracurricular activities, school engagement, perceived science ability, science enjoyment, science effort/motivation, science anxiety, interest in STEM careers, and educational aspirations, $p>.05$. However, there was one significant difference on perceived importance of science, with treatment students reporting greater perceptions about the importance and usefulness of science than control students, $p<.05$. In sum, as a result of baseline differences on the ITBS and Developed Science Test, analyses of program effects controlled for pretest differences.

Attrition Analysis

Both measurement attrition (i.e., missing data due to students not completing assessments) and dropout attrition (i.e., missing data due to students leaving the study) were examined. Details on the attrition analysis are presented in Technical Appendix A, and are summarized herein. There was a dropout attrition of 8.7% in Year 1 and 10.4% in Year 2[15] due to students leaving school during the study year. Unlike the first year of the study, during the second year of implementation, drop-out behavior was similar between groups and no significant relationship was observed. However, further analyses of the students constituting Year 2 participants showed a significant relationship by group. Specifically, there were a higher percentage of new students in the control group (39%) than treatment group (30%), and conversely, there was a higher percentage of students who participated in both years of the study in the treatment condition (70.3%) than the control condition (61%). The high number of "new" students can be attributed to the fact that data was not collected on all students within science classes in Year 1 (only randomly selected classes were included in the study). While schools were asked to try to move all Year 1 study students to participating study classes in Year 2, this did not always occur and therefore, study classes contained a mix of "new" students and students from Year 1.

Among students participating in both study years (n=381), 16.8% (n=64) students changed conditions (i.e., students who were in Science Fusion program in year 1 were placed in a class that was in the control condition, or vice versa). Chi-square results showed a significant relationship in the proportion of students who changed conditions. Namely, a higher proportion of students who were in the control condition in year 1 were transferred to a treatment class in year 2 (21.5%) as compared to treatment students moving to the control condition (13.5%).

In order to determine whether such differences were associated with *performance* differences (which could potentially bias results), additional analyses were performed using outcome data. Results showed significant interactions between group and condition change on the outcome measures. Specifically, treatment students who remained in the study in both years and did not change conditions had lower baseline scores than control students. As such, the threat is not in favor of the treatment group and any significant differences would have occurred despite having lower performing students in the

[15] Of note, the RI school district underwent a major reorganization for the 2012-13 school year and decided not to participate in Year 2 of the study as a result. These counts are excluded from this dropout attrition rate.

treatment group. Indeed, outcome analyses controlled for pretest scores.

With respect to measurement attrition, chi-square analyses showed significant relationships between the proportion of students who provided and did not provide data and group. Specifically, a higher proportion of *treatment* students did not provide Developed Science Spring 2013 posttests and a higher proportion of *control* students did not provide ITBS Spring 2013 posttests. Similar to the dropout attrition analysis, additional analyses were run to examine if there were any *performance* differences between those who completed tests and those that did not by group. Results showed no significant interaction which means that results are unlikely to be biased due to measurement attrition.

Teacher and Class Characteristics

There were 7 middle school science teachers who participated in the RCT during Year 2 (16 participated in Year 1). Teachers taught a total of 27 classes (14 treatment and 13 control). While 2 teachers were randomly assigned to conditions, 5 teachers had classes that were randomly assigned and therefore, these teachers taught Science Fusion and another science program depending on the class period.

Approximately 86% of Year 2 teachers were female and 57% were Caucasian. In regards to educational background, 43% of teachers held a Bachelor's degree and 57% of teachers held a Master's Degree, primarily in Science and/or Education. Teacher experience ranged from 1 to 7 years, with the average number of years taught being 4.

Given that primarily classrooms within teachers were randomly assigned, teacher differences were minimized. Nevertheless, teacher data was examined at the classroom level to determine if significant differences existed. Results showed no significant difference between teachers in terms of perceptions of autonomy in setting instructional goals, extent to which different types of students may hinder teaching, preparation to teach various science topics, pedagogical leanings, comfort with technology, access to resources to teach science and knowledge of NTSA standards, $p>.05$.

Classroom environment and implementation of various typical activities that occur in science classrooms were also analyzed based on information collected from the teacher logs and teacher surveys collected in the Fall of each school year (data from Year 1 and Year 2 teachers were combined). Results showed no significant differences between treatment and control classrooms in terms of student engagement, independent practice, lab activities, provision of differentiated instruction, assessment use, and prior technology use by teachers and students. However, differences were observed in the areas of classroom environment and teachers' engagement in intervention activities, $p<.05$. Treatment teachers reported that they had a more positive classroom environment and engaged in more intervention activities than control teachers at baseline. As a result, these two factors were accounted for in analyses of program effects.

In summary, randomization was reasonably successful in producing equivalent treatment and control groups in terms of student and teacher characteristics. However, given significant differences among a few variables including pretest differences, care was taken to include variables that differed across the treatment and control groups as covariates in the analyses of program effects. Specifically, the following covariates were identified for inclusion in the multilevel model of program effects: 1) classroom environment, 2) engagement in intervention activities, 3) school, and 4) pretest science performance.

Instructional Curricula

Researchers tried, to the extent possible, to select schools to participate in the study that used a control program that differed pedagogically from the intervention under study. Indeed, part of the site selection criteria included a review of the control curricula prior to approving a site for participation, to determine if the program was sufficiently distinct. For the Science Fusion RCT, participating schools used four distinct science programs. As well, teachers at Schools A and D used additional science resources such as other commercial science resources, websites, and teacher-created materials. Most teachers taught a spiral curriculum covering various aspects of Life, Earth and Physical Science; however, depending on the school and grade level, different science concepts were taught due to state and local curricular guidelines which are typically aligned to state assessments. It is also important to note that within schools, there were similarities in content covered between treatment and control classrooms as teachers had to cover similar concepts regardless of the program used. The focus of this study was to examine the effects of an entire core curriculum and as such, it must be compared to other core curricula that teach the same content area.

Science Fusion

The Science Fusion program is designed to give students a meaningful way to study science by blending hands on, print and digital components. According to the publishers, the program delivers a holistic science experience and promotes 21st century learning skills by providing students with a self–paced, individualized learning experience.

Based on Schema Theory, the theory that knowledge is organized and stored into units, Science Fusion promotes Enduring Understanding through Big Ideas and Essential Questions. Enduring Understanding is achieved by providing students with an opportunity for active involvement in their own learning. The Big Ideas define the units of Science Fusion that articulates the overarching teaching and learning goals. The Essential Questions define the conceptual focus of the lesson that contributes to the growing understanding associated with the Big Idea. The goal is to provide students with the experience and tools to develop enduring understanding in science.

The units and lessons within the Science Fusion program are organized around the 5E's: engage, explore, explain, extend and evaluate. The 5E model is designed to emulate the elements of how an actual scientist works. Other aspects of the Science Fusion program includes, various levels of inquiry , the variety of lab activities, embedded response to intervention techniques, project based learning as well as active reading sections to keep readers focused comprehending and remembering.

To accomplish the goals of the Science Fusion program, resources were designed to integrate digital technology, emphasize hands on inquiry, and provide differentiated learning, all of which are essential components of the program.

Resources include:

Student Resources

- Write-In Student Edition
- eLearning Curriculum

Teacher Resources

- Teacher's Edition
- Lab Resource Book
- Assessment Guide
- ExamView® Test Generator CD-ROM
- Professional Development at hmhelearning.com

Digital Resources

- Teacher Access
- Electronic Teacher Edition (eTE)
- Teacher Resource Bank
- Student Access English
- Electronic Student Edition (eSE)
- Lab Manuals
- Assessment Guide
- ExamView Test Banks
- Unit Quiz
- Online Assessments
- Power Notes Presentations
- Science Fusion Glossary
- History Channel Videos
- People in Science
- NSTA Learning Center and Science Links
- Video Based Projects

A key feature of the program is that the Student Edition is a write-in resource, which gives students the opportunity to write in their own book and keep track of notes without having additional papers or folders. Generally, the pacing of the program is about one lesson per 4-5 days for a typical 45 minute class or about 4 weeks per unit.

For a more detailed description of the program's key features and materials, see Appendix C-Implementation Guidelines.

Control Curricula

The type of control curricula used by teachers varied between sites. Table 6 shows the programs used at each of the sites. The control programs varied across the school sites with Schools A and C using program 1, School C also using program 2, School B using program 3, and School D (Year 1 only) using program 4. Teachers at Schools A and B used the same program across grade levels. Teachers at Sites C and D combined their control program with teacher-created materials, other commercial resources (e.g., from websites), and resources they had collected over the years.

Control program 1 (2005/2006) uses a modular chapter based arrangement of lessons built around big ideas and hands on learning activities that reinforce key concepts. The program emphasizes a research based approach to learning that connects big ideas to the real world. Each lesson begins with an engaging section opener that connects new learning to prior knowledge. Lessons include information rich visuals that connect to the text and support student learning. The program also includes built in assessment activities with student self-checks for comprehension and built in vocabulary activities. This program is a basal instructional program that provides the option to bring in additional hands on resources. With respect to usage at School C, the 7^{th} grade control teacher used this program while also supplementing regularly with teacher-created materials and other resources collected over the years. The 6^{th} grade control teacher at School C used this program along with program 2 (described subsequently) but also supplemented with other online resources. Program 1 was used almost exclusively within control classes at School A, with the teacher supplementing with outside resources only on occasion.

Table 6. Primary Control Curricula by Site

	Program 1	Program 2	Program 3	Program 4 (Year 1 Only)	Teacher- created / Web / Other resources
Site A: OH	Science-2005				
Site B: OH			Science-2005		
Site C: DC	Science 2006	Science 2001			X
Site D: RI (Year 1 Only)				Science 2002	X

Similar to control program 1, control program 2 (2001) uses a modular chapter based arrangement of lessons that include lab activities and opportunities for self-assessment. The program emphasizes a connection to other content areas of science to develop a greater understanding of science in real world contexts. Each chapter begins with a full length lab investigation to introduce the topic through a hands-on experience. Each lesson includes a quick lab activity, math activities that integrate math and science and an "Apply" feature to connect student knowledge to the real world. The program also includes feature articles following every chapter emphasizing Science and Technology. While the program is a blend of basal and inquiry teaching approaches, it leans more towards basal instruction for the core, while providing the option to bring in additional investigations as desired. As previously noted, the 6th grade control teacher at School C used this program along with program 1 and supplemented with other online resources.

Control program 3 (2005) uses a modular chapter based arrangement of lessons that include lab activities and opportunities for inquiry. These inquiry activities include chapter projects, discovery and exploration activities, activities that reinforce key concepts, inquiry skills practice and at home lab activities. The program emphasizes interdisciplinary exploration and the integration of other academic subjects. Each lesson includes an introduction to key lesson topics, engaging introductory activity, lab activities, reading guide, connections to other academic subjects, built in learning checks and review. The program also includes math practice activities and sections that emphasize Technology and Design. The program is a basal program with inquiry teaching approaches that includes a variety of hands-on projects and activities to enhance inquiry. The control teachers at School B used this program as their main resource, supplementing on occasion to address additional needs.

Control program 4 is a magazine style text that contains numerous nonfiction readings designed to present scientific ideas in a unique way to expand student understanding. Unlike control programs 1, 2 and 3, control program 4 is organized in a magazine style arrangement with stories and articles relating to the overall topic. The program includes links to other subject areas such as language arts and history and class investigative activities. Lessons also include visual learning activities and built in learning checks for review. During Year 1 teachers at School D used this program as their main science curriculum, although they supplemented regularly with teacher-made, commercial, and online resources. To reiterate, this school (and program) were not included in Year 2 of the study.

Teachers participating in the study were instructed to cover topics as required by their respective state and districts, so there was variability in which science topics were covered by each grade in each school. The control curricula, including resources available, are described in more detail in Appendix E.

Comparisons between Science Fusion and Control Program Content, Coverage and Practices

As a result of state and district scope and sequence guidelines prescribing what science content needed to be covered, treatment and control classes within schools generally taught similar content. While some topics were presented in a different sequence depending on the program used, study teachers within each school generally noted that by the end of the year, the content covered in both treatment and control classrooms was similar.

As shown in Table 7, comparison on the percent of science topic areas completed during the second study year showed that while treatment and control teachers covered approximately the same content areas, the amount covered varied. Two significant differences were observed across the study classes, with treatment teachers covering Science & Technology and Cells & Heredity significantly more than control teachers, $p<.05$. Furthermore, although differences were not significant, treatment teachers tended to provide more coverage of Motion, Forces & Energy. Given observed differences in coverage, only topic areas that matched treatment and control classes were assessed via the Developed Science Test, thus controlling for differences in content coverage.

Table 7. Percent Coverage of Science Topic Areas: Year 2

	Control	Science Fusion	# of classes covering topic area
Cells & Heredity	39.9%	68.4%	18
Diversity of Living things	36.1%	30.1%	14
Human Body	55.6%	55.6%	9
Ecology and Environment	21.1%	28.6%	12
Dynamic Earth	61.1%	72.2%	5
Water & Atmosphere	81.7%	82.3%	7
Space Science	86.7%	93.3%	4
Matter and Energy	46.3%	49.0%	9
Motion, Forces and Energy	64.3%	92.9%	7
Sound and Light	30.0%	30.0%	6
Science and Technology	45.8%	100.0%	8

It should be noted that not all topic areas were covered over the school year. On average, teachers covered 3-5 topic areas. Specific topics covered varied by grade level and school, with coverage based on district/state curriculum maps. As shown in Table 7, the topic areas taught in most classes included Cells & Heredity, Diversity of Living Things, Ecology and the Environment, Matter & Energy, and Human Body.

With respect to the textbooks and the pedagogical approaches employed by the various science curricula, there were notable differences between control and Science Fusion programs. As previously noted, all schools, except Schools C used traditional chapter-based, teacher delivered programs as their main control curricula. Teachers at Schools C tended to supplement their program on a regular basis. In comparison to the Science Fusion program, the control group did not incorporate technology to the degree that the Science Fusion program did. While three of the four of the control programs did have some digital resources, the teachers did not incorporate them into the main lesson, nor were they utilized by students for the practice, reference or differentiation activities they were designed for. The Science Fusion program also allowed students to fully access their interactive lessons and texts from home. In contrast, the control programs with digital resources allowed for general online access to lessons and chapters but were not interactive and it was not reported that students or teachers in control classrooms utilized these limited features.

When the pedagogy of the Science Fusion program is compared to the control programs there is a notable difference in the primary philosophy behind each program. Specifically, Science Fusion delivers content and lessons organized by large, overarching concepts that span the gap between science concepts and real world applications; specific skills and activities support the larger concept. Furthermore the Science Fusion program lessons are driven by a blend of print, digital and hands-on on resources that emphasize 21st century skills. While basal control program 1 is also organized by large overarching concepts, the program does not provide the blend of inquiry that the Science Fusion program offers. The Science Fusion program encourages students and teachers to begin each lesson by asking questions, whereas the control programs focus on a more traditional approach where students read text passages. The Write-In Student Edition used in the Science Fusion program also encourages students to be actively engaged in the material as opposed to passively reading as in the control programs. While the skills and content is very similar between programs, the inherent differences stem from the way the Enduring Understanding model asks students to look at the big picture first whereas the traditional pedagogy of the control programs attempt to move students from general concepts to a larger understanding. This, along with the blending of technology and hands-on activities that is built into the Science Fusion program, are the greatest differences between this program and the control programs.

In terms of a typical lesson schedule, lessons in both control and treatment class were relatively consistent with a few exceptions. Lessons usually started with a bell ringer or warm up activity and a homework check. Next teachers would introduce and begin the new lesson. Lessons included some lecture, discussion and reading. Some teachers also incorporated technology (e.g., videos). Depending on where they were in the chapter this was followed by a lab. Teachers would then assign book work or worksheet activities to be completed independently or in groups. Depending on the length of the class, students might have time to finish the majority of the assignment in class; if not it was generally sent as homework.

In terms of specific instructional activities, there were some significant differences observed. For example, treatment teachers reported a significantly stronger emphasis in science vocabulary and science review over the course of the school year. Furthermore, treatment classes tended to have higher levels of student engagement and students engaged in more lab activities than control classes over the school year.

These were the only notable differences observed across schools in terms of science instruction. Appendix E contains a crosswalk between Science Fusion content and the control programs' content. As is clearly evident, the Science Fusion program is more comprehensive, offering lessons on a wide variety of science topic areas. Nevertheless, and as previously discussed, while variations did exist in coverage for a number of topics, within schools and grade levels treatment and control teachers tended to cover similar areas.

Fidelity of Implementation

Three levels of implementation (low, moderate, and high) were assigned for teachers' implementation of key Science Fusion program components as noted in the implementation guidelines (see Appendix C). Triangulation of the available information[16] showed that one teacher did not typically follow the implementation guidelines which outlined the key components of the Science Fusion program. In particular, this teacher did not engage in the embedded lesson activities (e.g., Activities & Discussion, Think Outside the Book, and Lesson Quiz) with the requested frequency. Another teacher was at a moderate level and all remaining teachers were categorized as high implementers.

Appendix F provides a more detailed table describing the extent to which teachers utilized the various Science Fusion program components. Of note is that while most teachers did well in accessing and blending digital and print materials, some teachers rarely if ever used the digital resources due to a lack of suitable technology infrastructure at their school or because they did not feel comfortable using. For more information on how teachers implemented the Science Fusion program in their classrooms, see Appendix D: Case Studies.

Table 8. Level of Science Fusion Implementation

Level of SF Implementation	Completion of Key Program Components
Year 1	
High	80% or higher consistent completion of Science Fusion components= **11 classes**
Moderate	70%-79% consistent completion of Science Fusion components = **8 classes**
Low	Less than 70% of goals met = **8 classes**
Year 2	
High	80% or higher consistent completion of Science Fusion components= **7 classes**
Moderate	70%-79% consistent completion of Science Fusion components **= 5 classes**
Low	Less than 70% of goals met = **2 classes**

Approximately 86% of classrooms were exposed to the key Science Fusion program components with a moderate to high level of fidelity.

No evidence of contamination was observed between teachers or in classrooms. That is, teachers did not use any components of the Science Fusion program with control students. However, there was some movement of students from treatment to control classes (or vice versa) over the two study years. As previously noted, these students were excluded from the all program effect analyses that are subsequently reported.

[16] Information was analyzed from teacher logs, class observations, and exit interviews.

It should be noted that the potential for contamination was given careful consideration when determining the level of random assignment. Through years of research experience, PRES researchers have found that the benefits of random assignment at the teacher/classroom level (hence, controlling for school and teacher level factors) with careful monitoring of possible contamination, outweighs the risk of contamination. Procedures used to eliminate the threat of contamination included an in-depth study orientation with teachers, site visits made to both treatment and control classrooms to observe what was occurring in classrooms, and monthly teacher logs that monitored practices and materials used across both treatment and control classrooms.

Results

This section is organized by the key questions from the RCT and reviews major findings first, followed by a more detailed presentation of results.

Major Findings

Does science ability improve as a result of participation in the Science Fusion program?

Results showed significant growth in science performance over the course of both school years as measured by the national, standardized ITBS Science test and a developed science assessment aligned to the content covered during the school year as well as national and state standards. Science Fusion students grew by 14 percentiles on the ITBS Science test over the course of the two year study. In addition, while significant growth was observed during each study year on the Developed Science Test, growth during Year 2 was larger (14 points) than Year 1 (11 points). When tests for each ITBS content area were examined separately, Science Fusion students' showed significant improvement in Life Science, Scientific Inquiry and Earth Science. Furthermore, marginally significant growth was observed in Physical Science.

In addition, Science Fusion students experienced significant learning gains as measured by the science vocabulary and science application/reasoning items of the Developed Science Test, with larger gains observed during Year 2 as compared to Year 1. Taken together, these findings suggest a cumulative gain of the Science Fusion program on science performance, with higher levels of growth observed following two years of exposure to the program as compared to one year of exposure.

Do changes in science performance among Science Fusion students vary by different types of students (e.g., grade, gender, science level, economically disadvantaged status) and levels of implementation?

In contrast to Year 1 findings, results showed that Science Fusion students in all subpopulations examined showed significant learning gains on the Developed Science Test and ITBS, with one exception. High level students showed a significant decline on the Developed Science Test. In sum, during Year 2, males and females, 7^{th} and 8^{th} graders, students receiving free/reduced lunch those not, and students at various ability levels demonstrated significant learning gains in science.

Analysis by implementation fidelity showed that students whose teachers used the Science Fusion program with moderate fidelity showed the lowest gains as compared to teachers using the program with low and high fidelity. This was consistent across both the ITBS and Developed Science Test. This differs from Year 1 findings in which no significant relationship was observed between implementation fidelity and performance gains.

Does using Science Fusion result in increased student achievement as compared to other types of science programs?

Analyses were conducted on two samples: 1) all students participating in the second year of the study, including new students enrolled in participating study classes, and 2) students who participated in both study years and remained in the same study condition throughout. Results showed significant differences between students who used the Science Fusion program and students using other science programs as measured by the Developed Science Test (DST), after controlling for pretest differences. Specifically, Science Fusion students participating in both study years and those participating in Year 2 of the study showed significantly higher performance levels on the DST overall score as compared to students using other science programs. In addition, Science Fusion students outperformed control students on the fill-in-the-blank items designed to measure science vocabulary over both study years. As well, among students who participated in Year 2 of the study, marginally significant differences were observed on DST items measuring science application and reasoning. Given the lack of significant differences observed in Year 1 (only performance on the DST vocabulary items was significant), results suggest stronger effects of the Science Fusion program following two years of study participation.

On the ITBS Science test, although a similar trend was evident with Science Fusion students showing higher test scores than control students after controlling for pretest differences, no significant differences were observed. In summary, results from the second year of the study indicate that Science Fusion students outperformed control students as measured by the test designed to measure specific content areas that were covered over the course of the school year. It is also noteworthy that the effect sizes were moderate, with a range of .48 to .64. Indeed, all effect sizes obtained exceeded the threshold for educational significance (.25) which means that these findings are meaningful in terms of impacting a students' educational experience.

Do effects of Science Fusion on student science performance vary as a function of different student characteristics and control programs?

Results among subgroups of Year 2 participants (i.e., grade, gender, free/reduced lunch, and science level) showed that there were no significant subgroup effects. This means that there was no difference between treatment and control students within subgroups. It should be noted that the lack of significant differences may be due to the limited number of students within subgroups. For instance, in Year 1 when the sample size was larger, results showed that low-performing students who used the Science Fusion program demonstrated accelerated learning gains compared to control students on the ITBS and White students who used Science Fusion had higher test scores at post-testing on the Developed Science Test than White control students, after controlling for pretest differences.

Does participation in Science Fusion result in other positive outcomes (e.g., positive attitudes towards science, etc.)?

Affective positive outcomes were reported by both student and teacher users of the Science Fusion program. Science Fusion students reported more positive attitudes than control students relating to their science ability, enjoyment of science, and beliefs in the importance of science. More than control students and teachers, Science Fusion students and teachers agreed that the program positively impacted students' academic skills, especially problem-solving skills, scientific inquiry, and science-related math, reading, and writing ability. Science Fusion teachers also reported that their students were more interested and engaged in learning science than control teachers.

Teachers and students agreed that the Science Fusion program helped students make connections to the real-world more than the control program. The Science Fusion program was also reported to better prepare students to do well in high school, do well on science tests, and do well in future science courses than the control program.

Anecdotal information revealed that Science Fusion teachers were more prepared to give quality lessons, engage students, and provide differentiated instruction because of the Science Fusion program. Compared to control teachers, Science Fusion teachers reported that their science program helped them to minimize lesson preparation time and provided them with good ideas for activities.

What do users of the Science Fusion program think about the programs?

Users of the Science Fusion program were generally very positive about the program and all of its components. Teachers and students reported that the Science Fusion program was better than the previous science program and that they would like to use it again. Teachers gave high ratings to the program design and ease of use, commenting that it is an overall excellent system. Teachers were especially positive about the Teacher's Edition, student online textbook, and videos. Teachers and students agreed that the write-in student worktext is a valuable learning tool. The provided labs were well received by both students and teachers, with teachers noting the usefulness of the virtual labs. Teachers also reported that overall, using the Science Fusion technology was exciting and engaging for students and a good teaching experience for them.

Detailed Findings

Does science ability improve as a result of participation in the Science Fusion program?

To determine whether students who used Science Fusion showed significant learning gains over the course both years of the study, analysis on outcomes were conducted via repeated measures analyses. Science Fusion students participating in both years were examined to determine if gains from the first year (2011-12) were similar or different as gains observed in the second year (2012-13). Similar to results obtained in year 1, results showed significant growth in science performance on both the ITBS and Developed Science assessments during year 2, $p<.05$. In particular, results showed that on the ITBS overall scale score, students had similar gains during each of the study years (linear trend was significant but not the quadratic trend) with an overall gain of 14 percentiles. On the Developed Science Test, separate analyses were run for each school year since data from this assessment are not vertically scaled and the tests were unique each study year given that they were aligned to content taught in the science classes. As shown in Figure 2, while both gains were significant, gains in the second year (14 points) were larger than those observed in the first year (11 points).

It should be noted that the ITBS measures a variety of science content areas (Physical, Earth, Life, and Scientific Inquiry), which rarely were all covered by study classes. As a result, the ITBS overall scale score is not as sensitive as the Developed Science test, which was aligned to the content covered by treatment and control study classes. Taking this into consideration, results indicate that across all grade levels students showed significant growth from pre to post-testings on both assessments and time periods.

Figure 1. Pre- and Post ITBS Science Performance of Science Fusion Students

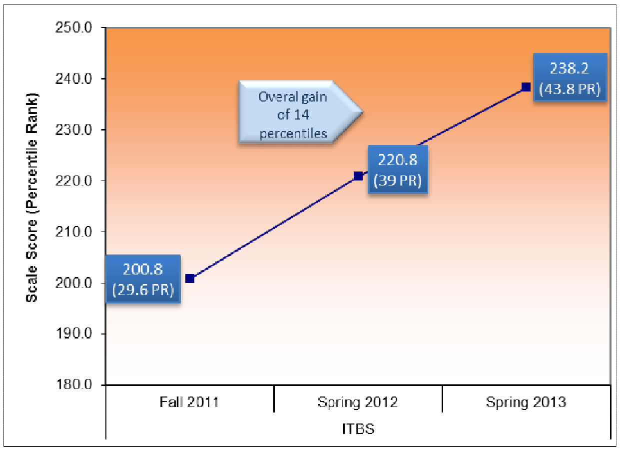
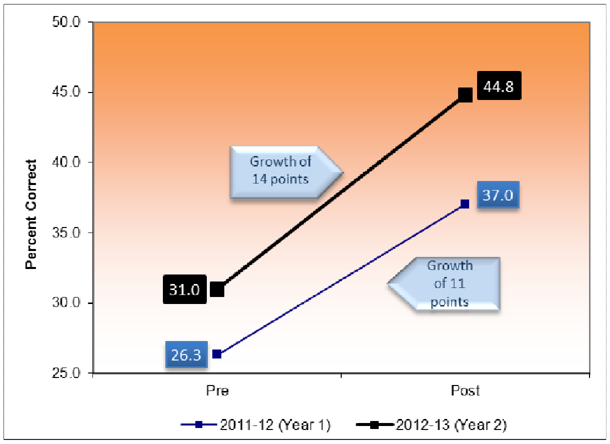


Figure 2. Pre- and Post Developed Science Test Performance of Science Fusion Students



Students who used Science Fusion showed significant growth in science performance as measured by the national, standardized ITBS Science test and a developed science assessment constructed from national and state science standards/items, and aligned to the science content covered in study classes. Significant gains occurred during both study years among treatment students.

To better understand the relationship between ITBS student science performance and the Science Fusion program, analyses were conducted for each content area (Physical, Earth, Life, and Scientific Inquiry) separately. Of note, because percent correct (and not a scale score) was available for analyses and such scores are not vertically scaled, overall growth across both study years (Fall 2011 to Spring 2013) was examined via paired sample t-tests. As well, only Science Fusion students participating in both study years were included in these analyses.

Similar to the prior year, results showed that Science Fusion students made significant learning gains in Life Science and Scientific Inquiry, $p<.05$. However, unlike the prior study year, students showed a significant gain in Earth Science, $p<.05$, and a marginally significant gain in Physical Science as well, $p<.10$, see Figures 3-6. As shown, students showed the greatest gains in the areas of Scientific Inquiry (15 points), Earth Science (13 points) and Life Science (9 points).

Figure 3. Pre- and Post ITBS Scientific Inquiry Performance of Science Fusion Students

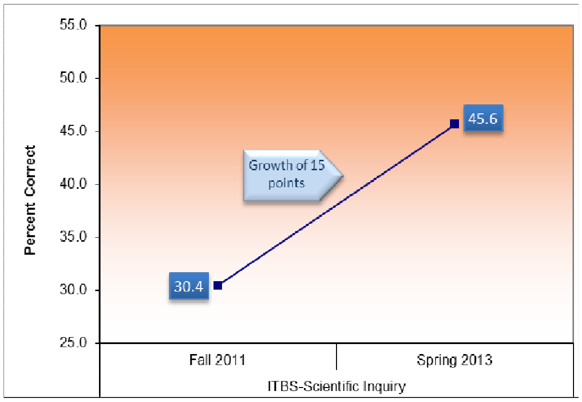


Figure 4. Pre- and Post ITBS Earth Science Performance of Science Fusion Students

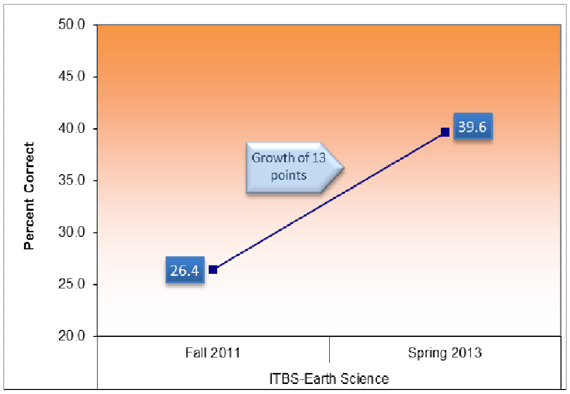


Figure 5. Pre- and Post ITBS Life Science Performance of Science Fusion Students

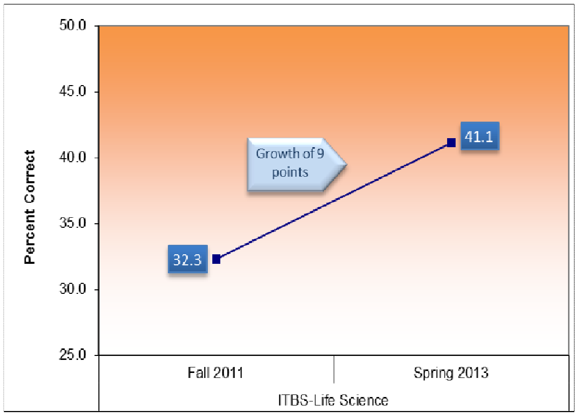
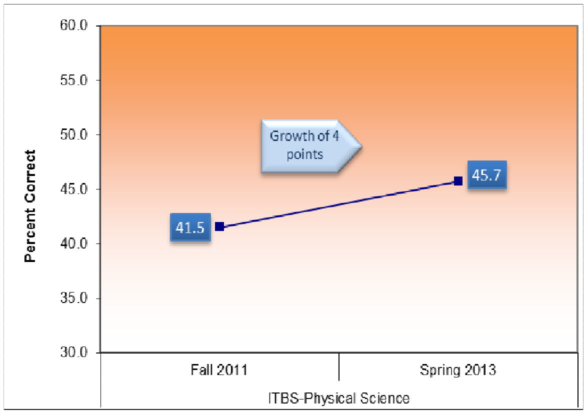


Figure 6. Pre- and Post ITBS Physical Science Performance of Science Fusion Students



When ITBS Science tests for each content area were examined separately, results showed significant growth from pretesting to post-testing (Spring 2013) in Scientific Inquiry (15 points), Earth Science (13 points), and Life Science (9 points) performance among Science Fusion students, and marginally significant growth on the Physical Science subtest (4 points).

Analyses were also conducted on the science vocabulary and science application / reasoning items of the Developed Science Test. As a reminder, analyses were conducted separately for each study year as students took different assessments at each study year (depending on science content areas taught) and scores are not vertically scaled. Among this sample of students who participated in both study years, results showed significant learning gains on vocabulary items in both study years, $p<.05$, see Figure 7. On the short answer items measuring science application, students did not show significant growth in year 1, but did show significant growth in year 2, $p<.05$, see Figure 8. Of note, in both subtests, students showed the greatest gains during the second year of the study.

Figure 7. Pre- and Post Developed Science Test – Vocabulary Performance of Science Fusion Students

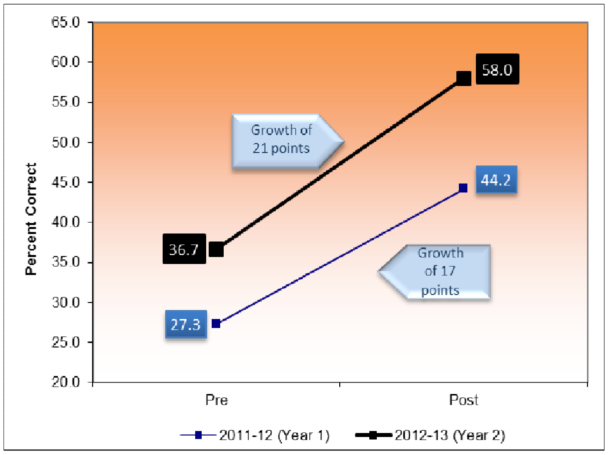
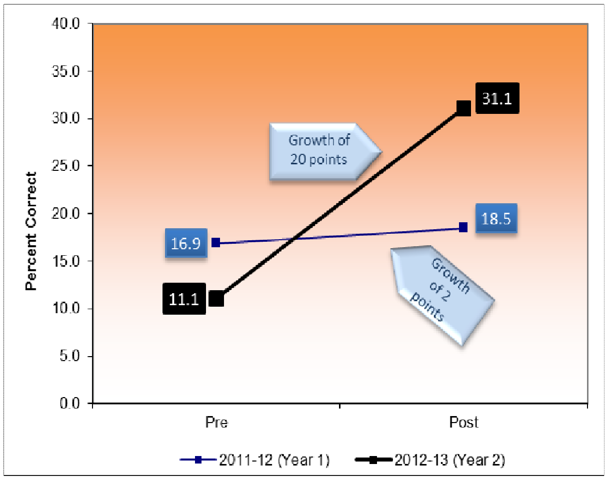


Figure 8. Pre- and Post Developed Science Test – Science Application Performance of Science Fusion Students



Science Fusion students showed more significant learning gains in 2012-13 as compared to 2011-12 as measured by items measuring science vocabulary and science application / reasoning.

Do changes in science performance among Science Fusion students vary by different types of students and levels of implementation?

In order to examine whether the Science Fusion program was associated with improvements among students of various subgroups, exploratory descriptive analyses were conducted. Only the performance of *treatment* students in specific student populations (i.e. students receiving free/reduced lunch and students not receiving aid, males and females, students of various science levels, and grade levels) was examined in these analyses. As well, due to the more limited sample size available for analyses for students participating in both years of the study[17], these analyses focused on students participating in Year 2 of the

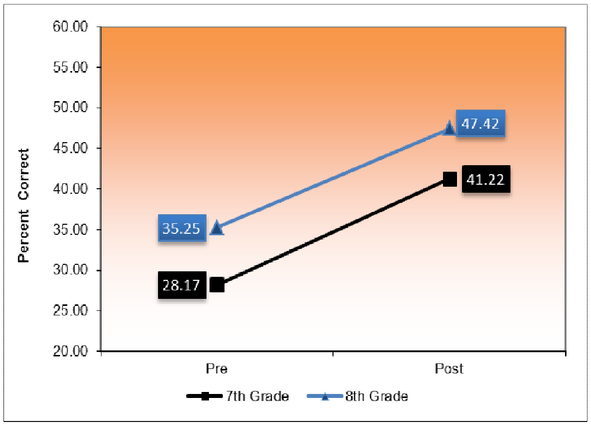
[17] More information on attrition observed in this study is provided in Technical Appendix A.

study (including new students in 2012-13). As a reminder, the Year 2 sample was both smaller and less diverse due to the loss of one large school during the second year of the study. As a result, these analyses do not include analysis by English Language Learner status, special education status, and ethnicity – there were insufficient numbers of students within these categories[18]. It should also be noted that the sample sizes in some of the subgroups are small and there are unequal sample sizes between those in the subpopulations and those not for a number of variables[19]. Therefore, with the caveat that these analyses are limited, this provides readers with preliminary, descriptive information on whether the program is associated with improvements among various subgroups. Figures 9 through 16 display the results for the various subgroups.

Results showed that students in all subgroups showed significant gains on both the Developed Science Test and ITBS. This is in contrast to last year's results in which subgroup growth findings on the ITBS were inconsistent with some groups showing gains and others decreases. In sum, males and females, 7th and 8th graders, students receiving free/reduced lunch those not, and students at various ability levels demonstrated significant learning gains in science.

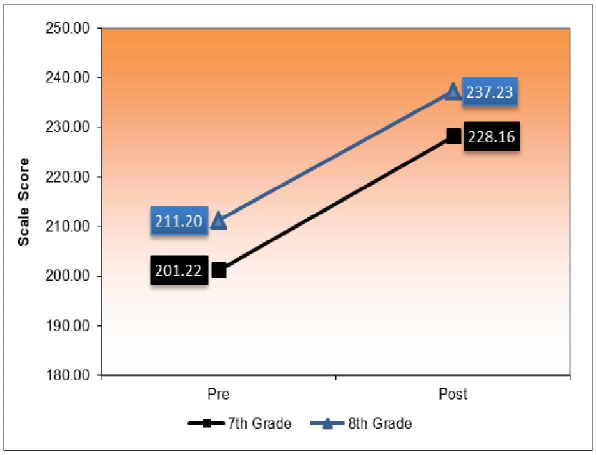
Science Fusion students in all subgroups (gender, free/reduced lunch status, grade level, and ability level) showed significant gains on the Developed Science Test and ITBS.

Figure 9. Science Fusion Students Performance Gains by Grade Level: Developed Science Test



Science Fusion students of all grade levels showed significant learning gains on the Developed Science Test.

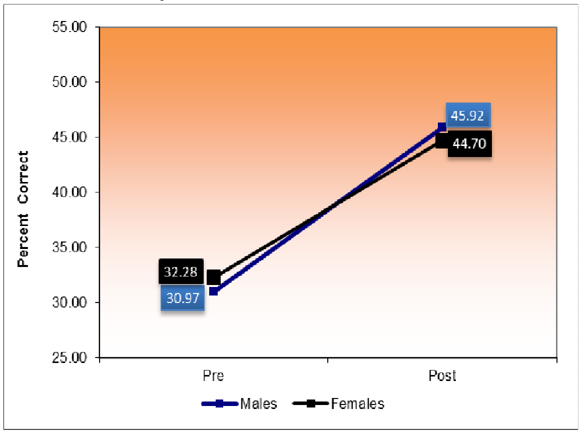
Figure 10. Science Fusion Students Performance Gains by Grade Level: ITBS



Both 7th and 8th grade Science Fusion students showed significant growth on the ITBS Science test.

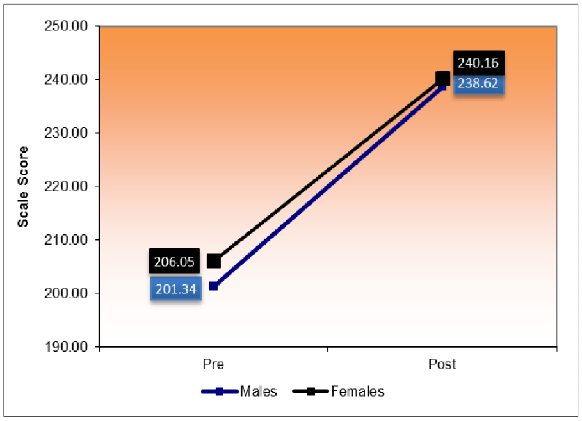
[18] The reader is referred to the Science Fusion Year 1 Final Report for results from these subgroups participating in Year 1 of the RCT.
[19] The reader is referred to the Technical Appendix A for detailed statistics.

Figure 11. Science Fusion Students Performance Gains by Gender: Developed Science Test



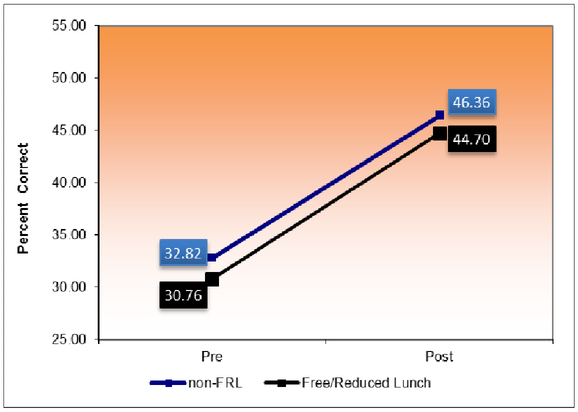
Females and males showed significant, similar learning gains on the Developed Science test.

Figure 12 Science Fusion Students Performance Gains by Gender: ITBS



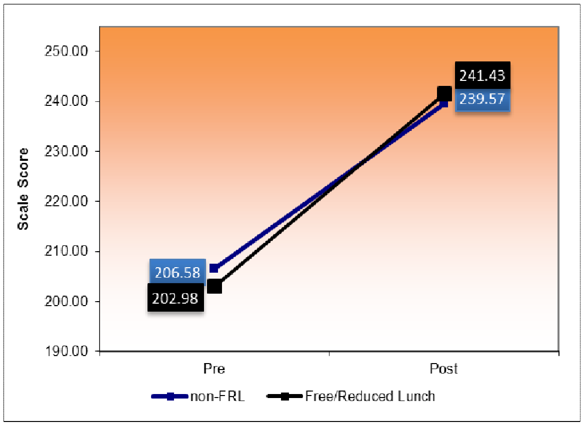
Similarly, both males and females showed significant gains on the ITBS Science test.

Figure 13. Science Fusion Students Performance Gains by Free/Reduced Lunch Status: Developed Science Test



Science Fusion students receiving free/reduced lunch and those not receiving this assistance showed significant improvement on the Developed Science test.

Figure 14. Science Fusion Students Performance Gains by Free/Reduced Lunch Status: ITBS

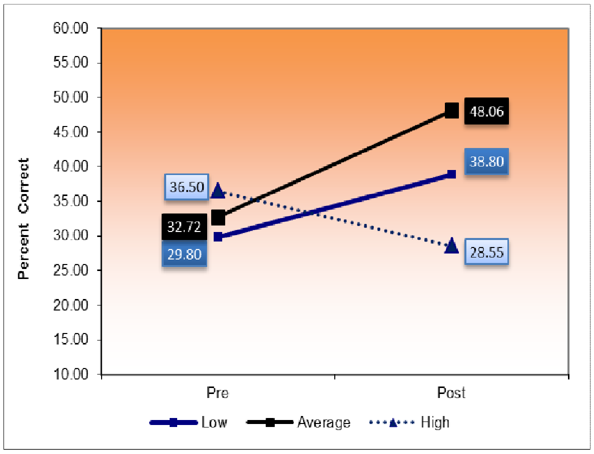


Similarly, students receiving free/reduced lunch and those not receiving this assistance demonstrated learning gains on the ITBS Science test.

SCIENCE LEVELS

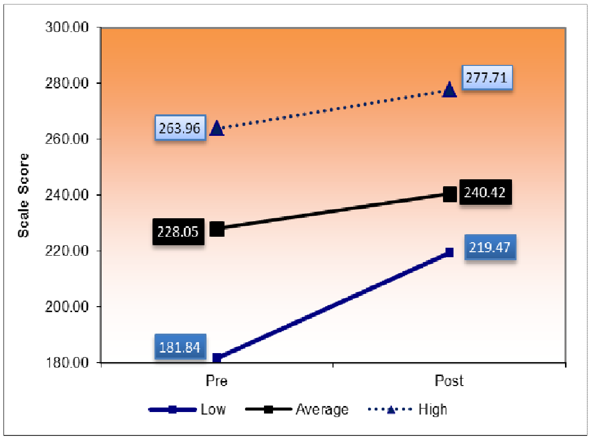
Performance results from the ITBS Science test administered during pre-testing were used to categorize students on initial science level, since it is a norm-referenced test. Students who were at or below the 33rd percentile were classified at a low science level, students who were at or above the 66th percentile were classified as high, and the remaining students were classified as average. Comparisons were made between the three identified science levels. Results showed that with one exception, students at all science levels demonstrated significant learning gains, see Figures 15-16. On the Developed Science Test, high level students showed a significant decline. However, all other students demonstrated significant learning gains.

Figure 15. Science Fusion Students Performance Gains by Science Level: Developed Science Test



Science Fusion students of low and average ability levels showed significant improvement over time on the Developed Science Test. However, high ability students demonstrated a significant decline.

Figure 16. Science Fusion Students Performance Gains by Science Level: ITBS



Science Fusion students at all ability levels showed significant improvement from pre to post-testing on the ITBS.

IMPLEMENTATION LEVELS

Exploratory analyses on the relationship between overall levels of Science Fusion implementation of key program components and student science performance were also conducted. These analyses provide preliminary information on whether low to high implementation fidelity of Science Fusion[20] components was associated with student performance. Note that sample sizes are uneven, with the majority of treatment teachers being high implementers.

Results showed significant relationships between overall Science Fusion implementation levels and improved performance on the ITBS and Developed Science assessments, $p<.05$. Specifically, results show that students whose teachers used the Science Fusion program with moderate fidelity showed the lowest gains as compared to teachers using the program with low and high fidelity, see Figures 17-

[20] See section on Fidelity of Implementation for how this categorization was determined.

18. This was consistent across both the ITBS and Developed Science Test. This differs from year 1 findings in which no significant relationship was observed between implementation fidelity and performance gains.

Figure 17. Science Fusion Students Performance Gains by Implementation Level: Developed Science Test

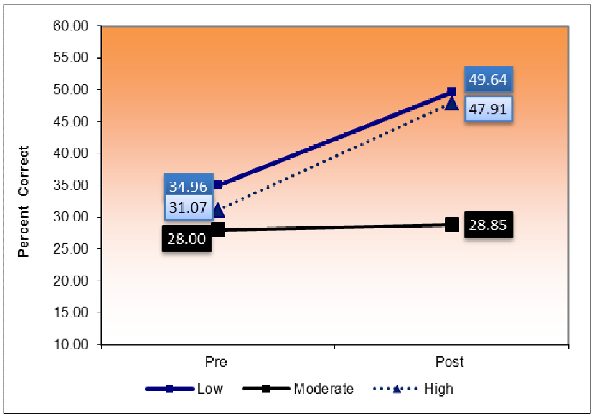
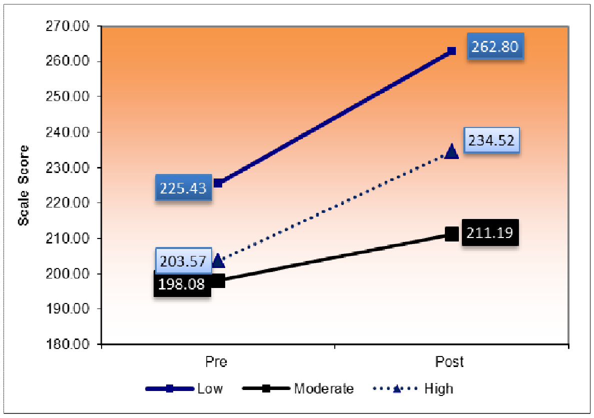


Figure 18. Science Fusion Students Performance Gains by Implementation Level: ITBS



Exploratory analyses showed that Science Fusion students whose teachers used the program with high and low fidelity demonstrated the greatest learning gains.

Does using the Science Fusion program result in increased student achievement as compared to other types of science programs?

Prior to discussing the results found, it is important to understand the differences and similarities of the Science Fusion program and control curricula and classes. This will assist the reader in interpreting the results and effect sizes[21], a measure of the importance of an intervention.

COMPARISON OF SCIENCE FUSION AND CONTROL CLASSES

As previously noted, control and treatment classes generally were exposed to the same content within schools. This is due to teachers following curriculum pacing guides that dictate what content to cover at each grade level. While coverage was fairly homogenous within schools, across all study schools there was variation in the extent to which control and treatment teachers covered specific topic areas. For instance, treatment teachers covered Science & Technology and Cells & Hereditary significantly more than control teachers. Given observed differences in coverage, only topic areas that *matched* treatment and control teachers covered during the school year were included in the Developed Science Test, thus controlling for differences in content coverage.

Differences in the pedagogy of the control programs used were also observed. The Year 2 control programs used were structured in a traditional unit/chapter/lesson organization. While the Science Fusion

[21] Effect size (ES) is commonly used as a measure of the magnitude of an effect of an intervention relative to a comparison group. It provides a measure of the relative position of one group to another. For example, with a moderate effect size of d=.5, we expect that about 69% of cases in Group 2 are above the mean of Group 1, whereas for a small effect of d=.2 this figure would be 58% and for a large effect of d=.8 this would be 79%.

program is also structured in units and lessons, content is further organized according to the program's pedagogy with units focused on Big Ideas and lessons focused on Essential Questions to facilitate an enduring understanding of science. It should also be noted that control program 1 also encouraged students and teachers to begin each lesson by asking questions, however, these questions did not function as an organizational tool as with Science Fusion. Other differences between the Science Fusion program and control curricula included the integration of digital and hands-on materials. While the control programs may have offered an opportunity for digital and lab activities, teachers reported that they did not have access to these materials and/or any digital or lab activities that were completed were from other commercial, online, or teacher-created resources.

While the aforementioned differences were noted, similarities between Science Fusion and the control curricula were also observed. Control programs 1, 2 and 3 all included engaging section openers that required students to access prior knowledge of the topic, cross curricular learning, and built in lab and assessment activities. Similar to Science Fusion, control program 1 also included information rich visuals that connect to the text and support student learning and vocabulary strategies.

There were also only a few observed differences between the groups in terms of how the lessons were structured or delivered. In the Fall, treatment teachers reported having a more positive class environment and engaging in intervention activities to a greater extent than control teachers. Over the course of the school year, treatment teachers reported a significantly stronger emphasis in science vocabulary skills and science review. Furthermore, treatment classes tended to have higher levels of student engagement and students engaged in more lab activities than control classes over the school year. Other than these differences, the instructional sequence and practices employed were comparable across treatment and control classes, and from teacher to teacher. Generally lessons included bell work and a review of the previous day's homework or prior lesson. Depending on the day, this was followed with whole group instruction of the new concept or a lab activity. If time remained, the last part of the class typically involved independent practice.

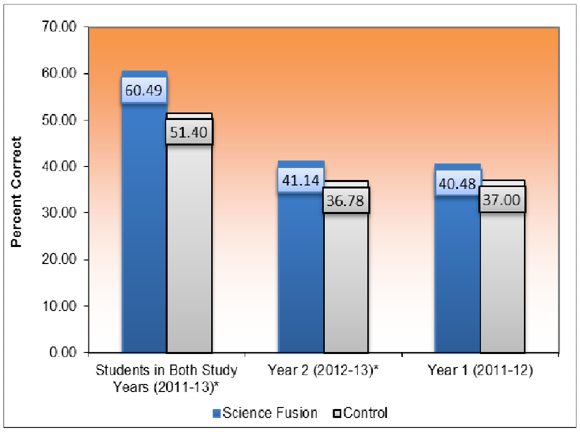
In summary, Science Fusion and control classrooms, with the exception of the program-based activities and coverage of certain topic areas, were similar to one another in terms of structure and science concepts. Given this information and the fact that the duration of the study and exposure to the program occurred during two school years, small to medium effect sizes, if any, were expected. Expanding the study over the course of two school years allowed for year 2 teachers to become better accustomed to the Science Fusion program and therefore, to be more familiar with the program, thereby reducing the learning curve experienced by teachers using a new curriculum during the first year. Thus, it was hoped that treatment and control teachers would be much more comparable in terms of experience and comfort in using their assigned curriculum.

RESULTS

Given the baseline differences observed among students participating in year 2 of the study and those participating in both study years, multilevel modeling was conducted to examine whether there were significant differences in student post-test performance after controlling for pretest performance. Such analyses directly control for pretest differences along with other covariates in which groups differed[22], thus equating groups. Analyses were conducted on two samples: 1) all students participating in the second year of the study, including new students enrolled in the participating study classes (Year 2 in graphs), and 2) students who participated in both study years and remained in the same study condition throughout. In addition, results from the first year of the study which included a larger student sample with the inclusion of School D are also presented (Year 1 in graphs).

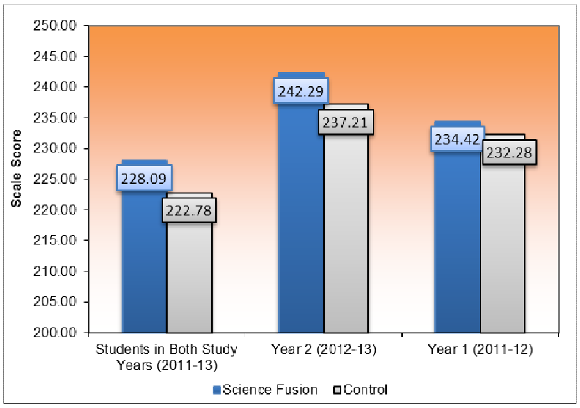
Results showed significant differences between students who used the Science Fusion program and students using other science programs as measured by the Developed Science Test, $p<.05$, after controlling for pretest differences. Specifically, Science Fusion students participating in both study years and those participating in Year 2 of the study showed significantly higher post-test scores as compared to control students. Such statistically significant differences were not observed in Year 1. While a similar trend was observed on the ITBS Science test (overall scale score), no significant differences were observed after controlling for pretest differences, see Figures 19-20. In summary, results from the second year of the study indicate that Science Fusion students outperformed control students as measured by the test designed to measure specific content areas that were covered over the course of the school year.

Figure 19. Developed Science Test Posttest Performance of Science Fusion and Control Students



*$p<.05$

Figure 20. ITBS Posttest Performance of Science Fusion and Control Students



Results yielded significant differences as measured by the Developed Science Test, with Science Fusion students outperforming students using other science programs. Significant differences were not observed on the ITBS Science test, however.

[22] Covariates for two level models include pretest, teacher engagement in intervention activities, class environment, and school.

When tests for each ITBS science content area were examined separately, results also showed no significant differences, $p>.05$, see Figures 21-24. As shown, while Science Fusion students on average showed higher post-test scores (after controlling for pretest scores) on Scientific Inquiry and Earth Science over all study years, such differences were not statistically significant. In contrast, control students tended to show higher Life Science posttest scores than Science Fusion students in both study years; however, these were not significant.

Figure 21. ITBS Scientific Inquiry Posttest Performance of Science Fusion and Control Students

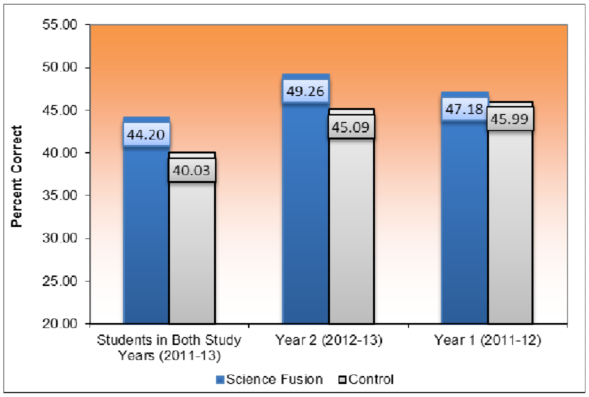


Figure 22. ITBS Earth Science Posttest Performance of Science Fusion and Control Students

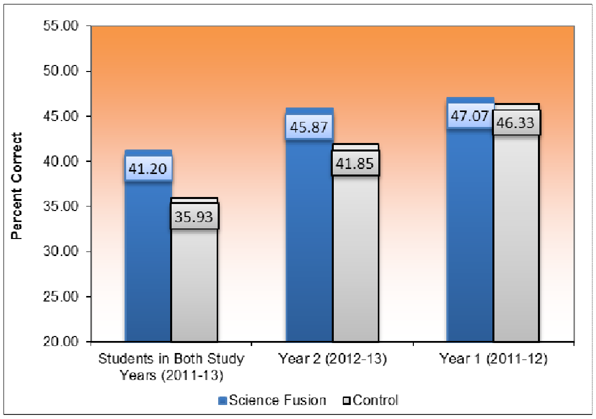


Figure 23. ITBS Physical Science Posttest Performance of Science Fusion and Control Students

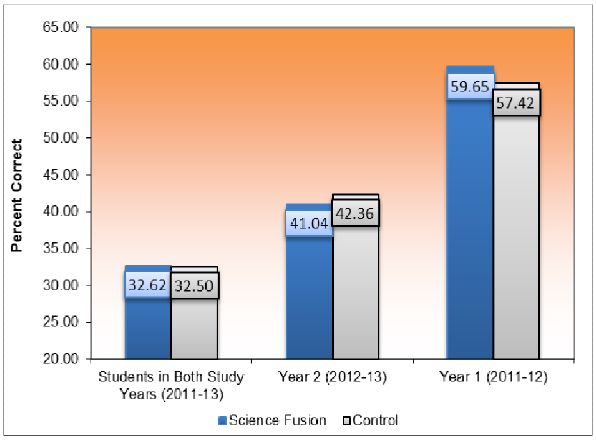
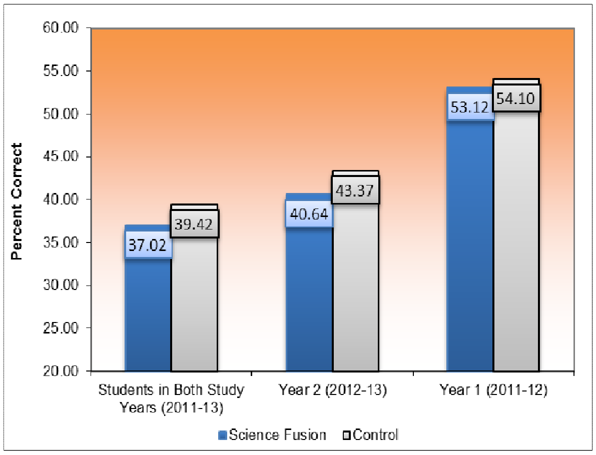


Figure 24. ITBS Life Science Posttest Performance of Science Fusion and Control Students



Science Fusion students showed similar levels of performance on the ITBS Scientific Inquiry, Life Science, Earth Science and Physical Science subtests as students using other science programs.

In order to explore if there were differences among treatment and control students with respect to their performance on specific scientific knowledge areas, further analyses of the Developed Science test were conducted. As a reminder, this assessment included fill in the blank items primarily measuring science vocabulary and short answer items primarily measuring scientific reasoning skills and application of science concepts.

Results controlling for pretest differences showed significant differences such that Science Fusion students outperformed control students on the vocabulary posttest items over both study years, $p<.05$. As well, among students who participated in Year 2 of the study, significant differences were also observed on items measuring science application and reasoning, $p<.10$. In sum, results suggest stronger effects of the Science Fusion program following two years of study participation.

Figure 25. Developed Science Posttest Vocabulary Performance of Science Fusion and Control Students

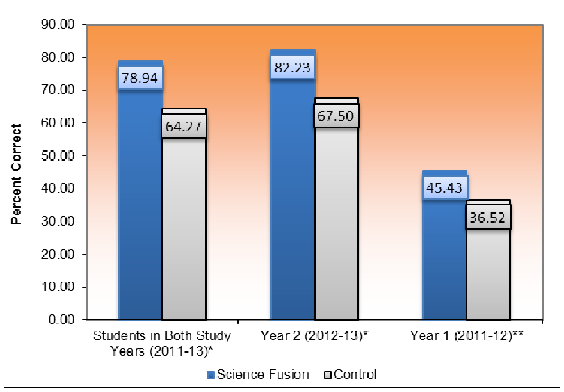
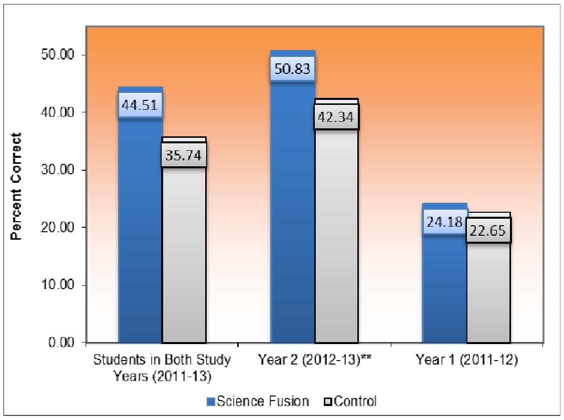


Figure 26. Developed Science Posttest Science Reasoning/Application Performance of Science Fusion and Control Students



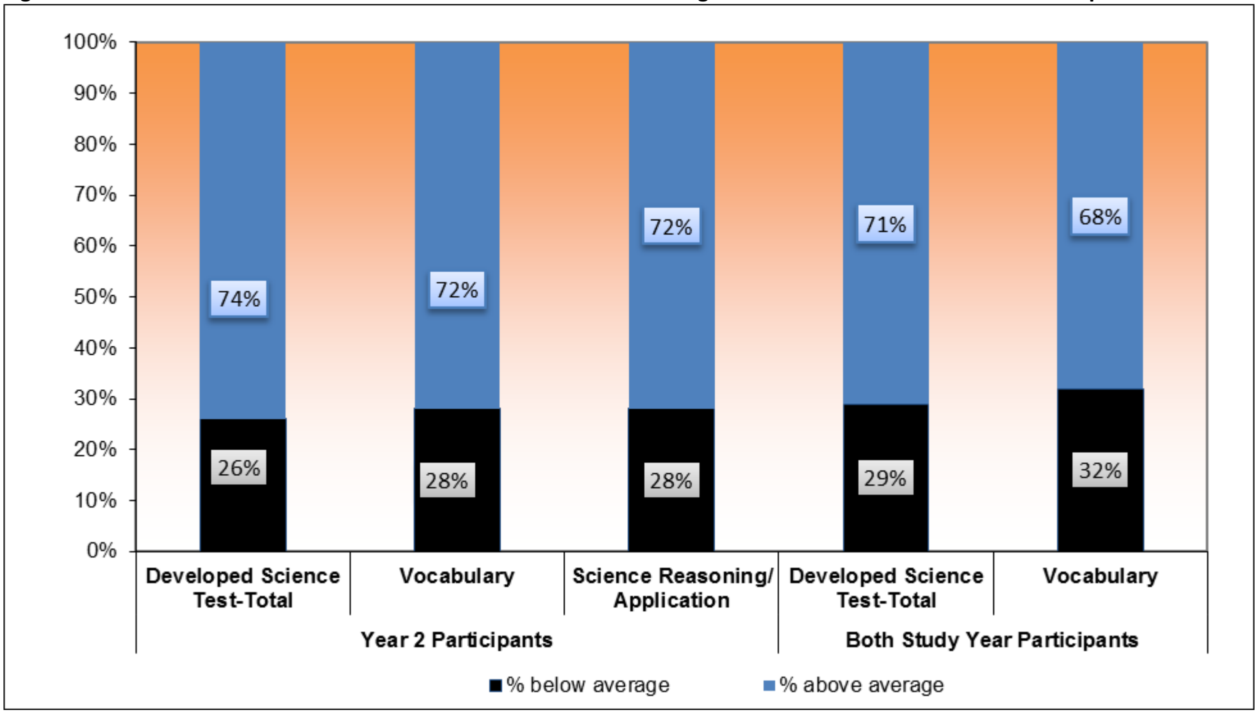
Students using the Science Fusion program performed better on the fill in the blank test items measuring science vocabulary and short answer items measuring science reasoning/ application as compared to students using other science programs, after controlling for pretest differences.

EFFECT SIZES

Effect size is a commonly used measure of the importance of the effect of an intervention (in this case, Science Fusion). Since the Developed Science Test showed significant differences, effect sizes were calculated for the obtained significant differences. The overall effect size for the Developed Science test is .64 among students participating in both years of the study and .57 for students participating in Year 2 of the study (which included both year study participants as well as new students). The effect sizes for the fill in the blank test items measuring science vocabulary are .57 and .55 respectively. Finally, the effect size for Year 2 students on short answer items measuring science application and reasoning is .48. All of these effect sizes exceed the threshold (.25) for educational significance. This means that these findings are meaningful in terms of impacting a students' educational experience.

Effect sizes can be translated to the percent of treatment students that can be expected to be *above* the average of the control group (see blue part of bar in Figure 27). As shown, on the Developed Science Test and its subtest areas, students using Science Fusion are more likely to have scored above the average of control students.

Figure 27. Percent of Science Fusion Students Above and Below Average Relative to Control Students: Developed Science Test



Results on the overall Developed Science Test, and the vocabulary and science reasoning/application subtests showed effect sizes that exceeded the threshold (0.25) for educational significance. Thus, learning gains observed can be interpreted as educationally meaningful.

Do effects on student science performance between Science Fusion and control students differ across types of students?

To examine if there were differences in performance between different subgroups of Science Fusion students and students using other science programs, subgroup effects were analyzed. Specifically, differences between Science Fusion and control students in the following subgroups were examined: grade, gender, free/reduced lunch status, and science ability level. In order to bolster the number of students within subgroups, Year 2 study participants were examined in these analyses[23]. As previously noted, multilevel models account for statistical issues that can affect the validity of the results. Furthermore, it is important to view these analyses as exploratory[24].

Unlike Year 1 findings, results among subgroups of Year 2 participants (i.e., grade, gender, free/reduced lunch, and science level) showed that there were no significant subgroup effects. This means that there was no difference between treatment and control students within these subgroups—subpopulations of Science Fusion students and those using other science programs performed similarly at post-testing after controlling for pretest performance. It should be noted that the lack of significant differences may be due to the limited number of students within subgroups. In Year 1, results showed that low-performing students who used the Science Fusion program demonstrated accelerated learning gains compared to control students on the ITBS , $p<.05$, and White students who used Science Fusion had higher test scores at post-testing on the Developed Science Test than White control students, after controlling for pretest differences, $p<.10$.

Analysis on Year 2 participants conducted to examine if there were differences between Science Fusion students and control students in specific subgroups (i.e., grade, gender, free/reduced lunch, and science level) showed that both Science Fusion students and control students within subgroups performed similarly at post-testing after controlling for pretest performance.

[23] The sample of students participating in both study years was smaller and would limit subgroup analyses further.
[24] Detailed information on why this is exploratory and non-causal and statistics, as well as these results are presented in Technical Appendix A.

Does participation in Science Fusion result in other positive student outcomes (e.g., positive attitudes towards science and so forth)?

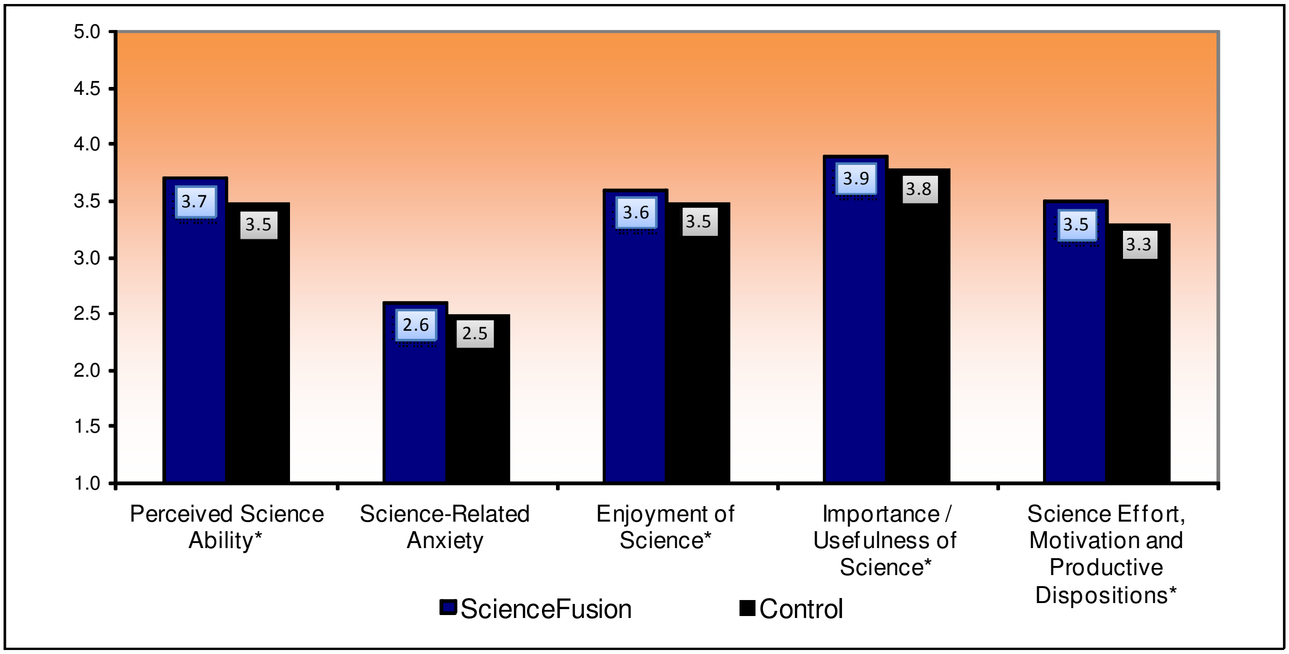
While the primary focus of the Science Fusion program is to improve students' science understanding and skills, the program incorporates a number of components that may have an effect on other important aspects of science education, including affective attitudes. Measures were included in the RCT to explore whether use of the Science Fusion program was associated with changes in student attitudes towards science as well as changes in teacher practices and attitudes. The following presents data collected during Year 2 of the study[25].

STUDENT ATTITUDES TOWARD SCIENCE

Comparison of data collected on science-related student attitudes (see Figure 28) showed a significant effect for perceived science ability, $p<.05$, enjoyment of science, $p<.05$, importance/usefulness of science, $p<.05$, and science effort/motivation, $p<.05$. Specifically, Science Fusion students were more likely to agree that they were good at science, enjoyed science, found science to be useful and that they are motivated to learn science, as measured by the Spring 2013 student survey. No significant differences were observed for science-related anxiety.

Results showed that students using Science Fusion were more likely to report a high science ability, they enjoyed science, they believed that science was important, and that they were motivated to learn science as compared to students using other science programs.

Figure 28. Student Science-Related Attitudes by Group



Higher scores indicate more positive attitudes. Based on a scale of 1-5. * $p<.05$.

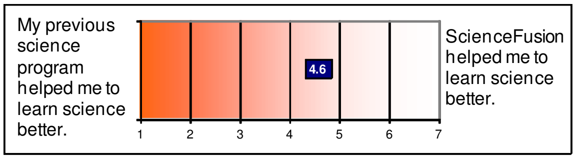
[25] For Year 1 results, the reader is referred to the Year 1 Science Fusion Final Report.

In the following sections, more detailed information is presented on how the science program impacted students in terms of: 1) science learning, 2) engagement and motivation, 3) application of science, and 4) preparation for future tests and science courses.

PERCEIVED IMPACT ON STUDENT ACADEMIC SKILLS

When asked to compare their prior science program to the Science Fusion program, students felt that the Science Fusion program was somewhat more effective in helping them learn science (see Figure 29). Qualitative feedback from Science Fusion students revealed that the design of the student book (a write-in edition) was especially useful in helping them easily organize notes. They also reported that science concepts were clearly explained and the visual aids assisted with their understanding.

Figure 29. Student Perceptions of the Degree to Which the Science Fusion Program Helped Them to Learn Science



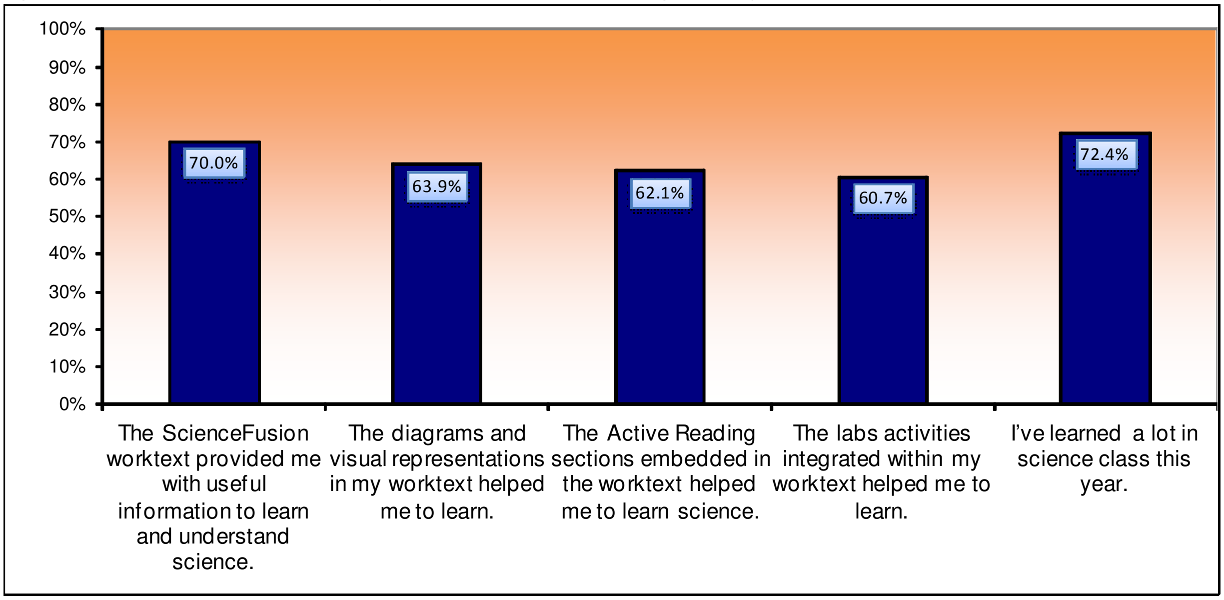
- *"It [ScienceFusion] helped me learn science better because it was easier to understand and organized." – Student, OH School*
- *"The book [worktext] had better explaining graphs, pictures, and paragraphs." – Student, OH School*
- *"Science Fusion made science interesting, describing what big scientific words mean in an easy way, showed pictures of science in nature and human activity (made connections)." – Student , OH School*

As shown in Figure 30, most students (72%) reported that they learned a lot in class this year with the Science Fusion program. Furthermore a majority of students (70%) reported that the Science Fusion student book provided them with useful information to learn and understand science and that the diagrams and visual representations in the student book helped them learn (64%). Some students also indicated that the lab activities integrated with the student book helped them learn (61%) and that the Active Reading sections embedded in the student book helped them learn science (62%). This represents an increase from Year 1 in which only a little over half the students agreed that the Science Fusion worktext provided useful information to learn science.

In sum, students reported only moderate agreement with statements about the usefulness of the book and program components and only a slight preference for Science Fusion compared to the previous program. Nevertheless, the majority of students reported positive feedback about specific ways in which the program was engaging and helped them learn.

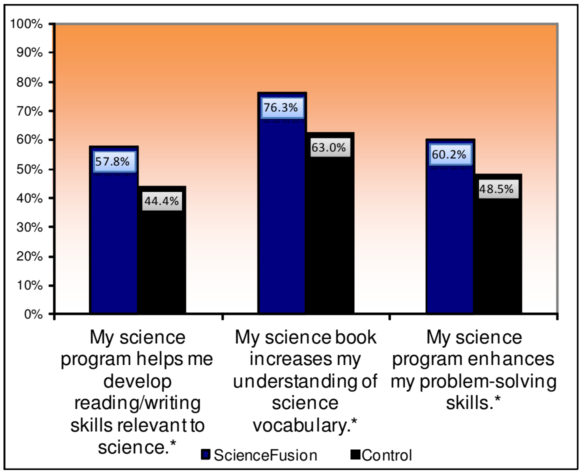
- *"The lab activities are more interactive and engaging." – Student, OH School*
- *"Science Fusion helped me learn better because it explained details in-depth." – Student, OH School*
- *"I like the review. I love the amount of reading embedded into the Science Fusion program. I also like the labs we complete." – Student, OH School*

Figure 30. Percent of Students Who Agreed the Science Fusion Program Helped Them Learn Science



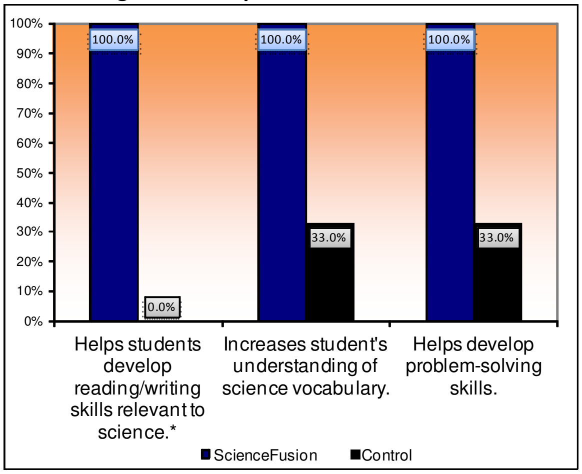
Comparing student perceptions on the degree to which their science program developed their academic skills (Figure 31), Science Fusion students were more likely to indicate that their book increased their understanding of science vocabulary (76%) than students in other science programs (63%), $p<.05$. As well, Science Fusion students perceived their science program's ability to develop reading/writing skills (58%) and problem solving skills (60%), as significantly greater than students in other science programs (44% and 49%), $p<.05$.

Figure 31. Percent of Students Who Agree That Their Science Program Developed Academic Skills



When teachers were asked about the degree to which their science program developed their students' academic skills, Science Fusion teachers were significantly more likely to report their program developed student reading and writing skills (100%) as compared to teachers of other science programs (0%), $p<.05$, (see Figure 32). Science Fusion teachers were also more likely to report that their program helped students' understanding of science vocabulary and problem-solving skills (100%) as compared to the teachers of other science programs (33%), but these differences were not statistically significant. Note, however, that small sample sizes (7 teachers in Year 2) reduce the likelihood of obtaining statistically significant differences even though differences may be large. Therefore, it's important to look at the pattern of results as well.

Figure 32. Percent of Teachers Who Agree That Their Science Program Developed Student Academic Skills

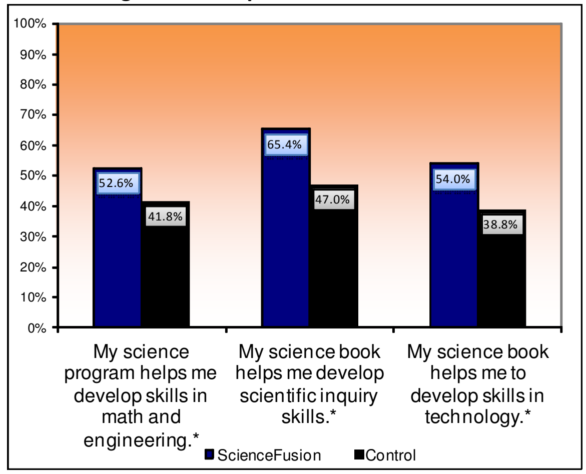


* $p<.05$

- *"They have made great gains in problem solving, vocabulary, and reading for content." – Teacher, D.C. School*
- *"They have the ability to actually pull information and skim the text easier and then write a great answer without quoting straight from the book." – Teacher, OH School*
- *"Most students really improved in inferring and reflecting on concepts of what they read." – Teacher, D.C. School*

Student perceptions on the degree to which their science program helped develop STEM skills (see Figure 33) were significantly different in treatment groups versus control groups, $p<.05$,. Over half of the students in the Science Fusion program reported that their science program helped them develop math and engineering skills (53%), science inquiry skills (65%), or skills in technology (54%), while less than half of the students in the control science program reported the same (42%, 47%, and 39% respectively).

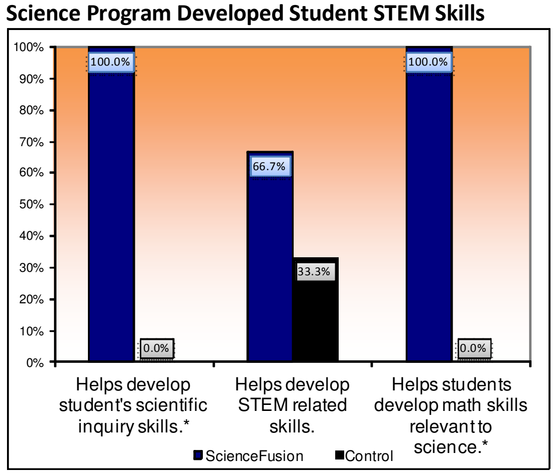
Figure 33. Percent of Students Who Agree That Their Science Program Developed STEM Skills



* $p<.05$

The majority of Science Fusion teachers also reported that their program had a positive effect on students' STEM and scientific inquiry skills (see Figure 34). Specifically, a significantly higher percentage of Science Fusion teachers agreed that their program helped develop student scientific inquiry skills (100% versus 0%), $p<.05$, and math skills relevant to science (100% versus 0%), $p<.05$. While not a significant difference, a higher percentage of Science Fusion teachers also reported that their program developed STEM related skills (67% versus 33%).

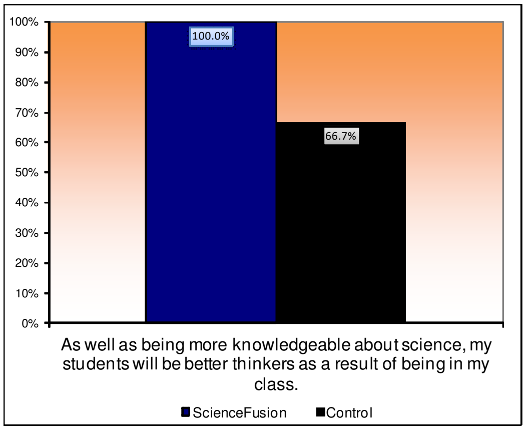
Figure 34. Percent of Teachers Who Agree That Their Science Program Developed Student STEM Skills



* $p<.05$

All of the Science Fusion teachers reported that their students would be better thinkers as a result of being in their class as compared to 67% of control teachers (see Figure 35).

Figure 35. Percent of Teachers Who Agree That Their Students will be Better Thinkers



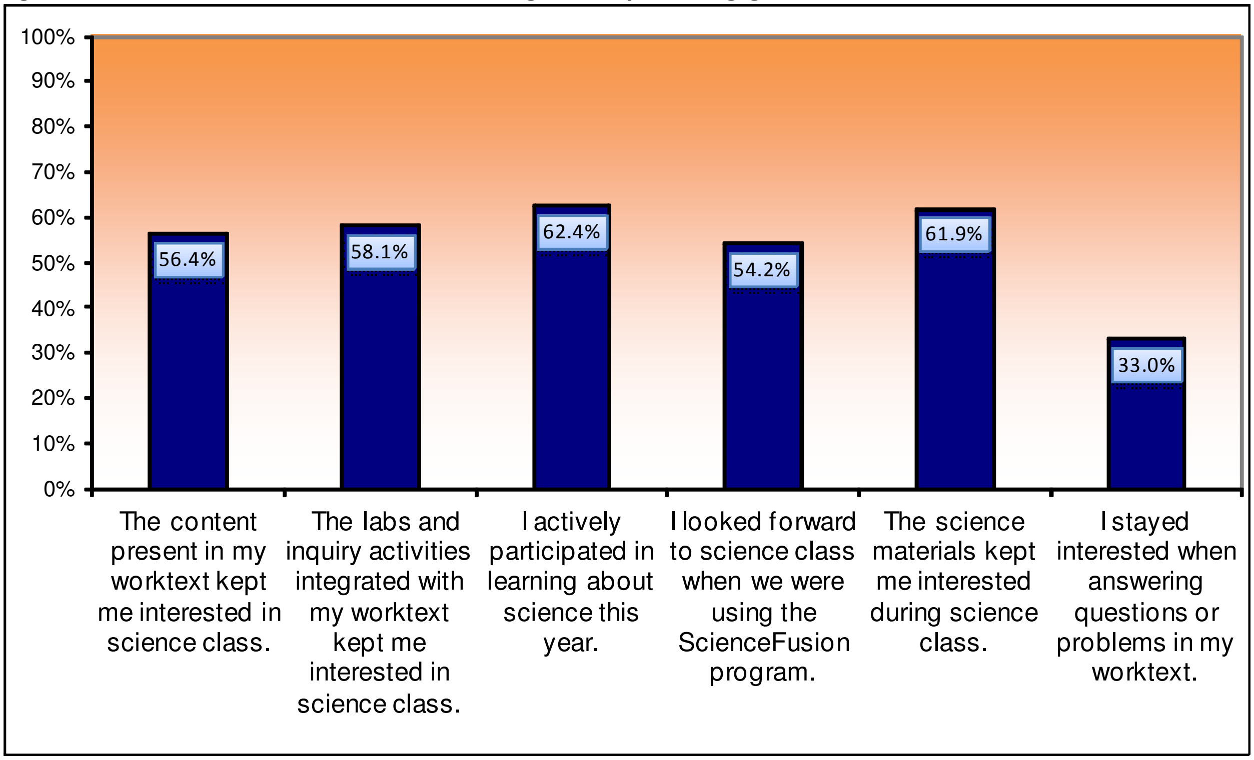
- *"It's awesome to see the students evolve, especially the 8th graders because they are able to discuss things like an adult. But my class that is not doing Science Fusion cannot do that." – Teacher, OH School*

The majority of Science Fusion students and teachers reported that the program helped students understand science and science vocabulary, and developed problem-solving, scientific inquiry, and STEM-related skills. Science Fusion students and teachers rated the impact of their science program on academic skills significantly higher than those using another program.

STUDENT ENGAGEMENT AND MOTIVATION

In general, about half of the Science Fusion students reported that they were engaged in science while using the Science Fusion program (see Figure 36). Treatment students commented that interactive aspects

Figure 36. Percent of Science Fusion Students Who Agreed They Were Engaged in Science



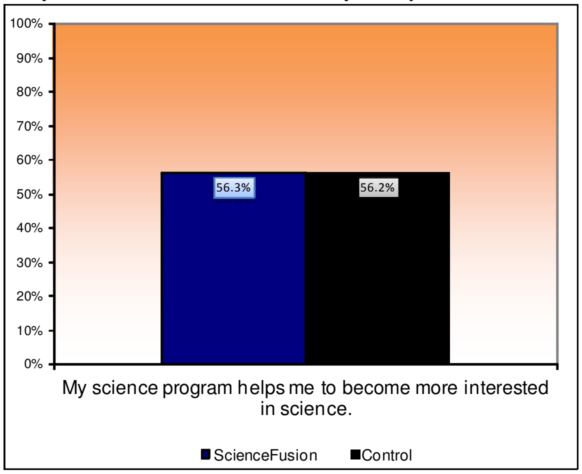
of the book along with the integrated lab activities were engaging. Some students also liked the interactive pictures and diagrams noting that it helped them feel more involved in learning science. Still, as noted in Figure 36, a large proportion of students also did not feel interested when answering questions from the worktext.

- *"The lab activities are more interactive and engaging." – Student, OH School*
- *"It is hands-on and very interactive." – Student, OH School*
- *"I liked the different pictures, I liked how the book gave examples." – Student, OH School*

As shown in Figure 37, a comparison of treatment and control students showed that both students using the Science Fusion program (56%) and students using the other science program (56%) felt that the content presented in their program kept them interested in science. In general, only half of students felt their program provided interesting science content. These findings are not surprising as research shows that student interest in science decreases as they move into secondary education (Rani, 2006). Thus, the lack of interest in their program's content may be more reflective of a general disinterest in science as opposed to their specific science program.

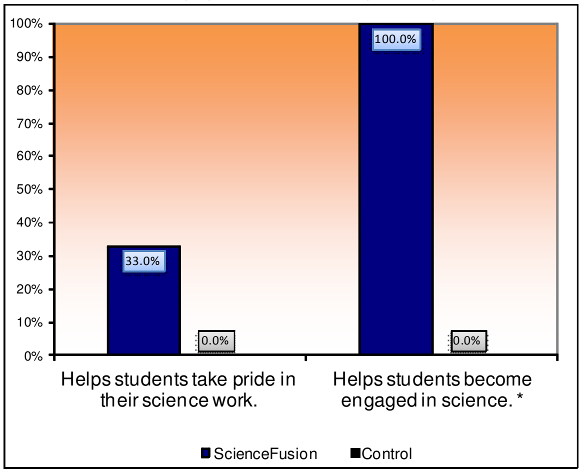
- *"Overall the Science Fusion kids are much more interested in science. The models and simulations are so much more beneficial for the kids, and it keeps them interested." – Teacher, OH School*

Figure 37. Student Perceptions of the Degree to Which They Were Interested in Science by Group



When teachers were asked about their students' interest in science (see Figure 38), all of the Science Fusion teachers reported the program helped students engage in science as compared to 0% of control teachers, $p<.05$. Also, though not statistically different, more Science Fusion teachers reported that the program helped students take pride in their science work (33%) as compared to control teachers (0%).

Figure 38. Teacher Perceptions of the Degree to Which Students Were Engaged in Science By Group



* $p<.05$

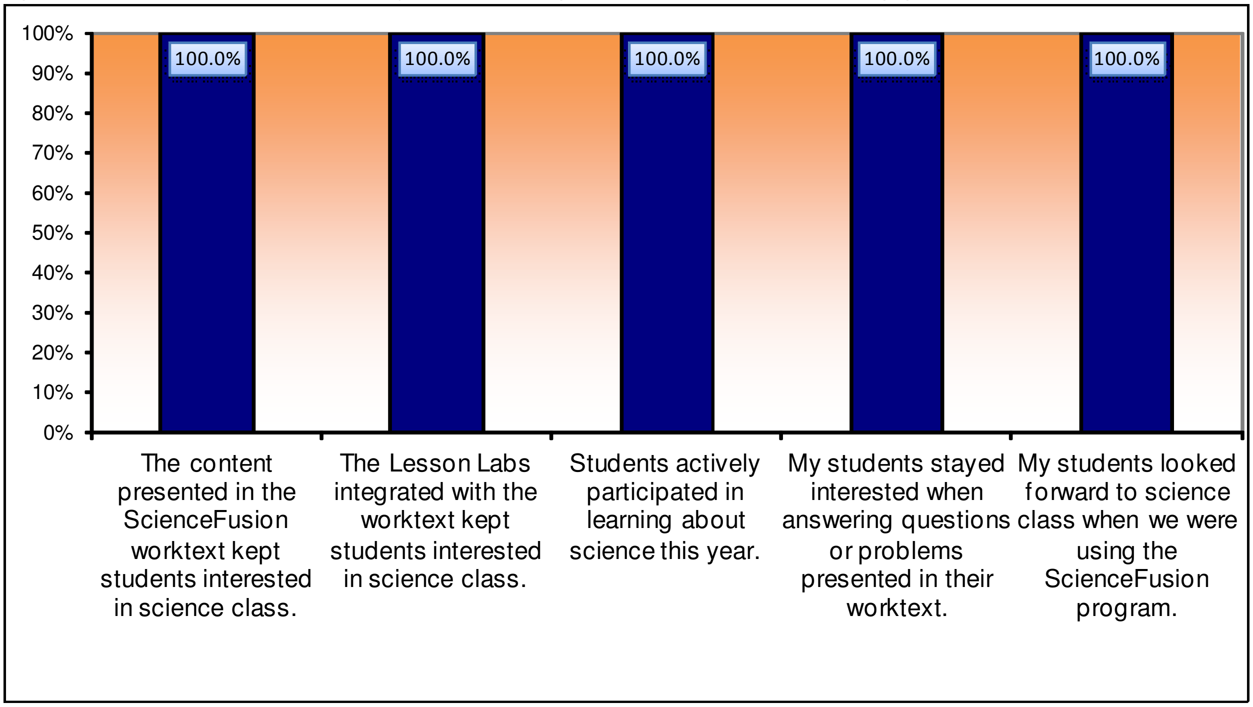
All of the Science Fusion teachers reported that their students looked forward to science and that the content in the Science Fusion student book kept their students interested in science class, (see Figure 39).

- *"The kids are reading the parts I don't even tell them to read and they are highlighting important parts and they actually really like learning." – Teacher, OH School*
- *"The students seemed to be more involved and interested in the content." – Teacher, OH School*

"The students that do not have Science Fusion are really sad, and very jealous. My students love this program." – Teacher, OH School

While only half of Science Fusion students reported that the content in their program kept them interested in science, the majority of their teachers felt that the Science Fusion program helped students engage in science and to take pride in their science work.

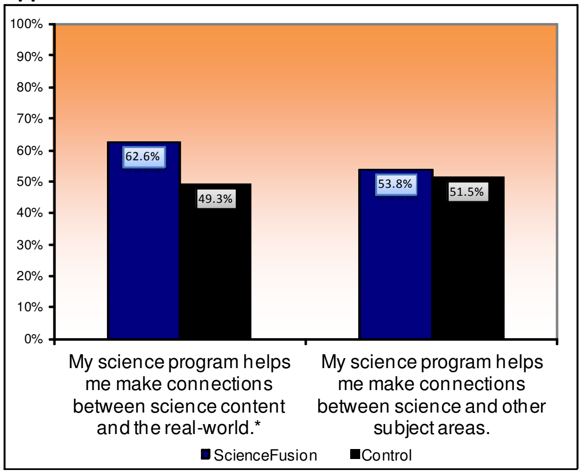
Figure 39. Science Fusion Teacher Perceptions of the Degree to Which Students Were Engaged in Science



CONNECTIONS WITH SCIENCE

When students were asked about the extent to which their science program helped them to apply and make connections between science and the real-world, significant differences were observed between treatment students (63%) and control students (49%), $p<.05$, (see Figure 40). Approximately half of the Science Fusion and control students reported that their science program helped them make connections to other subjects.

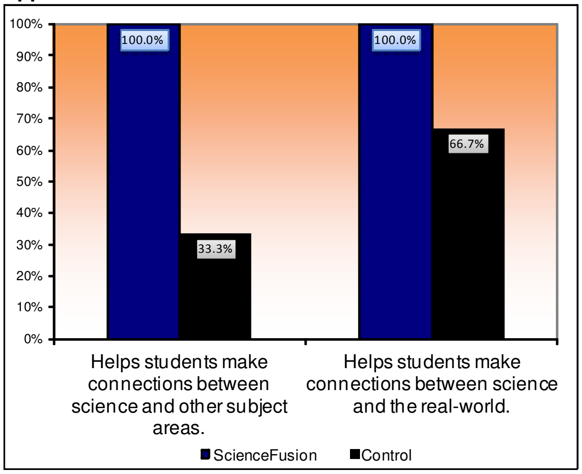
Figure 40. Perceptions of the Degree to which Science Program Helped Students with Science Connections and Applications: Science Fusion and Control Students



* $p<.05$

Though not statistically significant differences, Science Fusion teachers more often reported that their science program helped students make connections between science and the real world as well as other subjects (see Figure 41).

Figure 41. Perceptions of the Degree to which Science Program Helped Students with Science Connections and Applications: Science Fusion and Control Teachers



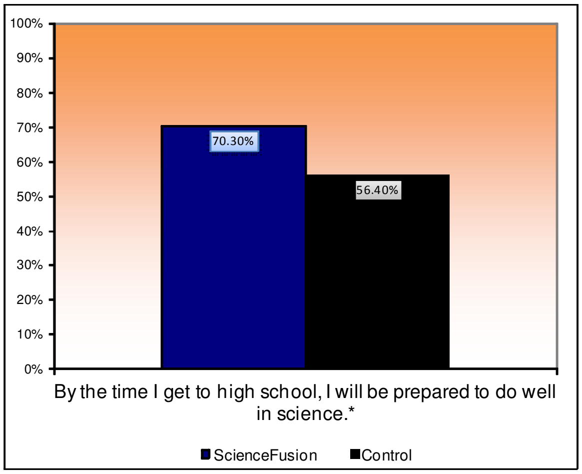
- *"Science Fusion made it easy to apply a lot of the science to my students' everyday lives; it made them really focus and engage." – Teacher, D.C. School*

Science Fusion students and teachers reported that their program helped them to make a connection between science and the real world.

PREPARATION FOR FUTURE TESTS AND SCIENCE COURSES

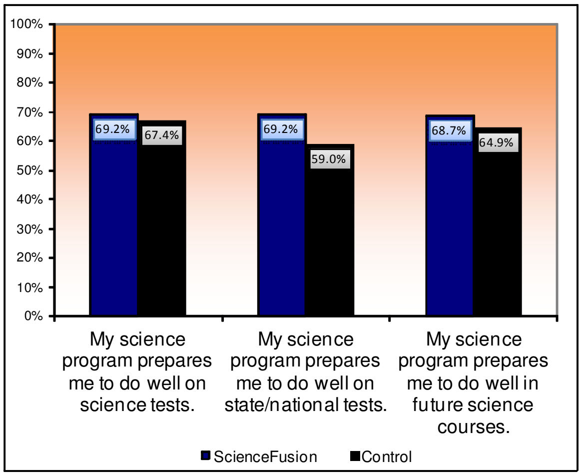
Analysis of student surveys revealed that Science Fusion students (70%) were significantly more likely to report that they would be prepared for science in high school than control students (56%), $p<.05$ (see Figure 42). As shown in Figure 43, slightly more treatment students agreed that Science Fusion prepared them to do well in future science classes and on state, national, and science class tests as compared to control students, although the difference was not statistically significant.

Figure 42. Perceptions of the Degree to Which Science Program Helped Students Prepare for High School Science By Group



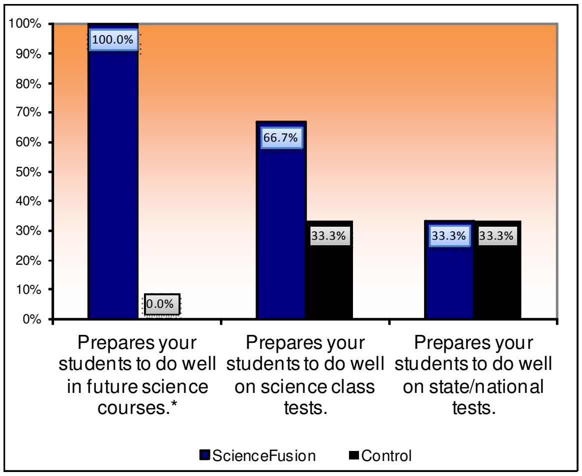
* $p<.05$

Figure 43. Perceptions of the Degree to Which Science Program Helped Students Prepare for Future Tests and Courses By Group



In contrast to student reports, Science Fusion teachers reported that their science program prepared students to do well in future science courses more often than control teachers, $p<.05$ (see Figure 44). Though not a statistically significant difference, Science Fusion teachers were also more likely to report that their program prepared students to do well on science class test. Less than half of the Science Fusion or control teachers reported that their students were prepared to do well on state or national tests.

Figure 44. Perceptions of the Degree to Which Science Program Helped Students Prepare for Future Tests and Courses: Science Fusion and Control Teachers



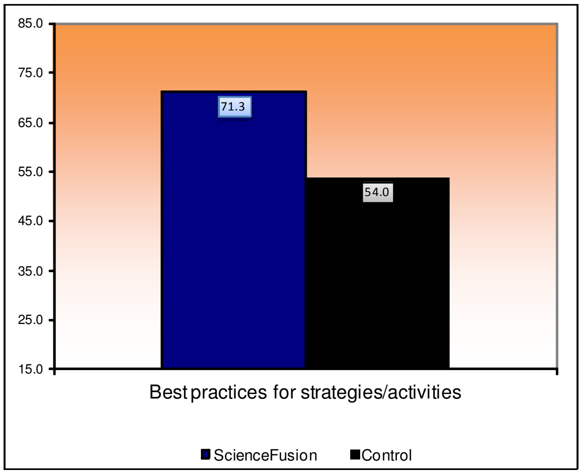
- *"Unit test scores seemed to increase." – Teacher, OH School*

The majority of both Science Fusion and control students indicated that they were prepared to do well in high school and other future science courses and that their program prepared them to do well on tests. Science Fusion teachers were more likely to report that their program prepared students for future science courses, however, neither group of teachers reported that their students were prepared for state tests.

TEACHER LEVEL OF PREPAREDNESS

Teachers were asked how prepared they felt to utilize best practices and strategies in science instruction such as providing concrete examples of concepts, applying concepts to a variety of contexts, using technology, and assessing student understanding. Results showed that both Science Fusion and control teachers reported that their science programs left them sufficiently prepared to utilize best practices (see Figure 45). While not a statistically significant difference, Science Fusion teachers reported higher overall levels of preparedness to use best practices.

Figure 45. Teacher Preparedness for Best Practices by Group



Mean teacher score on a summative scale with a range from 17 to 85 created from 17 items scored from 1-Not at all Prepared to 5-Very Prepared

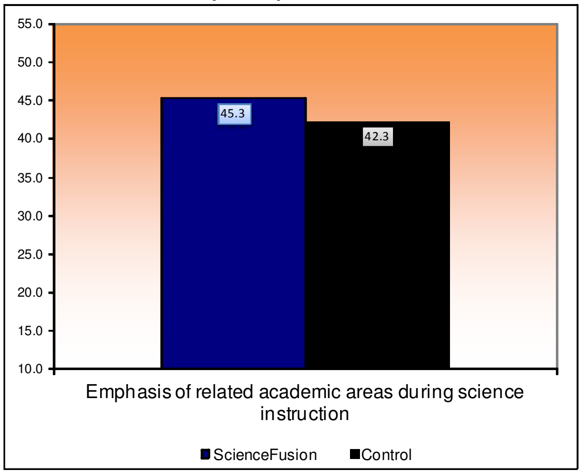
- *" I loved that the Teachers Edition gives you a refresher and overview; it helps me be better prepared to teach." – Teacher, DC School*
- *Using this particular program helped me learn the 8th grade curriculum so much better than the other books ever would have. I transitioned from 7th grade to 8th grade and this was a great way for me to transition and made me a better teacher." – Teacher, OH School*
- *"It has helped me be more structured in my teaching with all the resources available. I really enjoy how complete it makes my lessons." – Teacher, OH School*

INSTRUCTIONAL PRACTICES

Teachers were asked how much they emphasized areas such as vocabulary, writing, reading, STEM, and problem-solving during science instruction. Analyses of the Year Two data showed no statistically significant differences between Science Fusion and control teachers in the amount that teachers reported emphasizing each area (see Figure 46). However, in survey comments and interviews, treatment teachers reported that the content of the Science Fusion program encouraged or enabled them to include more focus on reading skills and STEM.

- *" I focus more on reading strategies, because Science Fusion has really made me hone in and I don't have to compare and contrast on my own anymore; it's built in." – Teacher, D.C. School*
- *"I wouldn't have taught as much STEM stuff if it wasn't built in to Science Fusion." – Teacher, D.C. School*

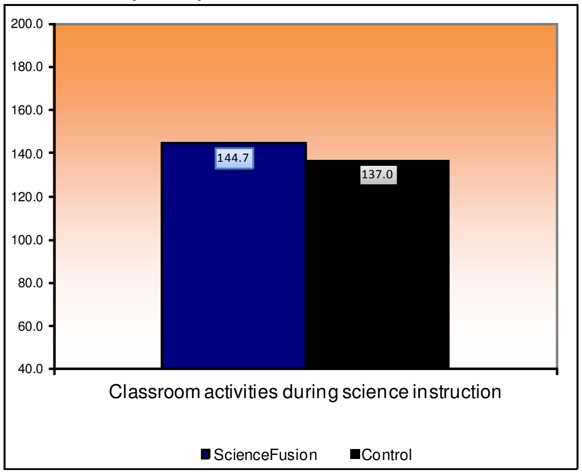
Figure 46. Teacher Emphasis in Science-Related Instructional Areas by Group



Mean teacher score on a summative scale with a range from 11 to 55 created from 11 items scored from 1-No Emphasis to 5-Very Strong Emphasis

Teachers were also asked about the typical classroom practices they incorporated into their instructional day, such as using computers or having students design experiments. Analyses of the Year Two data showed no significant differences in the instructional activities of Science Fusion and control teachers (see Figure 47).

Figure 47. Teacher Emphasis on Activities during Science Instruction by Group



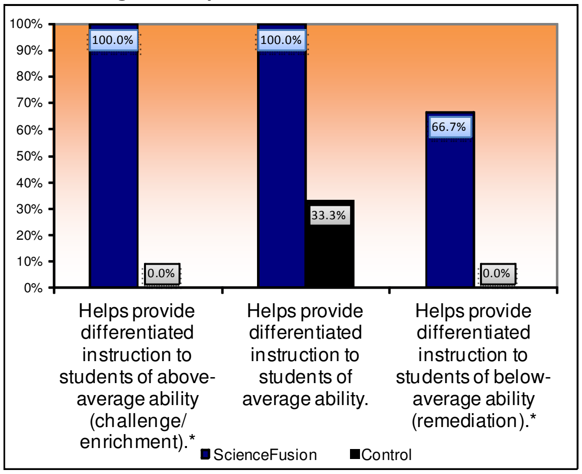
Mean teacher score on a summative scale with a range from 40 to 200 created from 40 items scored from 1-Never to 5-Every Day

Science Fusion teachers did report more preparedness to use best practices and strategies and placing more emphasis on reading strategies and STEM.

Differentiated Instruction

In general, most Science Fusion teachers reported that their science program provided them with assistance to provide differentiated instruction to students at all levels (low, average and advanced) (see Figure 48). Science Fusion teachers reported that their science programs provided more assistance with differentiated instruction for above-average students (100%) than control teachers (0%), $p<.05$ and for below-average students (67% compared to 0%), $p<.05$. While not a statistically difference, more Science Fusion than control teachers reported that their science program provided assistance with differentiated instruction for average students as well.

Figure 48. Teacher Perceptions of the Degree to which Science Program Helped with Differentiated Instruction



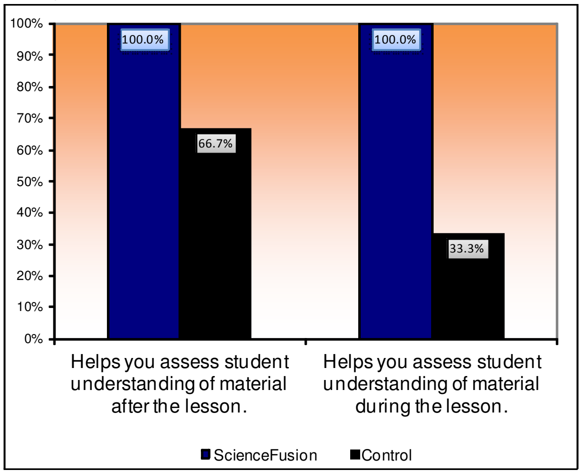
*$p>.05$

Over three-quarters of the Science Fusion teachers felt that the program provided them with assistance to provide differentiated instruction to students at average levels, and all Science Fusion teachers for above-average and below-average students.

Progress Monitoring

When analyzing perceptions about the assistance science programs provided in assessing student progress and learning, results from teacher surveys indicated that Science Fusion teachers generally perceived greater assistance from their programs (see Figure 49). Though not a statistically significant difference, 100% of Science Fusion teachers reported that their science program helped them to assess student understanding during and after the lesson compared to 67% of control teachers assessing understanding after the lesson and 33% of control teachers during the lesson.

Figure 49. Percent of Teachers Who Agreed their Science Program Helped Them Monitor Student Progress By Group

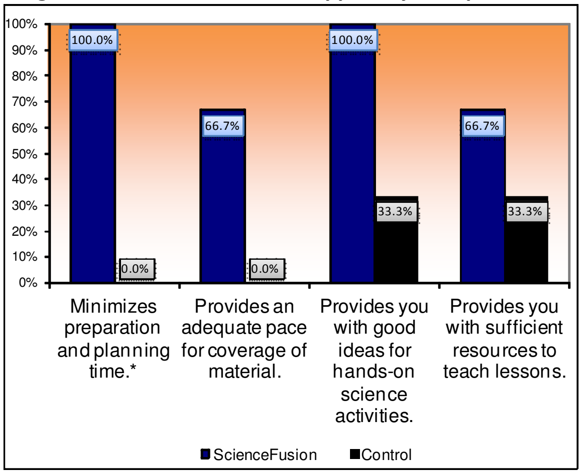


A higher percentage of Science Fusion teachers perceived assistance from their science program in helping them assess student understanding of science as compared to teachers using other science programs.

Teacher Support

Teachers were asked about the extent to which their science programs provided them with support (e.g., lesson planning, providing resources and ideas, etc.). As shown in Figure 50, Science Fusion teachers reported more support from their science program than control teachers. Science Fusion teachers were significantly more likely to report that their science program minimized preparation and planning time (100% vs. 0%), $p<.05$. Though not statistically significant differences, Science Fusion teachers were also more likely to report that their science program provided ideas for activities, adequate resources, and adequate pacing. (The percentage differences across treatment and control groups are quite large but not statistically significant; this is likely due to the small sample size for the groups in this study.) The patterns among Science Fusion teachers' responses are consistent with anecdotal comments. In particular, teachers indicated that the Science Fusion program gave them enough resources to choose the ones they wanted and quickly put together a complete and engaging lesson. Teachers were especially positive about the supplements such as the video series.

Figure 50. Percent of Teachers Who Agree Their Science Program Provided Them with Support By Group



*$p>.05$

- *"It has all the resources; it's one-stop shopping. Go to one website and everything you would ever need is right there. No more looking everywhere for things to implement." – Teacher, OH School*
- *"Science Fusion helps me teach quality lessons; I am using all the resources. It only takes me 10 or 15 minutes to see everything and have a great lesson ready to go." – Teacher, OH School*

Science Fusion teachers agreed that there were sufficient available resources and indicated that the program helped expedite prep time.

In summary, results showed that the Science Fusion program led to students being more likely to enjoy science, positively rate their own abilities, and to be motivated to learn science as compared to students in other programs. Furthermore, Science Fusion students and teachers were more likely than control students and teachers to report that the program helped develop students' academic, scientific inquiry, and STEM skills. Science Fusion teachers also reported that their students were more interested and engaged in science than control students.

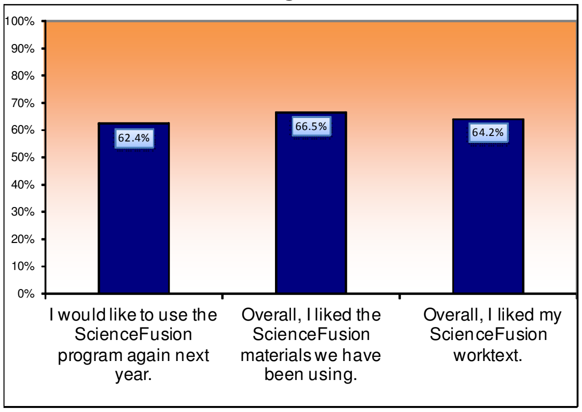
Science Fusion teachers reported that they were prepared to use best practice strategies and give quality lessons with differentiated instruction. Finally, the majority of the Science Fusion teachers reported that they found the program material helped facilitate their instruction and was sufficient to effectively teach their lessons.

What do users of the Science Fusion program think about the program? What aspects of the program do they find most useful? Least useful? What, if any, suggestions for program improvement do they have?

STUDENT PERCEPTIONS

Many students using the Science Fusion program (62%) indicated they would like to use the Science Fusion program next year, (see Figure 51). Over half of the treatment students agreed that they liked the Science Fusion materials they had been using (67%), and a similar amount said they liked their Science Fusion Student book (64%). This was an increase from Year 1 in which only a little over half the students agreed that they liked the Science Fusion materials they have been using.

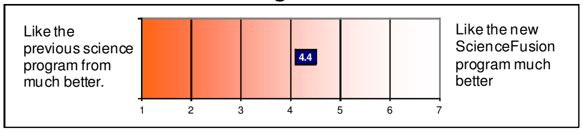
Figure 51. Percentage of Students Who Agreed They Liked the Science Fusion Program and Materials



- *"The new Science Fusion program gives me a better understanding than the old science program." – Student, OH School*
- *"I liked the hands-on activities, I enjoyed the digital lessons, I also enjoyed the classroom discussion on subjects connected to science." – Student, D.C. School*

Analysis of student surveys also showed that students in Science Fusion classrooms liked it somewhat better than their previous science program (see Figure 52).

Figure 52. Science Fusion Student Average Rating on the Extent to Which the Science Fusion Program Compares to the Previous Science Program



Students were split between favoring Science Fusion (36%), favoring the previous program (26%), and rating them as equal (38%). Students who favored Science Fusion often cited clear explanations, as illustrated in the comments below:

- *"Science Fusion helps me learn better because the concept is more understandable for me from my experience of using it for my 7th grade and 8th grade year compared to my 6th grade year." – Student, OH School*
- *"I think I'm doing better this year in science and really understand better." – Student, OH School*
- *"Science Fusion is better than last year because it gets in more depth in science terms." – Student, OH School*

Many of the "middle-of-the-road" students explained that it is the subject of science that they do not understand or that they don't find fun or interesting. For these students, the Science Fusion program did not change their perceptions about learning science.

- *"I don't like science. Nobody helps me and I never did well in science." – Student, OH School*
- *"I struggle with science unlike other subjects." – Student, D.C. School*
- *I believe they are both the same because they seem to provide all the information I need but just different concepts or material." – Student, OH School*

The remaining proportion of students felt the prior program was better for a variety of reasons, including the organization.

- *"Vocabulary needs to be explained and more evident, too many questions and answers are at the bottom of certain pages." – Student, OH School*

The majority of the students using the Science Fusion program indicated they were able to maintain good science notes over the school year (66%), see Figure 53. Most students reported they thought the science activities included in the program were fun and interesting (62%), and that they enjoyed reading their science worktext (58%). Furthermore, more than half of the treatment students indicated they used their worktext when studying for tests and quizzes (54%). These figures represent an increase from Year 1 in positive student attitudes toward the Science Fusion worktext.

Figure 53. Percentage of Students Who Agreed They Liked the Science Fusion Student Book

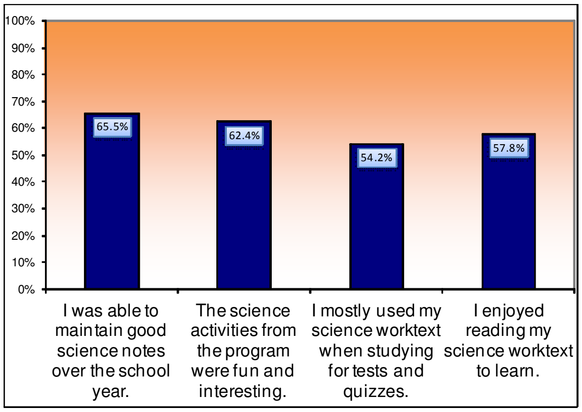
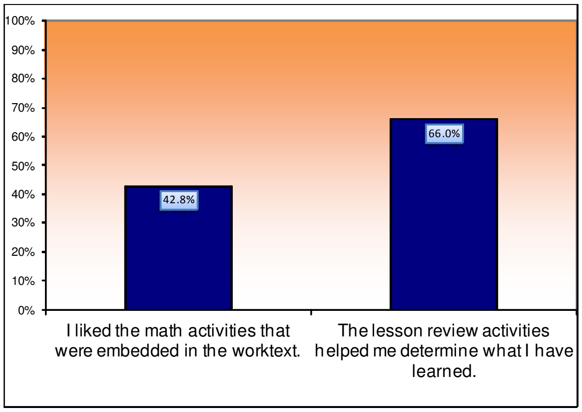


Figure 54. Percentage of Students Who Agreed They Liked the Math Activities and Review in Book



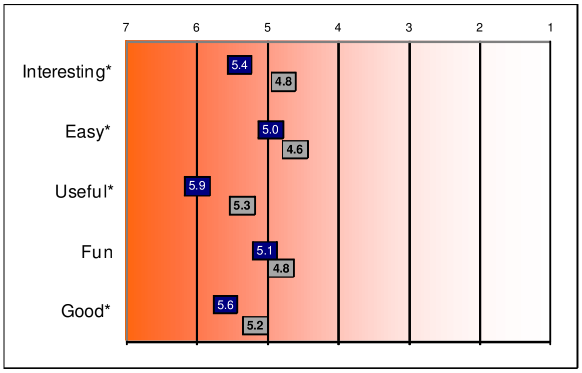
In sum, the majority of students liked the Science Fusion program. Students commented that they especially liked the ability to write in the books and the program's ability to engage students with interesting text and descriptive pictures.

- *"The [Science Fusion] book had better explaining graphs, pictures, and paragraphs." – Student, OH School*
- *"I liked the different pictures and I liked how the book gave examples." – Student, OH School*
- *"Science Fusion helped better because it was more explicit and it helps that I can write in the book." – Student, OH School*

Many students enjoyed using the Science Fusion program and 62% would like to use the program during the following school year. In general, students liked being able to write in their book.

Students were also asked to rate their respective science programs according to specific adjectives. Specifically, students were asked to rate the program on a scale (7=positive, 1=negative) from interesting to boring, easy to difficult, useful to useless, fun to not fun, and good to bad. Figure 55 shows the results of these ratings. There were significant differences in ratings between treatment and control students on four of the scales. In particular, Science Fusion students were more likely to rate the program as being good, useful, easy and interesting as compared to control students, $p<.05$.. Although treatment students were also more likely to rate their program as fun, these differences were not found to be statistically significant.

Figure 55. Average Descriptive Ratings by Science Fusion and Control Students



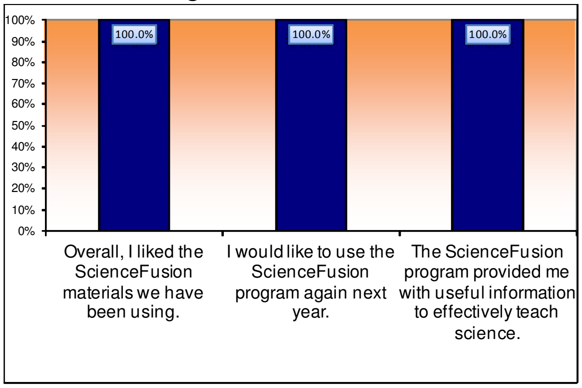
* $p<.05$ Science Fusion Control

Comparisons between treatment and control students on descriptive ratings (e.g., good, interesting, fun) revealed that Science Fusion students were more likely to rate their program as good, useful, interesting, and easy as compared to students using other science programs.

TEACHER PERCEPTIONS

Science Fusion teachers reported very positive perceptions of the Science Fusion program (see Figure 56). At the end of the second year, 100% of Science Fusion teachers agreed that they liked the materials, they would like to use Science Fusion again the next year, and that Science Fusion provided them with useful information to effectively teach science.

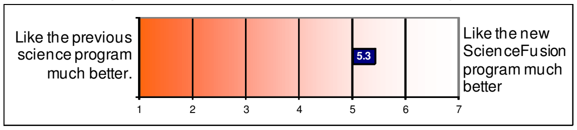
Figure 56. Teacher Overall Perceptions of the Science Fusion Science Program



- *"Students were receptive to this program and it was comprehensive for them." – Teacher, D.C. School*
- *"I feel that the curriculum is very well put together and encourages learning for students." – Teacher, OH School*
- *"The Science Fusion program is awesome." – Teacher, OH School*

Science Fusion teachers indicated that they liked the Science Fusion program somewhat better than the science program they had used previously (see Figure 57). They especially liked the visual appeal of the technology, the organization, and that students were able to write in their consumable student books.

Figure 57. Teacher Average Rating on Quality of Science Fusion Program Relative to Prior Science Program



- *"This Science Fusion program is more hands-on and effective for student learning." – Teacher, OH School*
- *"Compared to the other science program I've used I like Science Fusion more; it's just a better program all around." – Teacher, D.C. School*

Science Fusion teachers also reported that their students had positive perceptions about the program (see Figure 58). All of the Science Fusion teachers indicated that the science activities their students did from the program were very engaging. Additionally, they all reported that their students spoke about the student book in positive terms, enjoyed reading the worktext, and used it for studying for tests and quizzes.

Figure 58. Teacher Perceptions about Student Attitudes Toward Their Science Program

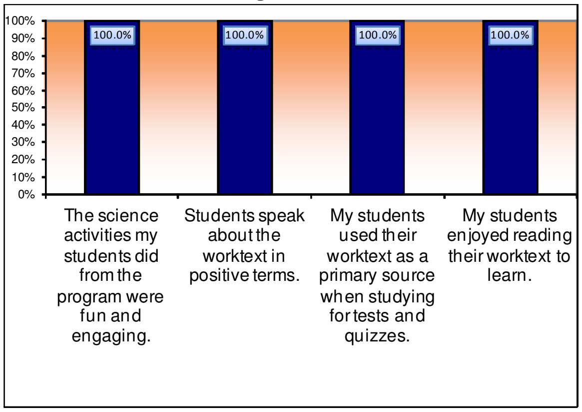
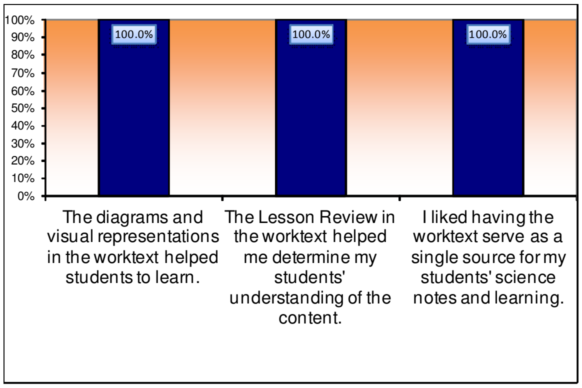


Figure 59. Teacher Perceptions about Elements of the Science Fusion Student Book



In rating elements of the worktext (see Figure 59), all of the Science Fusion teachers reported that the diagrams and visual representations in the student book helped their students learn and that the Lesson Review in the student book helped them determine their students' understanding of the content. All of the teachers also agreed that they liked having the student book serve as a single source for their students' science notes and learning.

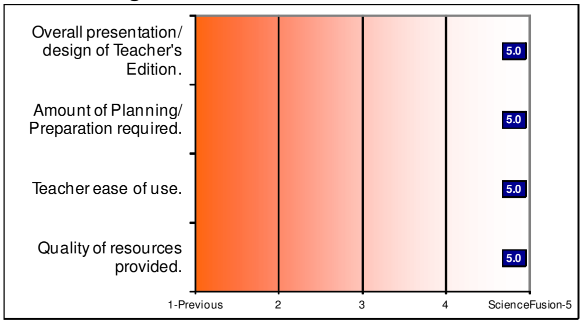
- *"The kids love the worktexts; it's bendable, they can write in it, it's lightweight and it makes it easy for them to pack around so they seem to always have it ." – Teacher, OH School*
- *"Being able to write in the book is amazing." – Teacher, OH School*
- *"They like that the pictures are in color and the homework is in color which makes the students who don't use this program jealous. I feel it gives them an advantage because they have the ability to reference charts, diagrams and pictures a lot easier. They can draw in the books, so if I am labeling something they can too right in their book which makes it much more user-friendly." – Teacher, OH School*
- *"The books are so catchy; they look like a magazine instead of a textbook and the students really like that!" – Teacher, OH School*

Science Fusion teachers rated the program very highly and agreed that they would like to use it again the next year. They also rated the elements of the worktext highly, and reported that they liked having the worktext for students' science notes.

SCIENCE FUSION RESOURCES AND PROGRAM COMPONENTS

When asked to compare the Science Fusion program to their previous science program in terms of quality and design, the treatment teachers felt that the Science Fusion program was better than their previous program (See Figure 60). Specifically, teachers preferred Science Fusion for the overall presentation/design of the Teacher's Edition, the amount of planning and preparation required, ease of use, and quality of resources provided.

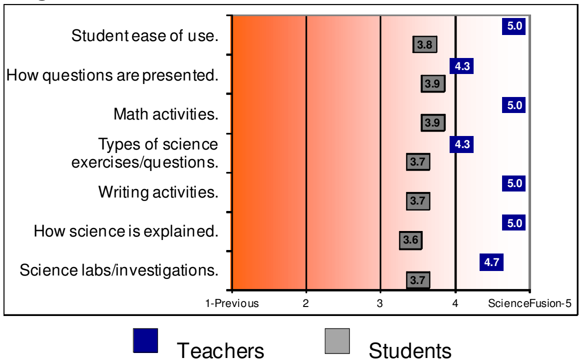
Figure 60. Teacher Attitudes about Ease of Use and Resources of Science Fusion Program Relative to Prior Science Program



Treatment teachers preferred Science Fusion to their previous science program for design of the Teacher's Edition, the amount of planning and preparation time, ease of use, and quality of the resources.

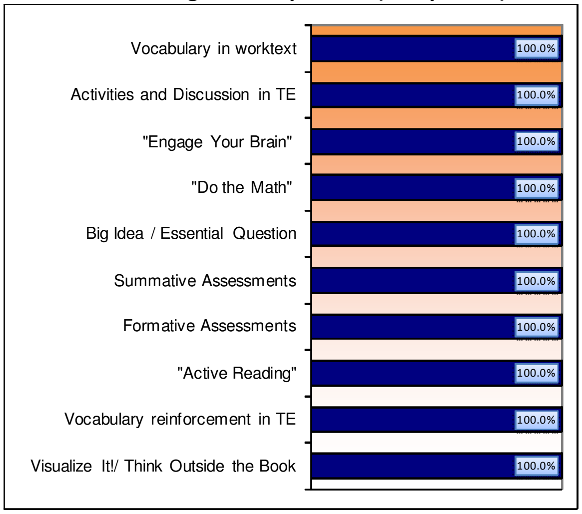
Teachers and students were asked to compare additional program activities available in the Science Fusion program to the previous science program (see Figure 61). Both teachers and students rated Science Fusion components as preferable to the components of the science program they used previously (i.e., ratings are above the midpoint of 3.0). Though teachers rated all components more highly than students, teachers and students preferred Science Fusion over the previous program in student ease of use, how science is explained, how questions are presented, types of science exercises, writing activities, math activities, and science labs/investigations.

Figure 61. Teacher and Student Attitudes about Science Fusion Program Activities Relative to Prior Science Program



Ratings of specific Science Fusion components (see Figure 62) indicated that all Science Fusion teachers found key components of the program to be useful. Specifically, teachers indicated that the vocabulary section of the student book, discussion sections, Engage your Brain, Do the Math, Big Idea, Essential Question, summative assessments, formative assessments, Active Reading activities, and Visualize It/Think Outside the Box activities were useful components of the program.

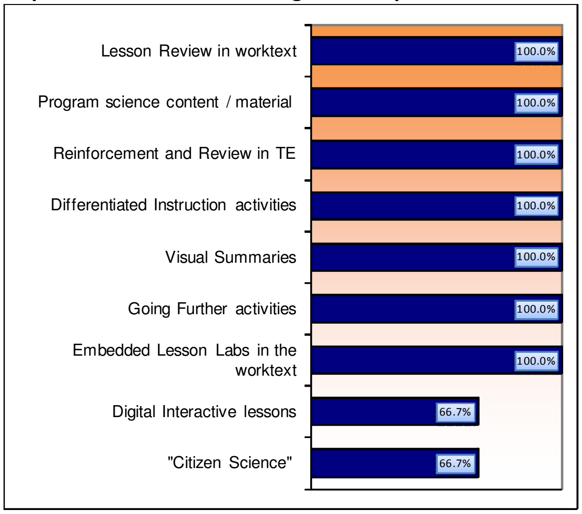
Figure 62. Teacher Ratings of the Usefulness of the Core Science Fusion Program Components (except Labs)



- *"I think Science Fusion is an excellent system; I love the pictures and the vocabulary." – Teacher, D.C. School*

A review of ratings of the Science Fusion components that were considered optional indicates that the majority of teachers found many of these components to be useful as well (see Figure 63). Specifically, all teachers rated the Lesson Review, science content / material, Reinforce and Review activities, differentiated instruction activities, visual summaries, going further, and embedded lesson labs as useful. The lowest rated items (67%) were the Digital interactive lessons and the Citizen Science Unit project. In anecdotal comments, teachers reported that the online student edition of the textbook was very useful, and that while they loved the visual summaries, they did not like that the answers were given to students.

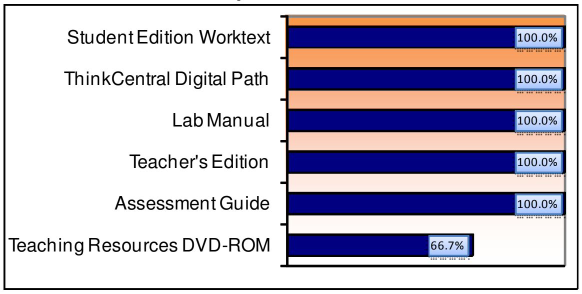
Figure 63. Teacher Ratings of the Usefulness of the Optional Science Fusion Program Components



- *" I really liked the electronic student edition, because it has a good sequence with a lot of engaging components; there was really a connection there." – Teacher, D.C. School*
- *" My students became more knowledgeable and worked more independently, especially using the electronic student edition." – Teacher, D.C. School*
- *" Being able to walk through the lessons at home online if a student is absent so they can catch up is great." – Teacher, OH School*
- *"The visual summary is a great conclusion or assessment for teachers and so I really dislike that the answers are in the book. Even if they were in the back of the book it would help." – Teacher, OH School*

Teachers also rated the usefulness of print materials and ancillary resources (see Figure 64). All of the teachers indicated that the Write-In Student Edition, ThinkCentral Digital path, Lab Manual, Teacher's Edition, and Assessment Guide were useful resources. About two-thirds of the teachers found the Teaching Resources DVD-ROM useful. While teachers found the assessments useful, many teachers made anecdotal comments about ways to improve assessment options.

Figure 64. Teacher Ratings of the Usefulness of Print Materials and Ancillary Resources



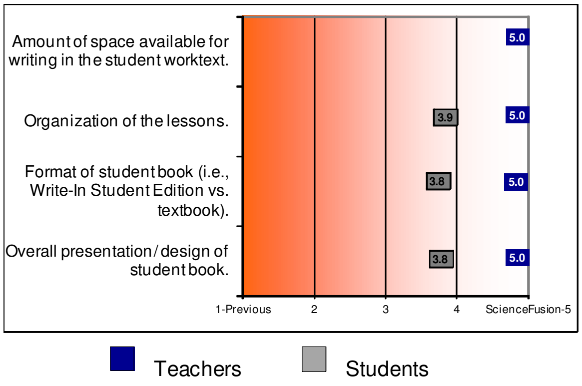
- *"I didn't find the assessment tool useful at all; pulling from the bank was a hassle and took a lot of time." – Teacher, D.C. School*

- *"The only thing that would make it better is having access to a bunch of multiple choice questions!" – Teacher, OH School*

Write-In Student Edition

Treatment teachers and students were asked to rate aspects of the Science Fusion Write-In Student Edition in comparison to their previous program's textbooks (see Figure 65). Both teachers and students rated the format and organization and the overall presentation/ design of the Science Fusion student book as better than their previous science textbook, although teachers showed a greater preference towards Science Fusion than students. Teachers also indicated that the amount of space for writing in the student book was better than the previous student text (students were not asked about the amount of writing space).

Figure 65. Teacher and Student Attitudes about Organization and Design of Science Fusion Write-In Student Edition Relative to Prior Science Program



- *"From a student point-of-view the students love the books. The students that do not have the Science Fusion books are always jealous of the students who do have the books." – Teacher, OH School*

In general, students reported that the material was easier to understand, which can be attributed at least in part to the design of the text. Many students explained that because they could write in their text, it made the Science Fusion student book better than last year's science book. Some students also described the length of the passages, the use of visuals, or how other elements of the design of the text encouraged their learning.

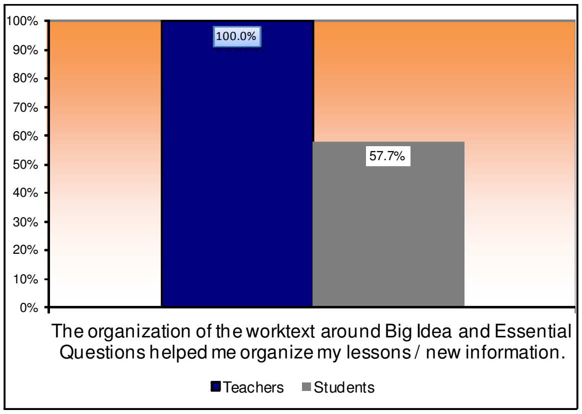
- *"Science Fusion helps me understand science more, helpful/interesting pictures, easy note taking" – Student, OH School*

- *"It [ScienceFusion] explains it [science] clearly, it helps me take notes easier, and it's one of the first things that I study off of." – Student, OH School*

Findings also show that a little more than half of the Science Fusion students (57%) and all of the teachers agreed that the organization of the Write-In Student Edition around the Big Question helped them to organize their lessons or new information (see Figure 66). Feedback from teachers and students indicate that it is not so much the organization of the write-in student book around the Big Idea or Essential Questions that they found beneficial, but the fact that the students had their own workbooks that they could take notes in and use to study for tests and exams. Multiple comments illustrate that teachers and students value the write-in student book.

- *"The students love being able to write in the books and everything we needed was right there." – Teacher, D.C. School*
- *"My students liked being able to carry around the worktext and they liked taking ownership of it." – Teacher, D.C. School*

Figure 66. Percentage of Science Fusion Teachers and Students That Agreed They Liked the Organization of the Student Write-In Student Edition

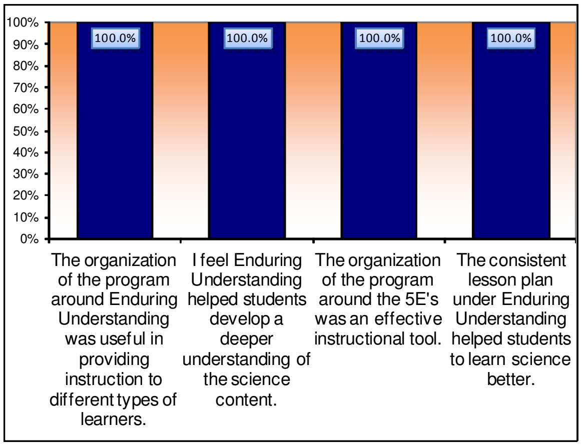


Science Fusion teachers rated nearly all of the program components as useful. They reported preferring the Science Fusion components to the previously used program. Students and teachers agreed that the Science Fusion textbook was preferred to the previously used textbook, and the majority found the organization of the program to be helpful.

Program Pedagogy

Science Fusion's pedagogical design, Enduring Understanding, was perceived as useful and promoting a deeper level of student understanding by all of the treatment teachers. While the results of Year One data showed mixed reviews, Year Two data and comments reflect an appreciation for the organization and overall design of the program.

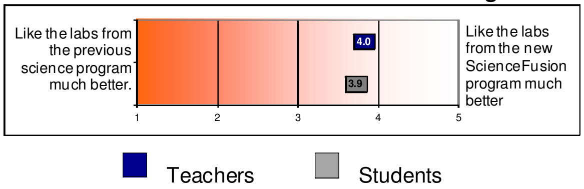
Figure 67. Treatment Teacher and Student Rating of Science Fusion Labs Relative to Prior Science Program



Science Labs

On average, treatment teachers and students reported liking the Science Fusion labs somewhat better than the labs from their previous science program (see Figure 68). Furthermore, all of the Science Fusion teachers and about half of the Science Fusion students (56%) reported that they liked the labs and investigations used as part of the program (see Figure 69).

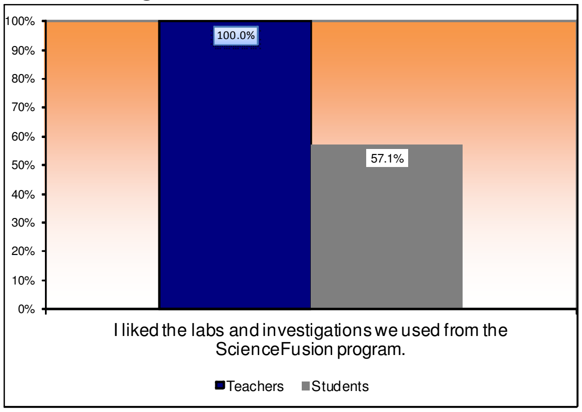
Figure 68. Treatment Teacher and Student Rating of Science Fusion Labs Relative to Prior Science Program



- *"The labs go right with the lessons, which I love. They make my life as a teacher easier in terms of trying to gather materials - those kits are wonderful." – Teacher, OH School*

- *"Some of the labs from the other program were easier for students to complete on their own; had very detailed instructions with pictures." – Teacher, D.C. School*

Figure 69. Teacher and Student Ratings of Labs included in Science Program



- *"I had all the materials, data sheets, and instructions at my fingertips." – Teacher, OH School*

- *"The stuff is all ready for you in the labs. No more fishing around to find materials; it makes you want to use the labs." – Teacher, OH School*

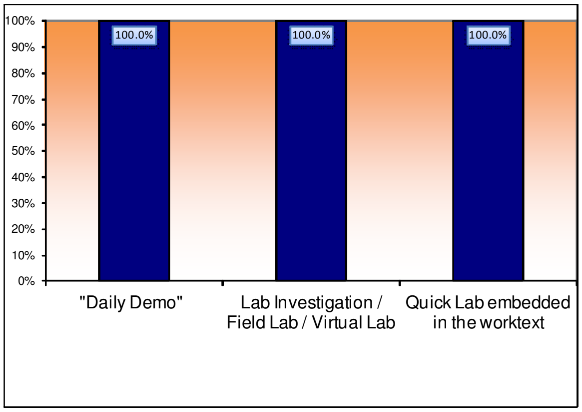
Some students reported not finding the Science Fusion labs as interesting as the labs they used in their prior year's program. Still other students were much more positive and described what they liked about the Science Fusion program's labs. This variation in student perspectives about the Science Fusion labs is reflected in the comments below.

- *"The labs were much more fun than last year's labs." – Student, OH School*

- *"The labs helped me to know how the subject worked and it made me understand it more." – Student, OH School*

Figure 70 presents teacher ratings of the usefulness of various types of Science Fusion labs. All of the treatment teachers indicated that they found the "Daily Demo" activity, the lab investigation / field lab / virtual lab, and the Quick Lab in the worktext useful.

Figure 70. Teacher Ratings of the Usefulness of Science Fusion Lab Activities



- *"The most useful parts of Science Fusion for me are the digital lessons and the virtual labs. The students are so keen with technology they always want to be on the computers." – Teacher, OH School*

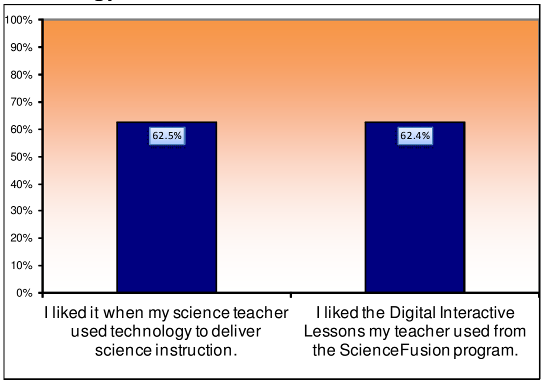
- *"The virtual labs break it down and show them examples of what the terms mean so you really know what you are learning." – Teacher, OH School*

While a little over half of the students reported liking the Science Fusion labs, all of the teachers reported liking the labs and preferring the Science Fusion labs to previously used labs.

Technology

Similar to teacher and student feedback on the Science Fusion labs, perceptions about the program's technology varied. About two-thirds of the students reported that they liked when their science teacher used technology to deliver science instruction (63%) and that they liked the Digital Interactive Lessons from the Science Fusion program (62%) (see Figure 71).

Figure 71. Student Ratings of the Science Fusion Technology

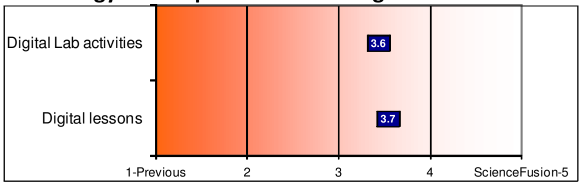


- *"I liked the hands-on activities, I enjoyed the digital lessons. I also enjoyed the classroom discussion on subjects connected to science." – Student, D.C. School*

- *"I like the labs, online lessons and textbooks." – Student, D.C. School*

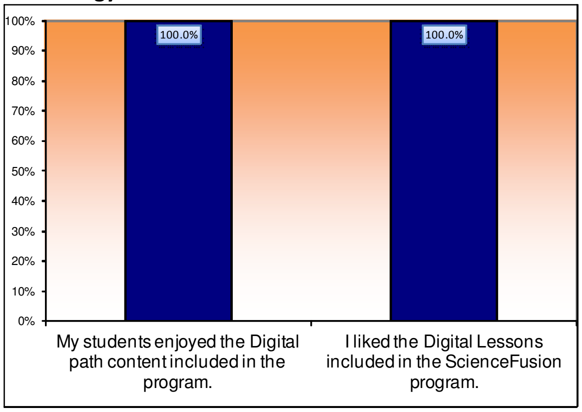
When asked how technology features of the Science Fusion program compared to last year's technology components, students reported that digital lab activities and lessons were slightly better than last year's program, (see Figure 72).

Figure 72. Student Ratings of the Science Fusion Technology as Compared to Prior Program



All of the Science Fusion teachers reported that their students enjoyed the digital path content and they reported liking these components themselves (see Figure 73).

Figure 73. Teacher Ratings of the Science Fusion Technology



- *"With all the technology and the mini projects it helped the students get excited about it. They were competing with each other to learn. It made me as a teacher excited to do the projects with the students as well." – Teacher, OH School*

Many students reported that the technology and digital components were one of the three things they liked best about the course. Some students listed "virtual labs" and "digital lessons" as the particular elements they liked.

Both students and teachers reported liking the technology used in the Science Fusion program.

Teachers noted a variety of specific program components when asked to identify the three things they liked best about the Science Fusion program. However, a couple of items emerged as favorites from many teachers, including:

- Organization and design
- Write-in student text
- Labs

Teachers also noted the quality of the videos, the alignment with state objectives, and the innovation and reflection of current trends in science.

While overall the teachers liked the Science Fusion program, they also had some feedback about the program and potential areas for improvement. The primary area(s) that teachers noted as needing improvement were:

- Variety of test bank questions
- Removal of answers in visual summaries

Teachers also requested a notes section to accompany the videos. One teacher mentioned that students did not like switching books and another teacher requested more training on using the Science Fusion program.

Conclusion

Results obtained over the course of the two year randomized control trial designed to look at the effects of the HMH Science Fusion program on student learning showed significant overall effects on student science learning. Science Fusion students demonstrated higher test scores on the Developed Science Test, an assessment developed to measure content areas taught over the course of the school year. On the ITBS Science test, although a similar trend was evident with Science Fusion students showing higher test scores than control students after controlling for pretest differences, no significant differences were observed. It is also noteworthy that the effect sizes for the effect on the Developed Science Test were moderate, with a range of .48 to .64. Indeed, all effect sizes obtained exceeded the threshold for educational significance (.25) which means that these findings are meaningful in terms of impacting a students' educational experience.

Survey information obtained from teachers and students supports these findings as well. Science Fusion students and teachers were more likely than control students and teachers to report that the program helped develop students' academic, scientific inquiry, and STEM skills.

Affective outcomes also showed that the Science Fusion program led to students being more likely to enjoy science, positively rate their own abilities, and to be motivated to learn science as compared to students in other programs. Science Fusion teachers reported that their students were more interested and engaged in science than control students. These teachers also reported that they were prepared to use best practice strategies and give quality lessons with differentiated instruction. Finally, the majority of the Science Fusion teachers reported that they found the program material helped facilitate their instruction and was sufficient to effectively teach their lessons.

In sum, results from this two-year RCT show that students who use the Science Fusion program perform significantly better than students using other science programs as measured by an assessment designed to measure specific content taught over the course of the school year. Such positive treatment effects were observed in multiple areas (vocabulary and scientific reasoning) and findings suggest a stronger effect following two years of usage of the Science Fusion program. This is to be expected given that it takes time for teachers and students to become accustomed to using a specific program and for effects to be realized. To conclude, the Science Fusion program has a positive impact on student science performance relative to other science programs.

Appendix A:

Technical Appendix

Overview of the Technical Appendix

The purpose of this appendix is to provide fellow researchers with additional technical information to fully evaluate the scientific rigor of this study. Specifically, this appendix is written for technical audiences so that they may examine the statistical procedures employed as well as make more informed judgments of the internal and statistical conclusion validity of this study. It is *not* written for lay people. This *Technical Appendix* contains the following information:

- Analytical goals of these analyses
- Analytical framework
- Results of data analyses by analytical framework

Analytical Goals

The evaluation of the Science Fusion program focuses on the following broadly-framed goals:

1. *Assessment of effectiveness of the Science Fusion Program:* The Science Fusion program is examined in comparison to other middle school science programs. The analytical framework used to identify the effectiveness of the Science Fusion program is causal in a numbers of ways:

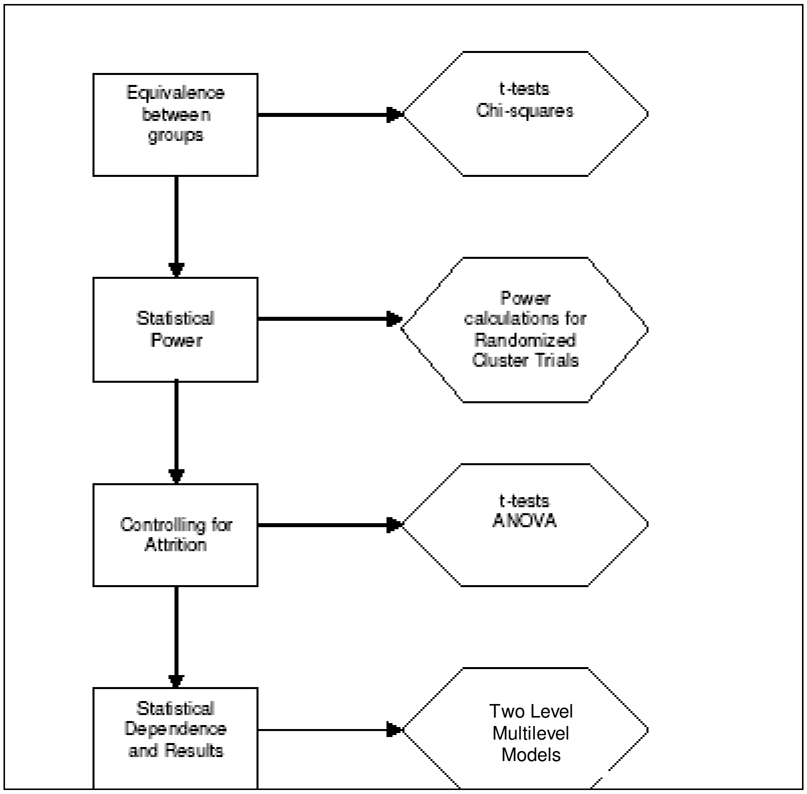
   (i) As described in the body of this final report, a well-planned randomized control trial was implemented;
   (ii) The analytical procedures pay close attention to multiple threats to internal validity including selection effects and attrition (Shadish, Cook, and Campbell, 2002);
   (iii) Given that students are "nested" within classrooms, the data are unlikely to be independent across students; dependence in outcomes is modeled by implementing hierarchical linear models (Raudenbush and Bryk, 2002);

2. *Knowledge development:* The implemented design also provides an opportunity to examine student and classroom/program measures that may be associated with program effectiveness for the Science Fusion program. This relationship between student and classroom characteristics and program effectiveness is viewed as primarily associative and not causal for two reasons: (a) The implemented design is focused on estimating causal *main effects* for the program; the statistical power to identify program effects within subgroups is much lower; (b) There have been very few studies that have examined subgroup effects of the Science Fusion program as well as science interventions as a whole. In the absence of a strong program theory, the subgroup effects are viewed as empirical patterns that need theoretical frameworks and other rigorous experimental designs in the future to be estimated "causally."

Analytical Framework

Figure A1 below and accompanying narrative show the four-step analytical procedures that were implemented to evaluate the effectiveness of the Science Fusion program.

Figure A1. Description of Analytical Framework



(i) *Establishing group equivalence*: The differences in the treatment and control group were examined by conducting t-tests and chi-square analyses at the student, class and teacher levels on a range of baseline outcomes and other student and teacher characteristics. Care was taken to ensure that measures on which the groups differed significantly were used as covariates in subsequent analyses.

(ii) *Statistical power*: Dependency in the data decreases the statistical power to detect significant differences. Specifically, increased values of intra-class correlations (higher dependency in the data) results in reductions in statistical power. The power to detect significant differences in clustered random trials was calculated for a range of intra-class correlations and effect sizes, and also with and without a cluster covariate.[26]

(iii) *Controlling for attrition*: In this step, consideration is given to attrition as a potential threat to both internal and external validity of the study (Cook and Campbell, 1979).

[26]The use of a cluster-level covariate that is correlated with the outcomes of interest increases the power of the test (Raudenbush et al., 2005).

Both issues of measurement attrition (i.e., missing data due to student absences or lack of test administration) and dropout attrition (i.e., missing data due to students leaving the study) were examined.

Measurement Attrition

First, chi-square analysis was performed to determine if the proportion of measurement attrition was equivalent among both groups. In other words, this analysis examined whether there was a significant relationship between students who provided and did not provide data (at *each* time point) and group assignment (treatment vs. control). Second, ANOVAs were run to determine whether there were performance differences between those who completed the tests and those who did not by group using posttest measures (to examine those not providing pretest measures) and pretest measures (to examine those not providing posttest measures). An interaction between group and test completion status would be indicative of a bias because the type of treatment students who did not complete the test would be different than the type of control students who did not complete the test.

Dropout Attrition

The potential problems of overall attrition and differential attrition due to students leaving the study was first "diagnosed" using a simple statistical procedure; specifically, chi-square analysis was conducted to determine if the proportion of dropout attrition was equivalent among both groups. Second, in order to determine whether there was differential attrition on pretest measures, ANOVAs were run to determine if there was (1) a significant interaction between group and attrition status, and (2) a significant main effect for attrition status (Cook and Campbell, 1979). A significant interaction would indicate a threat to internal validity because the type of student dropping out of the treatment group would be different than the type of student dropping out of the control group. A significant main effect would indicate a threat to external validity because the students remaining in the study would be different than the students who dropped out of the study.

(iv) *Statistical Dependency and Results:* Two-level multilevel models were implemented to estimate program effects. In the two-level model, student outcomes are modeled at level 1 while controlling for pretest differences. Teacher characteristics are modeled at level 2. Appendix B describes the mathematical equations representing the two-level multilevel models.

Results

This section is organized according to the aforementioned analytical framework.

1. Establishing Group Equivalence

a) The relationship between various student demographic variables and group status was examined. Results showed that one demographic variable was significantly associated with group. In particular, there were more special education students in the control group than the treatment group, $p<.05$. That said, the number of special education students is small (n=22). For more information, see Table 4 within the main report.

a) Pre-test differences on the assessment measures were examined, see Table A1. Student level t-test analyses revealed two significant differences on the ITBS and Developed Science Tests, $p<.05$. Treatment students had significantly lower pretest scores than control students on the ITBS Science overall scale score and significantly higher scores on the Developed Science Test-short answer items. Thus, treatment and control students were not equivalent with respect to pretest science performance on these measures.

Table A1. Sample Size, Means, Standard Deviations, and *t*-test (Student Level) Results for Assessments at Pre-testing

Pretest*	Group	N	Mean	Std. Dev.	*t*	Sig. Level
ITBS Science Test - Overall	Control	236	211.57	30.14	2.513	.012
	SF	288	204.59	32.80		
ITBS Scientific Inquiry	Control	236	32.74	18.95	1.112	.267
	SF	288	30.92	18.26		
ITBS Life Science	Control	236	32.03	19.11	1.159	.247
	SF	288	30.07	19.31		
ITBS Earth Science	Control	236	29.31	19.53	-.033	.974
	SF	288	29.37	22.64		
ITBS Physical Science	Control	236	38.82	20.41	-.347	.729
	SF	288	39.47	22.35		
Developed Science Test (DST)-Overall	Control	232	30.61	12.62	-.728	.467
	SF	288	31.38	11.46		
DST Vocabulary (fill in the blank items)	Control	232	36.85	24.10	-.803	.422
	SF	288	38.54	23.61		
DST Science Application (constructed-response items)	Control	232	6.77	10.42	-3.237	.001
	SF	288	9.90	11.37		

b) Data from the pre student survey were also examined. Results showed no significant differences between treatment and control students in perceived parental support, mother's educational background, father's educational background, amount of English spoken at home, participation in extracurricular activities, school engagement, perceived science ability, science enjoyment, science effort/motivation, science anxiety, interest in

STEM careers, and educational aspirations, $p>.05$. However, there was one significant difference on perceived importance of science, with treatment students reporting greater perceptions about the importance and usefulness of science than control students.

c) Recall that for the most part, classrooms within teachers were randomly assigned. As such, teacher differences were minimized. Nevertheless, teacher data was examined at the classroom level to determine if significant differences existed. Results showed no significant difference between teachers in terms of perceptions of autonomy in setting instructional goals, extent to which different types of students may hinder teaching, preparation to teach various science topics, pedagogical leanings, comfort with technology, access to resources to teach science and knowledge of NTSA standards, $p>.05$.

d) Classroom environment and implementation of various typical activities that occur in science classrooms were also analyzed based on information collected from the teacher logs and teacher surveys collected in the Fall of each school year (data from Year 1 and Year 2 teachers were combined). Results showed no significant differences between treatment and control classrooms in terms of student engagement, independent practice, lab activities, provision of differentiated instruction, assessment use, and prior technology use by teachers and students. However, differences were observed in the areas of classroom environment, $t(25)=2.76, p=.01$, and teachers' engagement in intervention activities, $t(25)=2.24, p=.03$. Treatment teachers reported that they had a more positive classroom environment and engaged in more intervention activities than control teachers at baseline.

In summary, randomization was reasonably successful in producing equivalent treatment and control groups in terms of student and classroom characteristics. However, given significant differences among a few variables including pretest differences, care was taken to include variables that differed across the treatment and control groups as covariates in the analyses of program effects. Specifically, the following covariates were identified for inclusion in the multilevel model of program effects: 1) classroom environment, 2) engagement in intervention activities, 3) school, and 4) pretest performance.

2. Statistical Power

The following assumptions were used to calculate the power to detect effects:

- Significance level (α) = 0.05;
- 47 clusters (classes) with an average class size of 20.
- Calculations were done both without and with a cluster covariate. Our prior research has shown that this value can range from 0.32 to 0.80. The power analysis with a moderate cluster-level covariate was set at 0.50.
- The calculations were done on a range of intra-class correlations. Research conducted by PRES Associates has shown that this value can range from 0.07 to 0.55. In addition, the What Works Clearinghouse has set a default value of 0.20 when adjusting statistics for clustering.

The *Optimal Design* software was used in the calculations in this section (Raudenbush et al., 2005). This program is designed to determine the power of longitudinal and multilevel research.

Figure A2 describes the power for a cluster randomized trial for a range of intra-class correlations *without* any cluster covariate for low, medium and high power (effect sizes corresponding to 0.2, 0.5, and 0.8 respectively). Figure A3 describes the power for a cluster randomized trial with a correlated cluster variable (r = 0.50). The key point from the graphics below is that there is enough power to reasonably detect a moderate to large effect size; however, there is not sufficient power to detect a small effect size of .20.

Figure A2. Power vs. Intra-Class Correlations for a Range of Effect Sizes (No Cluster-Level Covariate Included)

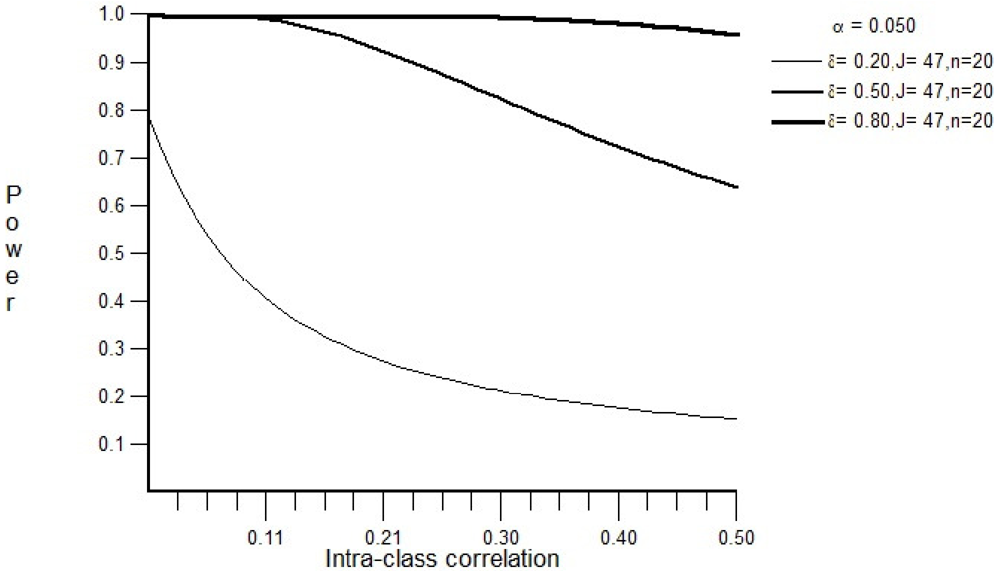
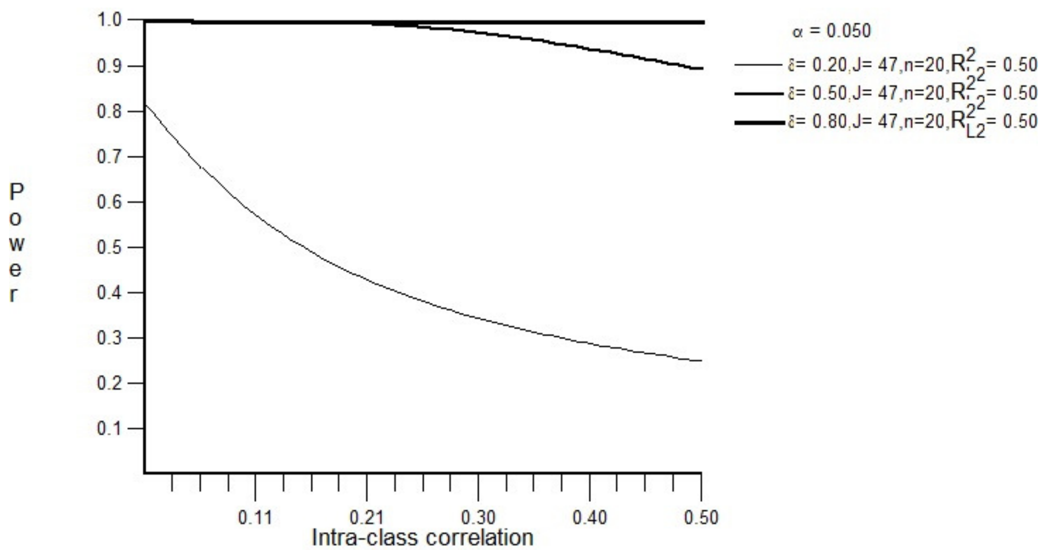


Figure A3. Power vs. Intra-Class Correlations for a Range of Effect Sizes (Cluster-Level Covariate Included)



Note: In figures A2 and A3, J refers to number of clusters, n refers to the average cluster size, δ refers to the effect size, α is the significance level, and r^2 is the correlation coefficient between the cluster-level covariate and the individual-level outcomes.

3. Attrition Analysis

As previously noted, both measurement attrition (i.e., missing data due to students not completing assessments) and dropout attrition (i.e., missing data due to students leaving the study) were examined. The approach taken in this project was to seek a consistent pattern of results of program effects across a range of methods. In this section, the observed pattern of differential attrition is examined to determine if it can explain the pattern of the observed results.

Dropout Attrition

Among students who were available to participate in both study years, there was an overall dropout attrition of 8.7% in Year 1 and 10.4% in Year 2 due to students leaving school. Of note, the RI school district underwent a a major reorganization for the 2012-13 school year and decided not to participate in Year 2 of the study as a result. These counts are excluded from the table below for Year 2. Analysis of dropout attrition during the 2011-12 school year found a significant relationship between attrition and group with more control students dropping out than treatment students. Unlike the first year of the study, during the second year of implementation, drop-out behavior was similar between groups and no significant relationship was observed. However, further analyses of the students constituting Year 2 participants showed a significant relationship. Specifically, there were a higher percentage of new students in the control group (39%) than treatment group (30%), and conversely, there was a higher percentage of students who participated in both years of the study in the treatment condition (70.3%) than control condition (61%). The high number of "new year" students can be attributed to the fact that data was not collected on all students within science classes in Year 1 (only randomly selected classes were included in the study). While schools were asked to try to move all Year 1 study students to participating study classes in Year 2, this did not always occur and so study classes contained a mix of "new" students and students from Year 1.

Table A2. Number of Students by Enrollment Status

		Students			
		Control	**Treatment**	**Total**	**Chi-square**
Year 1 of RCT	**Total students enrolled in Fall 2011**	**435 (100.0%)**	**602 (100.0%)**	**1037 (100%)**	
	Left in 2011-12 School Year	52 (12.0%)	38 (6.3%)	90 (8.7%)	$\chi 2$ (1)= 10.14, p =.001
	Enrolled at end of Spring 2012 *(Year 1 Participants)*	383 (88.0%)	564 (93.7%)	947 (91.3%)	
Year 2 of RCT	**Total students enrolled in Fall 2012**	**295 (100%)**	**348 (100%)**	**643***[27] **(100%)**	
	Left during 2012-13 school year	35 (11.9%)	32 (9.2%)	67 (10.4%)	$\chi 2$ (1)=1.22, p =.270
	Total Enrolled at end of Spring 2013	**263 (79.5%)**	**313 (76.2%)**	**576 (77.7%)**	
	New Students in 2012-13 school year	102 (38.8%)	93 (29.7%)	195 (33.9%)	$\chi 2$ (1)=5.25, p =.02
	Students from Year 1 of the Study *(participated both years)*	161 (61.2%)	220 (70.3%)	381 (66.1%)	

Not all students who left the study dropped out in a manner described above. Some simply changed conditions (i.e., students who were in Science Fusion program in year 1 were placed in a class that was in the control condition, or vice versa). Table A3 shows the number (and

[27] There were 499 students lost due to School D dropping out.

percent) of students who changed conditions and those who remained in the same condition during both years by group. Results showed a significant relationship in the proportion of students who changed conditions. Namely, a higher proportion of students who were in the control condition in year 1 were transferred to a treatment class in year 2 (21.5%) as compared to treatment students moving to the control condition (13.5%). These students were excluded from the main outcome analyses using the sample of students participating in both years of the study.

Table A3. Number of Students Changing Conditions*

	Students		
	Control	Treatment	Total
Students from Year 1 of the Study *(participated both years)*	161 (100%)	220 (100%)	381 (100%)
Students who stayed in the same condition during both years of the study	124 (78.5%)	193 (86.5%)	317 (83.2%)
Students who changed conditions	34 (21.5%)	30 (13.5%)	64 (16.8%)

*$\chi 2$ (1)=4.31, p =.04

Additional analyses were performed to examine whether baseline performance differences existed between students who remained in the study and those who left or changed conditions, and group assignment (see Tables A4-A6 for more information). Of interest in these ANOVAs were the interactions of group assignment and attrition status and the main effect for attrition status. A significant interaction would indicate a threat to internal validity. Similarly, a main effect for attrition status would suggest a threat to external validity.

For Year 1 of the study, examination of the *interactions* showed no significant group by attrition status interaction on science skills. However, a main effect for attrition was observed on the ITBS pretest score (it was marginal for the Developed Science Test). Those who left on average had lower test scores that those who remained, see Table A4.

Table A4. ANOVA Results for Pre-Tests by Group and Attrition Status for 2011-12 School Year

Measure	Attrition Status	Group	N	Mean	Sd.	ANOVA for interaction	ANOVA for main effect
ITBS-Scale Score	Attrition	Control	52	194.06	27.09	F(1, 970)=0.205, p=0.65	F(1, 970)=13.08, p<0.001
		Treatment	33	199.24	23.65		
	No change	Control	368	208.98	30.12		
		Treatment	521	210.84	33.25		
Developed Science Test	Attrition	Control	37	22.54	8.44	F(1, 948)=0.133, p=0.72	F(1, 948)=3.29, p=0.07
		Treatment	34	25.76	10.52		
	No change	Control	364	25.81	12.23		
		Treatment	513	27.94	12.38		

For Year 2 of the study, examination of the interactions nor main effects showed no significant group by attrition status interaction on science skills, see Table A5.

Table A5. ANOVA Results for Pre-Tests by Group and Attrition Status for 2012-13 School Year

Measure	Attrition Status	Group	N	Mean	Sd.	ANOVA for interaction	ANOVA for main effect
ITBS-Scale Score	Attrition	Control	31	216.34	24.50	F(1, 579)=0.763, *p*=0.38	F(1,579)=0.025, *p*=.875
		Treatment	24	207.67	28.77		
	No change	Control	234	211.76	30.08		
		Treatment	290	204.48	32.80		
Developed Science Test	Attrition	Control	28	30.50	9.70	F(1, 575)=0.009, *p*=0.92	F(1,575)=0.004, *p*=0.95
		Treatment	27	31.78	12.25		
	No change	Control	229	30.44	12.53		
		Treatment	291	31.51	11.53		

There were significant interactions and main effects observed with respect to students who changed conditions and group, however. Students who changed conditions tended to have higher ITBS pre-scores (main effect). In addition, treatment students who changed conditions tended to have higher baseline scores than control students who changed conditions. In contrast, treatment students who remained in the study in both years and did not change conditions had lower baseline scores than control students. Therefore, any significant effects observed will have occurred despite the fact that treatment students had lower baseline scores.

Table A6. ANOVA Results for Pre-Tests by Group and Second Year Group Status Changes

Measure	Attrition Status	Group	N	Mean	Sd.	ANOVA for interaction	ANOVA for main effect
ITBS-Scale Score	Changed Status	Control	33	211.73	21.50	F(1, 357)=6.77, *p*=0.01	F(1, 357)=7.91, *p*=0.005
		Treatment	26	223.04	32.29		
	No change	Control	116	210.82	29.82		
		Treatment	182	199.64	31.10		
Developed Science Test	Changed Status	Control	33	26.30	10.60	F(1, 355)=9.66, *p*=0.002	F(1, 355)=0.95, *p*=0.33
		Treatment	25	33.52	12.72		
	No change	Control	119	30.02	12.52		
		Treatment	178	26.40	11.78		

In summary, while there was no evidence for differential dropout attrition, there was some evidence for baseline performance differences among those who changed conditions (while not technically "dropouts," these students were dropped from analyses). Treatment students who remained in the study in both years and did not change conditions had lower baseline scores than control students. As such, the threat is not in favor of the treatment group and any significant differences would have occurred despite having lower performing students in the treatment group. Indeed, outcome analyses controlled for pretest scores.

Measurement Attrition

Among Year 2 study participants[28], a small portion of the students did not have data available at pre or post test due to absences on test administration days or because the teacher did not administer assessments[29]. Table A7 lists the number (and percent) of students who were in the study throughout the school year but did not provide pre or post tests. Chi-square analyses showed two significant relationships. Specifically, there were more treatment students who did not take the Developed Science Test Spring 2013 posttest *and* more control students who did not complete the ITBS Spring 2013 posttest.

Furthermore, to examine if there were any *performance* differences between those who completed tests and those that did not by group, ANOVAs were run on the post-test measures (to examine those not providing pretest measures) and on pretest measures (to examine those not providing posttest measures). Significant interactions between measurement attrition status and group assignment would suggest a bias. Results showed no significant interactions on the ITBS and Developed Science tests. Thus, results are not likely to be biased due to measurement attrition.

Table A7. Number of Students Who Did Not Provide Pre and Post Data

	Admin Time	N (%) Who Did Not Take Test			Chi-Square	ANOVA for interaction
		Control	Science Fusion	Total		
ITBS Scale Score	Pre Fall 2011 (N=538)	20 (3.5%)	18 (3.1%)	38 (6.6%)	$\chi2(1)=0.92$, $p=0.34$	F (1, 524)=0.14, p=.71
	Post –Spring 2013 (N=453)	81 (13.9%)	65 (11.3%)	146 (25.3%)	$\chi2(1)=19.26$, $p<0.001$	F (1, 453)=3.34, p=.07
Developed Science Test	Pre Fall 2012 (N=520)	31 (5.4%)	25 (4.3%)	56 (9.7%)	$\chi2(1)=2.82$, $p=0.11$	F (1, 520)=0.09, p=.77
	Post – Spring 2013 (N=398)	50 (8.7%)	128 (22.2%)	178 (30.9%)	$\chi2(1)=30.24$, $p<0.001$	F (1, 398)=3.18, p=.08

In summary, there was some evidence for dropout attrition. A higher proportion of students who were in the control condition in year 1 were transferred to a treatment class in year 2 (21.5%) as compared to treatment students moving to the control condition (13.5%). As well, treatment students who remained in the study in both years and did not change conditions had lower baseline scores than control students. As such, the threat is not in favor of the treatment group and any significant differences would have occurred despite having lower performing students in the treatment group. There was also no evidence of bias due to measurement attrition since there were no significant performance differences between those who completed tests and those that did not by group.

[28] Measurement attrition results from the Year 1 sample is available in the Science Fusion Year 1 Final Report.

[29] School C had two teachers who failed to administer the post ITBS or Developed Science Test. While researchers repeatedly contacted the teachers and school liaison about administration, no data was provided.

4. Statistical Analysis of Outcomes Measures

Analysis of Growth among Treatment Students

Paired t-tests for Change from Pretest to Posttest

Table A8 presents the means obtained for treatment students using Science Fusion at pre- and posttest as measured by the ITBS and Developed Science tests. Paired sample t-tests were conducted to examine whether there was significant change from pretest to posttest. Results showed significant growth (i.e., improvement in performance) on all outcome measures, with the exception of Year 1 Developed Science Test scientific reasoning items (note: Physical Science was marginally significant). However, this analysis is only intended to be descriptive.

Table A8. Pre-Post Scores for Treatment Students (Paired Sample t-test Results)

Test	Time	Mean	Std. Deviation	N	t	df	Sig.
ITBS Overall Science Scale Score	Pre – Fall 2011	200.79	31.78	144	Flinear (1, 143) = 167.17, p<.001 Fquadratic (1, 143) = 0.25, p=.62		
	Post – Spring 2012	220.84	37.44	144			
	Post – Spring 2013	238.15	36.30	144			
ITBS Scientific Inquiry SS	Pre – Fall 2011	30.40	16.85	146	-8.315	145	.000
	Post – Spring 2013	45.55	21.03	146			
ITBS Life Science SS	Pre – Fall 2011	32.29	18.62	146	-4.454	145	.000
	Post – Spring 2013	41.10	21.89	146			
ITBS Earth Science SS	Pre – Fall 2011	26.37	22.19	146	-5.975	145	.000
	Post – Spring 2013	39.58	23.75	146			
ITBS Physical Science SS	Pre – Fall 2011	41.46	23.54	146	-1.838	145	.068
	Post – Spring 2013	45.66	22.49	146			
Developed Science Test Overall	Pre – Fall 2011	26.33	11.76	166	-11.053	165	.000
	Post – Spring 2012	37.04	14.37	166			
	Pre – Fall 2012	30.95	11.21	130	-10.285	129	.000
	Post– Spring 2013	44.83	14.41	130			
Developed Science Test: Vocabulary (Fill in Blank)	Pre – Fall 2011	27.25	19.47	153	-8.637	152	.000
	Post – Spring 2012	44.25	25.02	153			
	Pre – Fall 2012	35.67	22.42	127	-9.107	126	.000
	Post– Spring 2013	58.03	23.97	127			
Developed Science Test: Science Application & Reasoning (Short Answer)	Pre – Fall 2011	16.85	15.74	124	-.928	123	.355
	Post – Spring 2012	18.55	17.47	124			
	Pre – Fall 2012	11.15	11.00	122	-11.506	121	.000
	Post– Spring 2013	31.15	20.17	122			

Growth Analysis of Subgroups of Treatment Students

Exploratory analysis was also performed to examine the relationship between Science Fusion and subgroup performance. That is, the results summarized in this section deal with the performance among treatment students only. Analyses were performed for the following subgroup categories: gender, free/reduced lunch status, grade level, and students at various science levels. Due to the more limited sample size available for analyses for students participating in both years of the study (n=317), these analyses focused on students participating in Year 2 of the study (including new students in 2012-13, n=576). As a reminder, the Year 2 sample was both smaller and less diverse due to the loss of one large school during the second year of the study. As a result, these analyses do not include analysis by English Language Learner status, special education status, and ethnicity – there were insufficient numbers of students within these categories[30].

The accompanying tables (A9-A13) include the paired t-tests' results. Results showed that students in all subgroups showed significant gains on both the Developed Science Test and ITBS, with one exception. High level students showed a significant decline on the Developed Science Test.

<u>Gender</u>

Table A9. Paired t-test Results for Treatment Students by Gender

Test	Time	Mean	Std. Deviation	N	t	df	Sig.
Male							
ITBS Scale Score	Pre	201.34	33.70	92	-11.244	91	.000
	Post	238.62	35.37	92			
Developed Science Test	Pre	30.97	11.54	76	-9.701	75	.000
	Post	45.92	13.96	76			
Female							
ITBS Scale Score	Pre	206.05	33.30	105	-10.118	104	.000
	Post	240.16	36.36	105			
Developed Science Test	Pre	32.28	9.88	92	-7.831	91	.000
	Post	44.70	14.44	92			

[30] The reader is referred to the Science Fusion Year 1 Final Report for results from these subgroups participating in Year 1 of the RCT.

Grade Level

Table A10. Paired t-test Results for Treatment Students by Grade Level

Test	Time	Mean	Std. Deviation	N	t	df	Sig.
7^{th}							
ITBS Scale Score	Pre	201.22	33.04	131	-7.027	130	.000
	Post	228.16	38.18	131			
Developed Science Test	Pre	28.17	10.44	95	-6.808	94	.000
	Post	41.22	16.68	95			
8^{th}							
ITBS Scale Score	Pre	211.20	32.31	116	-9.136	115	.000
	Post	237.23	38.84	116			
Developed Science Test	Pre	35.25	9.38	83	-12.004	82	.000
	Post	47.42	11.92	83			

Free/Reduced Lunch Status

Table A11. Paired t-tests Results for Students by Free/Reduced Lunch Status

Test	Time	Mean	Std. Deviation	N	t	df	Sig.
Not Eligible for Free or Reduced Price Lunch							
ITBS Scale Score	Pre	206.58	33.92	88	-8.726	87	.000
	Post	239.57	38.28	88			
Developed Science Test	Pre	32.82	11.40	78	-8.091	77	.000
	Post	46.36	13.86	78			
Eligible for Free or Reduced Price Lunch							
ITBS Scale Score	Pre	202.98	30.68	93	-12.031	92	.000
	Post	241.43	33.22	93			
Developed Science Test	Pre	30.76	9.95	89	-9.454	88	.000
	Post	44.70	14.04	89			

Science Levels

Table A12. Paired t-test Results for Treatment Students by Science Skill Level at Pretest

Test	Time	Mean	Std. Deviation	N	t	df	Sig.
Low							
ITBS Scale Score	Pre	181.84	19.49	137	-11.802	136	.000
	Post	219.47	34.66	137			
Developed Science Test	Pre	29.80	10.79	88	-4.988	87	.000
	Post	38.80	15.02	88			
Average							
ITBS Scale Score	Pre	228.05	10.60	86	-3.114	85	.003
	Post	240.42	36.74	86			
Developed Science Test	Pre	32.72	9.63	64	-10.697	63	.000
	Post	48.06	12.25	64			
High							

Test	Time	Mean	Std. Deviation	N	t	df	Sig.
ITBS Scale Score	Pre	263.96	13.936	24	-2.570	23	.017
	Post	277.71	23.064	24			
Developed Science Test	Pre	36.50	10.874	20	3.489	19	.002
	Post	28.55	5.539	20			

Implementation Fidelity Levels

Table A13. Paired t-test Results for Treatment Students by Level of Implementation

Test	Time	Mean	Std. Deviation	N	t	df	Sig.
Low							
ITBS Scale Score	Pre	225.43	26.28	46	-8.936	45	.000
	Post	262.80	26.06	46			
Developed Science Test	Pre	34.96	10.37	50	-11.893	49	.000
	Post	49.64	12.09	50			
Moderate							
ITBS Scale Score	Pre	198.08	34.85	78	-2.483	77	.015
	Post	211.19	34.30	78			
Developed Science Test	Pre	28.00	11.91	40	-.277	39	.783
	Post	28.85	14.85	40			
High							
ITBS Scale Score	Pre	203.57	31.42	123	-10.394	122	.000
	Post	234.52	37.19	123			
Developed Science Test	Pre	31.07	9.44	88	-12.412	87	.000
	Post	47.91	11.56	88			

Analysis of Program Effects

Independent Sample t-tests

Table A14 describes the means for the treatment and control groups for the eight outcomes at post-testing in Spring 2013 among Year 2 study participants. Independent sample t-tests were conducted for each of the outcomes. Statistically significant differences in favor of the treatment group were obtained for the Developed Science Test, including the portion of the test measuring science vocabulary and scientific reasoning/application. However, these differences do not account for clustering. The multilevel models described below incorporate dependency issues described above as a result of the hierarchical nature of the data.

Table A14. Sample Size, Means, Standard Deviations, and *t*-test (Student Level) Results for Assessments at Post-testing

Test	Group	Mean	Std. Deviation	N	t	df	Sig.
ITBS Overall Science Scale Score	Control	231.80	36.07	184	.479	451	0.63
	SF	230.05	39.43	269			
ITBS Scientific Inquiry SS	Control	40.31	19.13	184	-.788	451	0.43
	SF	41.85	21.32	269			
ITBS Life Science SS	Control	39.50	20.78	184	.309	451	0.76
	SF	38.84	23.00	269			
ITBS Earth Science SS	Control	37.90	23.69	184	.939	451	0.35
	SF	35.77	23.81	269			
ITBS Physical Science SS	Control	38.02	20.12	184	-1.705	451	0.09
	SF	41.53	22.43	269			
Developed Science Test Overall	Control	18.24	7.46	210	-4.61	396	0.00*
	SF	21.76	7.75	188			
Developed Science Test: Vocabulary (Fill in Blank)	Control	46.56	28.35	209	-4.770	386	0.00*
	SF	59.61	25.05	179			
Developed Science Test: Science Application & Reasoning (Short Answer)	Control	17.27	16.16	209	-6.535	379	0.00*
	SF	29.19	19.42	172			

* = $p<.05$

Multilevel Models

Given observed baseline difference, two-level multilevel models were implemented to estimate program impacts. For the two-level model, the first level incorporates student level covariates and the second level includes class/teacher level covariates. The two-level model focuses on the *levels* of program outcomes at the follow-up using the baseline level of program outcomes as a control.

Separate multilevel models were run for each of the following outcomes:

- ITBS Overall Scale Score
- ITBS Life Science Percent Correct
- ITBS Earth Science Percent Correct
- ITBS Physical Science Percent Correct
- ITBS Scientific Inquiry Percent Correct
- Developed Science Test: Overall Percent Correct
- Developed Science Test: Vocabulary Percent Correct
- Developed Science Test: Science Applications/Reasoning Percent Correct

Student level covariates in the model included:

- Group (Treatment=1; Control=0)
- Pretest performance

Other individual level covariates including gender and free/reduced lunch status were also available. However, due to small sample sizes and/or extensive missing data for these variables, these covariates were excluded from the multilevel analysis as this would significantly reduce the analytical sample. Teacher/class level covariates included in the model include school, teacher perceptions of classroom environment, and teacher engagement in intervention activities.

Analyses were conducted on two samples: 1) all students participating in the second year of the study, including new students enrolled in the participating study classes (Year 2 Students), and 2) students who participated in both study years and remained in the same study condition throughout.

Tables A15 - A17 summarize the results of the two-level models. Note that each row in the tables corresponds to the program effect coefficient estimated for that dependent variable from a separate multilevel model. In Year 2, results showed three significant differences. Science Fusion students showed higher post-test scores on the Developed Science Test overall score and vocabulary items as compared to control students, and a marginally significant positive effect was obtained on the short answer items. Year 1 results from two-level models are provided in Table A17 for reference purposes.

Note that unlike the results presented in Table A14, these analyses control for the baseline differences between control and treatment groups. Given the pre-existing differences (see section "Establishing Group Equivalence"), it is critical to include covariates corresponding to the pretest scores and other important characteristics on which groups differed in order to equate the two groups. When this is done via multilevel modeling, significant differences are obtained as described above. The effect sizes are also calculated; the effect size for the effect of the program on Developed Science Test is moderate (.48 to .64).

Table A15. Main Program Effects from the Two-level Model: Year 2 Participants (N=576)

Outcome Measures	Coefficient	Std. Error	t-ratio	Sig. Level	Effect Size
ITBS Overall Science Scale Score	5.08	10.82	0.469	0.639	
ITBS Scientific Inquiry Percent Correct	4.17	5.38	0.775	0.439	
ITBS Earth Science Percent Correct	4.02	4.59	0.877	0.381	
ITBS Life Science Percent Correct	-2.73	5.79	-0.472	0.637	
ITBS Physical Science Percent Correct	-1.32	6.53	-0.202	0.840	
Developed Science Test Overall Percent Correct	4.36	1.76	2.485	0.013	0.57
Developed Science Test: Vocabulary	14.73	6.58	2.237	0.026	0.55
Developed Science Test: Science Application and Reasoning	8.49	4.54	1.872	0.062	0.48

Table A16. Main Program Effects from the Two-level Model: Participants in Both Study Years (N=317)

Outcome Measures	Coefficient	Std. Error	t-ratio	Sig. Level	Effect Size
ITBS Overall Science Scale Score	5.31	13.05	0.407	0.684	
ITBS Scientific Inquiry Percent Correct	4.17	6.40	0.652	0.515	
ITBS Earth Science Percent Correct	5.27	6.10	0.864	0.388	
ITBS Life Science Percent Correct	-2.40	7.18	-0.334	0.789	
ITBS Physical Science Percent Correct	0.12	7.90	0.015	0.988	
Developed Science Test Overall Percent Correct	9.09	4.11	2.213	0.028	0.64
Developed Science Test: Vocabulary	14.67	7.45	1.970	0.050	0.57
Developed Science Test: Science Application and Reasoning	8.77	5.85	1.500	0.135	

Table A17. Main Program Effects from the Two-level Model: Year 1 Participants (N=947)

Outcome Measures	Coefficient	Std. Error	t-ratio	Sig. Level	Effect Size
ITBS Overall Science Scale Score	2.14	5.32	0.402	0.69	
ITBS Scientific Inquiry Percent Correct	1.19	2.77	0.430	0.67	
ITBS Earth Science Percent Correct	0.74	2.73	0.271	0.79	
ITBS Life Science Percent Correct	-0.98	2.89	-0.34	0.74	
ITBS Physical Science Percent Correct	2.23	2.92	0.766	0.44	
Developed Science Test Overall Percent Correct	3.48	2.92	1.19	0.23	
Developed Science Test: Vocabulary	8.91	5.11	1.74	0.08	.32
Developed Science Test: Science Application and Reasoning	1.53	2.64	0.58	0.56	

Multilevel Models with Subgroup Effects

Subgroup effects were analyzed for variables that had a sample size of 40 or greater (i.e., gender, ability grouping, free/reduced lunch status, and grade). The main effects two-level multilevel models were re-specified to re-estimate program effects for the following subgroups: gender (*female),* free/reduced lunch status*,* grade, and Ability *(Low* and *Average* with *High* as reference category). The subgroup effects were obtained by adding the interaction term(s) corresponding to each subgroup separately to the main effects model.

It is important to view these analyses as exploratory for a number of reasons: (i) the treatment and control groups were not randomized by subgroups; (ii) the sample sizes for some subgroups are quite small; and (iii) differences were obtained between the treatment and control groups at baseline for some of the subgroups. Results showed no significant subgroup effects. That is, subgroup of Science Fusion and control students performed similarly.

Appendix B

Mathematical Details of Multilevel Models

The Structure of the Two-level Multilevel Model for Program Effects

The two-level multilevel model had the following structure (note that the variable names are described in the text):

Level-1 Model

$\text{Outcome}_{tij} = \psi_{0j} + \psi_{1j}*(GROUP_{mj}) + \psi_{2j}*(PRE_{mj}) + e_{mj}$

Level-2 Model

$\psi_{0j} = \gamma_{00} + \gamma_{01}*(SCHOOLA_j) + \gamma_{02}*(SCHOOLB_j) + \gamma_{03}*(E1\_MEAN_j) + \gamma_{04}*(E2\_MEAN_j) + u_{0j}$

$\psi_{1j} = \gamma_{10}$

$\psi_{2j} = \gamma_{20}$

Note that γ_{10} is a measure of program impact. E1_MEAN represents average score for classroom environment, and E2_MEAN represents average score for teacher engagement in intervention activities.

Effect Size

Following the guidelines set forth by the What Works Clearinghouse (2008), the effect sizes were calculated using the following formula:

> *Hedges's g for intervention effects estimated from HLM analyses is defined in a similar way to that based on student-level ANCOVA: adjusted group mean difference divided by unadjusted pooled within-group SD. Specifically,*

$$g = \frac{\gamma}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}}}$$

> *where γ is the HLM coefficient for the intervention's effect, which represents the group mean difference adjusted for both level-1 and level-2 covariates, if any; n_1 and n_2 are the student sample sizes, and S_1 and S_2 are the posttest student-level SDs for the intervention group and the comparison group, respectively.*

Appendix C:

Science Fusion Implementation Guidelines

HOUGHTON MIFFLIN HARCOURT

Science Fusion RCT Study

Implementation Guidelines

INTRODUCTION

Welcome and thank you for participating in the Randomized Control Trial being conducted by PRES Associates[31], on the Houghton Mifflin Harcourt *Science Fusion* program. We hope your experience with our study will be a rewarding one. Not only will you contribute to cutting edge research, but you will also benefit from targeted professional development provided by Houghton Mifflin Harcourt professional training specialists.

We realize that it can be challenging to change former teaching practices and implement a new science program. We understand that there may be associated obstacles and challenges with the beginning of implementation of any new program. For these reasons, we want and need to hear from you so that we can help guide you through any initial challenges you might encounter. In fact, it is critical that any problems encountered are addressed as soon as possible to ensure that this program is being implemented to its full potential. Feel free to contact PRES Associates via e-email at studies@presassociates.com if you have any questions, problems or concerns. We greatly appreciate the time and effort you will contribute towards making this study a success.

The following provides answers to some common questions teachers may have related to this study. Please read through all of these and should you have further questions, please contact PRES Associates.

WHY IS THIS RESEARCH BEING DONE?

As you are aware, the No Child Left Behind Act (NCLB) of 2001 requires that educational materials and strategies used by educators in the classroom *must be proven by scientific research to improve student achievement in the classroom.* Houghton Mifflin Harcourt has developed a strong research model for determining that their programs are scientifically based. As part of this research agenda, Houghton Mifflin Harcourt has contracted with PRES Associates, an external educational research firm, to conduct a randomized control trial (RCT) focused on a rigorous evaluation the effectiveness of the *Science Fusion* program in helping middle school students (grades 6-8) attain critical science skills.

[31] PRES Associates is an external, independent, educational research firm with an established track record in conducting large-scale, rigorous evaluations on the effectiveness of research materials.

WHAT ARE THE TRAININGS FOR?

It takes more than a good curricular program to provide effective and meaningful lessons in science. It also takes good teachers with a thorough understanding of the curriculum, who are supported by professional development, school administrators, and parents/guardians. To this end, it is hoped that through the professional development training session provided by Houghton Mifflin Harcourt on the use of its science program, all teachers participating in the study will gain the knowledge and skills to successfully implement this program fully from the start.

As you will soon learn, this science program provides numerous teaching resources and supports. In order to implement this program successfully, it is essential that teachers have a thorough understanding of the resources provided by the *Science Fusion* program. Rather than having teachers figure it out on their own, professional trainers will guide you through this process, offering examples of when to use certain materials, how to structure and pace classroom instruction, what types of assessments to administer, and so forth.

WHY DO I NEED TO FOLLOW THESE IMPLEMENTATION GUIDELINES?

Teacher Implementation Guidelines have been developed as part of this research study on *Science Fusion* in order to promote full and effective use of the program. The guidelines are being provided to teachers as a reference to draw from when implementing the new program in their class(es). Specifically, the *Science Fusion* implementation guidelines point out key program components that *must* be implemented during science instruction because they are integral to the program and have the greatest influence on student learning and performance. In addition, it is critical to ensure that all teachers are implementing a similar instructional model. That is, if teachers are modifying the program to an extent that it no longer resembles the original program, the research study will not provide accurate information on the effects of the *Science Fusion* program. In sum, by providing these implementation guidelines, we are attempting to (1) maximize the potential of this science program to help your students, and (2) ensure that the program is being implemented with fidelity across all teachers using the program. To reiterate, *it is essential that all teachers using the program fully apply the following implementation guidelines as prescribed.* That being said, there are optional parts to the program as well as ancillary resources that provide you with the flexibility you need to address unique student needs or contexts. *We trust your professional judgment and ask that you try to implement the program as best you possibly can while meeting your students' instructional needs.*

Again, thank you for your participation in this study. You are an integral part of this endeavor and we appreciate your assistance. We look forward to working with you.

Teaching the Unit

Items within the list below are organized according to the order they appear in your TE.

- **Items in bold below are critical core instructional activities that have been identified as necessary for optimal use of the *Science Fusion* program and as a study participant we will need you to incorporate these instructional activities into your science lessons.**
- *Items italicized below have been identified as important activities, but are not required for use as part of the study;* if you are able to incorporate them great, but if not, that's ok too.

Review the Unit Opener at the beginning of every unit paying close attention to the following items that will help in planning the unit:

- ✓ **The Big Idea and Essential Questions** – Introduce the Big Idea for each unit and Essential Question for every lesson that is essential for understanding the content.
- ✓ **Options for Instruction** – Each lesson includes both a print path and a digital path. While we ask that teachers fully use the Science Fusion program as described in these guidelines, the teacher is free to choose which path they want to use (print, digital, or a combination of both print and digital). While the print and digital paths cover the same content, please note that the format and organization of the information presented in the digital path differs significantly from the print version – this has been done so as to take advantage of interactive options available only in a digital environment.
- ✓ **Assessment Options**– It is very important that teachers conduct both formative and summative assessment(s) so as to monitor understanding of the material presented and adapt instruction as necessary. The *Science Fusion* program offers a multitude of assessment options to teachers, however, which assessment(s) are used for formative and summative purposes are not prescribed. It is a requirement, however, that formative assessment(s) occurs during lessons and that summative assessments occur at the end of Units, at minimum.
    - Formative Assessment – Strategies, Lesson Reviews, Unit Pre Test
    - Summative Assessment - Alternative Assessment, Lesson Quizzes, Unit Tests, Unit Review, Practice Tests
    - Project Based Assessment
- ✓ *Differentiated instruction* – Depending on the range of abilities or the difficulty of the content, teachers can choose to follow the Differentiated Instruction suggestions for modifying lessons or activities.
- ✓ *Content Refresher* – While this is considered imperative at the K-5 levels, at the middle school level it is an optional memory refresher on key topics found in every lesson.
- ✓ **Advance Planning** – set the stage for each unit by asking students "What do you think"
- ✓ Unit Project "Citizen Science"– Citizen Science needs only to be completed if the teacher chooses to do so.

Lesson Activities

Each lesson is structured around the "5 E" instructional model that includes "Engage," " Explore", "Explain", "Extend", "Evaluate." *Science Fusion* was designed to provide students with numerous opportunities to write about & reflect on the processes they use to make sense of new scientific concepts. Throughout the program, students should write in response to prompts that ask them to engage in various types of thinking & reflection. Review the Lesson Overview and lesson activities and select the activities you need to engage your students.

1. **Overview**
    - **Engage Your Brain**
    - **Active Reading**

2. **Engage and Explore**
    - **Activities and Discussion** – Activities and discussion are an integral part of the program and some activity/discussion is expected to occur during every lesson. Teacher discretion is allowed in selecting activities for their classes.
    - **Labs and Demos** – There are 2 Quick Labs per lesson & unit level comprehensive labs. Each of these longer labs offers 2 sets of student data sheets which are differentiated by 3 levels of inquiry. Labs are an integral part of this program and it is very important that they are done regularly. That said, teachers may use their professional judgment in deciding on which labs to do and how frequently.
        - Quick Lab
        - Daily Demo
        - Exploration Lab
        - Virtual Lab

3. Explain
    - **Key Topics** – Students should complete all the following sections of the Lessons as follows :
        - Active Reading
        - Visualize It!
        - Think Outside the Book
        - Do the Math
        - Predict/Infer/Identify
    - *Differentiated Instruction* – Differentiated instructional activities should be used based upon the professional judgment of teachers.
    - **Lesson Vocabulary** – It is very important that vocabulary is reviewed and reinforced in every lesson.

4. **Extend** – Depending on the needs of students or classes, teachers may choose to use some or all of the following extension activities keeping in mind the Lesson Quiz & Unit Tests that are part of the Summative Assessment are required.
   - *Reinforce and Review*— Provides extra support for reinforcement & extensions with cross-curricular activities as needed.
     - Activity
     - Graphic Organizer
     - Visual Summary
   - *Going Further*
     - *Why it matters?* – This is optional based upon teacher choice.
   - *Evaluate* – Provides options for formative & summative assessment, as noted, formative and summative assessment is essential to the program. It is anticipated that, at minimum, teachers will use the Lesson Quiz and Unit Test. Formative Assessment
     - Strategies - throughout the TE
     - Lesson Review - SE
   - **Summative Assessment**
     - Alternative Assessment
     - **Lesson Quiz**
     - **Unit Tests**
     - Unit Review – SE

5. *Other Activities*
   - *Think Science*
   - *STEM Engineering & Technology*

Appendix D:

Case Study of Site Visits

School A (OH)

About the School: School A is a charter school located in an older urban community in Ohio. The school consists of a renovated industrial building that houses students in grades 6-8. This building is also attached to the Elementary School that houses students in K-5. During the 2011-2012 school year, enrollment at School A was 350, with a student to teacher ratio of 19 to 1.

In 2013, Ohio used the Ohio Achievement Assessments to test students in grades 5 and 8 in science. The tests are standards-based, which means they measure how well students are mastering specific skills defined for each grade by the state of Ohio. Results show that 67% of 8th grade students at School A were proficient or above in science, which is lower than the state average of 69%. The student population is predominantly Black:

- 92% Black
- 4% Two or more races
- 2% Hispanic
- 1% Asian
- 1% White

Approximately 82% of the students at the school were eligible for free or reduced-price lunches, and 9% were classified as Students with Disabilities.

Study Participants: During year one, two teachers participated in the study: each teacher was a treatment and a control teacher due to the small school population. The 6th grade teacher taught three treatment class periods and two control class periods for a total of five classes (2 control and 3 treatment). At the 7th grade level, the teacher taught three treatment class periods and two control class periods for a total of five classes (2 control and 3 treatment). Thus, there were 10 participating study classes. The 10 classes contained approximately 257 students, with an average class size of 25.7, and a range of 20 to 31.

During year two, two teachers participated in the study: each teacher taught both treatment and control classes. At the 7th grade level, the teacher taught two treatment class periods and one control class period for a total of three classes (2 treatment, 1 control). In the 8th grade, the teacher taught two treatment classes and two control classes for a total of four classes (2 treatment, 2 control). Thus there were 7 participating study classes in year two. The 7 classes contained approximately 196 students, with an average class size of 28 and a range of 24 to 32.

In year two teachers characterized their classes as a mix of students. Both teachers had students in their classes with Individualized Education Plans and a few students that were designated as English Language Learners. Overall classes had a larger number of students of average ability mixed with students that were of higher and lower abilities. The 8th grade teacher had one class that was considered high with students of average to high ability and one class that was considered low with students of low to average ability. Overall classes were noted as typical of the student population.

Science Curriculum and Resources: The control program consisted of a 2005 middle grades science textbook. In general, the teachers used this basal program as their main science curriculum. There were a few similarities between the control program used and the Science Fusion program.

Similarities included opportunities for hands-on explorations and an emphasis on a big question to connect big ideas to the real world. Both programs also begin with an engaging section opener that connects prior knowledge and information rich visuals to connect to the text. In general the Science Fusion program focused more on the blending of hands-on, digital and print materials and in general emphasizes the students' development of 21st century skills. Teachers also noted that students were more actively in engaged by reading the write-in student text and liked the ability to have all their notes in one place.

In treatment classes, the teachers were observed following the Science Fusion program exclusively and mostly adhering to the implementation guidelines. Teachers stated that they were able to complete the various types of the lab activities and utilize the Science Fusion digital components.

Instructional Practices and Strategies: Science instruction occurred throughout the day (the study teachers only taught science). Classes lasted for 55 minute periods and occurred every day during the same time for the duration of the year. While all treatment students had sufficient copies of student science textbooks, the control students sometimes only received copied pages from the control textbook and not the actual textbook.

Science instruction in the control classroom was relatively consistent. The teachers would usually spend the first 5-10 minutes doing a Do Now activity, which included practice questions from the Ohio Achievement Assessment. The teachers would then poll the students for answers and briefly discuss the correct answer. The teachers would then do class notes or guided practice for the next 25 minutes then students would complete an independent activity or lab activity for 15-20 minutes. The teacher would provide students with a packet with notes for guided practice and independent practice exercises. The class would end with an exit ticket, whereby students would have to answer a question relating to the days lesson before they could leave class. Anything that wasn't finished in class was assigned as homework. Additional homework would be assigned which typically included a teacher created worksheet.

Lessons in the treatment classrooms were very similar in structure to the control class. Lessons started with the same "Do Now" activity while teacher took roll and checked homework. Depending on where they were within a module, they would begin with the big idea and essential questions. Teachers would provide guided practice or class notes and then move on to independent practice or a lab activity. During independent practice students would work independently to complete the active reading sections of the text as well as additional questions provided by the teacher. The teachers would also use the Digital Path activities projected on their SmartBoards to illustrate concepts during guided practice or would utilize the virtual labs in place of a hands-on lab at least once per week. Anything not completed in class was assigned as homework. Students would then answer the Exit Ticket questions before leaving class to demonstrate their understanding of the lesson. Homework would also be assigned on a daily basis and included the Lesson Review questions for the Science Fusion text.

Assessment: In terms of assessment practice there was very little variation between the control and treatment classes. Teachers administered the same teacher created vocabulary quiz on a weekly basis and a midterm and final every 6 weeks. The midterm and final was largely teacher created and incorporated old questions from the OAA assessment for test prep. Student's daily "do now" activity also utilized old questions from the OAA assessment or Buckle Down test prep book. Informal assessment (i.e. observation, checking homework, discussion, etc.), also occurred with equal regularity and in similar ways in all classes.

Comparability: In terms of overall comparability, both the Science Fusion and control classrooms were similar. For example, teachers created daily lesson packets of questions and activities that were similar for both treatment and control classes. The only difference in the packet was that for treatment students this also included the completion of active reading sections and questions from the lesson review in the text. The sequencing of materials stayed relatively constant as the teachers adhered to the school curriculum map. However, control students had fewer opportunities for hands on lab activities (only occurred monthly), while treatment students completed a hands on lab activity at least once a week. Among the participating teachers' classes, no contamination was noted and student engagement and interest was average.

School B (OH)

About the School: School B is a public middle school located in an suburban residential community in Ohio. The school consists of an older building that houses students in grades 7-8. During the 2011-2012 school year, enrollment at School B was 627 with a student to teacher ratio of 14 to 1.

In 2013, Ohio used the Ohio Achievement Assessments to test students in grades 5 and 8 in science. The tests are standards-based, which means they measure how well students are mastering specific skills defined for each grade by the state of Ohio. Results show that 39% of 8th grade students at School B were proficient or above in science, which is lower than the state average of 69%. The student population is predominantly Black:

- 81% Black
- 12% White
- 5% Two or more races
- 2% Hispanic

Approximately 61% of the students at the school were eligible for free or reduced-price lunches, and 17% were classified as Students with Disabilities.

Study Participants: In year one, two teachers participated in the study: one teacher was treatment and one teacher was control due to the small school population. At the 7th grade level, the treatment teacher taught two class periods and the control teacher taught three class periods a total of five classes (3 control and 2 treatment). Thus, there were 5 participating study classes. The 5 classes contained approximately 101 students, with an average class size of 20.2, and a range of 17 to 27.

In Year 2, two teachers participated in the study: one teacher was treatment and one teacher was control due to the small school population. At the 8th grade level the treatment teacher taught two class periods and the control teacher taught three class periods for a total of five classes (2 treatment, 3 control). The 5 classes contained 112 students, with an average class size of 22.4, and a range of 14 to 29.

In Year 2, both teachers had classes that they characterized as mixed for the most part. Most students in the treatment classes were considered average with some students of above or below average abilities. The control teacher had one class that was considered advanced with 90% of the students classified as above average but also had one class that was an inclusion class with lower performing students and students with Individualized Education Plans. All classes were noted as typical of the student population.

Science Curriculum and Resources: The control program consisted of a 2005 middle grades science textbook. In general, the teacher used this basal program as the main science curriculum only supplementing with teacher created resources where needed. There were a few similarities between the control program used and the Science Fusion program. Similarities included opportunities for inquiry and built-in lesson checks throughout the lesson and at the end. Both programs also include connections to other subject areas (e.g., Math and Language Arts) and an emphasis on technology. In general the Science Fusion program focused more on big ideas and overarching themes and incorporated more opportunities for hands on and digital lesson support.

In treatment classes, the teacher was observed following the Science Fusion program exclusively and adhering to the implementation guidelines. The treatment teacher stated that had used all of the Science Fusion components, including print, digital, and hands-on activities.

Instructional Practices and Strategies: Science instruction occurred throughout the day (the study teachers only taught science). Classes lasted for 80 minute blocks and occurred every other day during the same time for the duration of the year. All students had sufficient copies of student science textbooks.

Science instruction in the control classroom was relatively consistent. The teachers would usually spend the first 5-10 minutes doing a bell ringer activity, which included practice questions from the Ohio Achievement Assessment. The teacher would then poll the students for answers and briefly discuss the correct answer. Next the teacher would review homework and complete a binder check. Then students would either take notes during a whole group lecture for 30 minutes or the teacher would introduce the lab activity. Next the students would complete the lab activity, watch a teacher demonstration or watch a science related video for 20 minutes. The teacher would then provide a lesson wrap up and have students complete a written response on what they learned in class that day. The control teacher would assign homework every two weeks which included a review of content questions.

Lessons in the treatment classrooms varied somewhat in structure to the control class. Lessons typically started with a 10 minute review of the previous day's lesson followed by a discussion or reading from the Science Fusion worktext. The teacher would then assign the class an independent practice activity that generally involved completing the activities and discussion section of the Science Fusion lesson or a lab activity. The teacher would then wrap up the days lesson with a brief review and assign homework. Anything unfinished in class was also assigned as homework.

Assessment*:* In terms of assessment practice there was very little variation between the control and treatment classes. Informal assessment (i.e. observation, checking homework, discussion, etc.) lesson quizzes and unit tests occurred with equal regularity and in similar ways in all classes; the main difference between the treatment and control classes was in the materials used for assessment. For treatment students, the teachers used Science Fusion program materials such as the lesson quiz, and custom unit tests from ExamView. For control classes, the teacher created their own lesson quiz and chapter tests based on the control textbook.

Comparability*:* In terms of overall comparability, both the Science Fusion and the control classrooms were similar. For example, science vocabulary and science facts/concepts were presented in both treatment and control classes and students in both treatment and control were taught the same concepts, although the sequence and materials used were different. Both classrooms emphasized hands-on and inquiry based scientific activities and included opportunities for group/partner work as well as individual practice. However, students in control classrooms had fewer opportunities for digital lesson support. Among the participating teachers' classes, no contamination was noted and student engagement and interest was average.

School C (DC)

About the School: School C is a public 6-9 charter school located in an urban neighborhood in Washington D.C. The school is housed in an older building that has undergone some renovations. The school offers a comprehensive curriculum that is aligned with the District of Columbia Public School Standards. ANet assessments are given in core subjects at regular intervals to measure growth and level of student mastery. In addition to their academic program, the school also offers a wide variety of elective courses in the areas of Technology, Band, Dance and Drama, Art, French, and Spanish.

During the 2011-2012 school year, enrollment at School C was 667 and during the 2012-2013 enrollment was 700. D.C. Public Charter Schools administer the DC-CAS to students in grades 1 to 11 each spring; however this test only measures reading and math. Therefore, school-wide science performance results are unavailable. The student population is predominantly African American:

- 93% African American
- 7% Hispanic

No data was available regarding number of students noted as Limited English Proficiency or eligible for free or reduced-price lunches.

Study Participants: In year one there were four participating teachers. Eight classes at School C participated in the study: three 6th grade treatment classes, one 6th grade control class, two 7th grade treatment classes, and two 7th grade control classes. These classes contained approximately 145 students, with an average class size of 22.

In year two there were three participating teachers, one 7th grade teacher, one 8th grade teacher and one teacher who taught 7th and 8th grades. Fifteen classes at School C participated in the study: five 7th grade treatment classes, two 7th grade control classes, three 8th grade treatment classes, and five 8th grade control classes. These classes contained approximately 331 students, with an average class size of 22 and a range of 17 to 26. During both years of the study classes were representative of the general student population.

During the course of the study technology was not emphasized in either year one or year two due to lack of technological resources or not finding the HMH Science Fusion technology resources useful. In year one some attempts to incorporate the Science Fusion technology were evident by the 6th grade treatment teacher, but was more rarely used by the 7th grade teacher. In year two more of an effort was made to incorporate technology, mostly in the form of videos, with the exception of the 7th grade teacher, who regularly incorporated technology into her science classes.

Science Curriculum and Resources: At School C control classes in both years of the study had access to two basal programs. One program (a 2006 edition) consisted of a modular chapter based arrangement of lessons built around big ideas and hands on learning activities that reinforced key concepts. The program emphasized a research-based approach to learning by connecting big ideas to the real world. Each lesson began with an engaging section opener that connected new learning to prior knowledge. Lessons included information rich visuals that connected to the text and supported student learning. The program also included built in assessment activities with student self-checks for comprehension and built in vocabulary activities.

Similarly the second control program (a 2001 edition) used a modular chapter based arrangement of lessons that included lab activities and opportunities for self-assessment. The program emphasized a connection to other content areas of science lending to a greater understanding of science in real world contexts. Each chapter began with a full length Lab Investigation activity to introduce the topic through a hands-on experience. Each lesson included a quick lab activity, math activities that integrated math and science and an “Apply” feature to connect student knowledge to the real world. The program also included feature articles following every chapter emphasizing Science and Technology.

In both years control teachers reported using these programs to plan instruction and for occasional student reading; however they supplemented these programs with other resources, such as content from websites (NASA website, NOAA, and ScienceSpot, etc.) and teacher found or created materials, while following district standards for planning and pacing guides.

In treatment classes during year one, one teacher was observed following the Science Fusion program almost exclusively and mostly adhering to the implementation guidelines, with the exception of not incorporating all lesson activities to a high degree. The other treatment teacher was noted as a low implementer due to lack of implementation of the several key program components such as the Big Idea, Essential Questions, lesson activities such as Think Outside the Box, Do the Math, lesson quizzes, and Science Fusion labs. Similarly in year two one teacher was observed following the Science Fusion program to a high degree, while the other treatment teacher was not a high implementer due to a consistent lack of key component usage.

Instructional Practices and Strategies: During both years science instruction occurred throughout the day (the study teachers only taught science). Classes lasted for 50 minute periods and occurred every day during the same time for the duration of the year. All students had sufficient copies of student science textbooks.

In year one and two science instruction in the control classrooms was similar across both teachers. Typically, classes began with a 5 minute Do Now as students copy homework assignment down. New material or a new activity (hands-on or group work) would then occur for 10-15 minutes, followed by a class discussion of the lesson's objectives. One teacher also reported showing up to 5 minutes of online videos toward the end of the period to keep students engaged. At the end of the class, homework was assigned. In-class independent work was assigned 4-5 times a week, with most work coming from the textbook, internet, teacher-created resources, or students' science journals. Homework was assigned 4-5 days a week with work coming from the student text book end of chapter reviews or teacher-created worksheets. With respect to labs, "Thought Labs" occurred every two weeks (frequency of physical labs varied depending on topics covered). It should be noted in year two one control teacher made a much greater effort to include more hands lab activities and teacher demonstrated mini labs.

In both years of the study lessons in the treatment classrooms varied somewhat as compared to the control classes. Lessons started with the brief Do Now activity which included writing of the homework assignment. Depending on where they were within a module, they would read from the student book with teacher modeling and whole-group discussion. This was followed by guided practice (e.g., Visualize It) and other lesson activities noted in the student book. Labs were typically held during their extended science class days (Tuesday and Wednesday). Lessons typically concluded with a short summary of the lessons objectives followed by a short 5 minute end-of-lesson "exit ticket" work. Homework was typically assigned every night except Fridays (SF activities such as Engage Your Brain, Visualize It, Active Reading, Extend).

Assessment*:* In terms of assessment practice there was very little variation between the control and treatment classes in either year of the study. Informal assessment (i.e. observation, checking homework, discussion, etc.) and unit tests occurred with equal regularity and in similar ways in all classes; in fact, the unit test (given every 9 weeks) were the same across all study classes (treatment and control) as they were required by the school. These tests included a mix of matching/multiple-choice, short answer, and essay questions. Typically, Fridays involved some form of assessment (pencil-paper quizzes, informal student presentations, or lab work). During year two 8th grade students were also assessed on a final group research project and related presentation.

***Comparability**:* In terms of overall comparability, both the Science Fusion and the control classrooms were similar. For example, science vocabulary and science facts/concepts were presented in both treatment and control classes and students in both treatment and control were taught the similar concepts, although the sequence and materials used were different. There were some disparities in the amount of lab activities that students engaged in, specifically treatment students had more lab opportunities (especially during year one in the 6th grade treatment classes) and engaged in more lesson activities as presented in the Science Fusion program than did their control counterparts. Among the participating teacher's classes, no contamination was noted and student engagement and interest was average. This was concurrent for both years of the study.

School D (RI) – YEAR ONE ONLY

About the Schools: School D is a large public school located in a low to middle class, suburban community in Rhode Island. The middle school is housed in two newer buildings. The school houses students in grades 6-8. During the 2011-2012 school year, enrollment at School D was 1336, with a student to teacher ratio of 11 to 1, which is somewhat higher than the state average of 13 to 1.

The student population is predominantly white, but still ethnically diverse.

- 54% White, not Hispanic
- 27% Hispanic
- <1% American Indian/Alaskan Native
- 11% Black, not Hispanic
- 8% Asian/Pacific Islander

Approximately 72% of the students at the school were eligible for free or reduced-price lunches, which is significantly larger than the state average of 40%.

Study Participants: During year one, ten teachers participated in the study. There were five teachers teaching 6th grade and three teachers teaching 7th grade. There was a total of sixteen 6th grade classes and eight 7th grade classes participating in the study, ten control and fourteen treatment. The 24 classes contained approximately 442 students, with an average class size of 26, and a range of 10 to 28. **<u>This school did not participate in Year 2 of the study.</u>**

For the most part teachers characterized their classes as mixed, with some classes being high performing and others being average or low performing, with a few exceptions. One treatment and one control teacher taught only classes consisting of lower performing and/or special education students with behavioral issues. Overall classes were noted as typical of the student population.

Technology was not overly emphasized in any of the classes, control or treatment, but both control and treatment teachers had students used technology in the form of laptops to research topics or work on special projects and to show science related movies.

Science Curriculum and Resources: Control teachers had access to a 2002 science program. This program consists of a magazine style text that contains numerous nonfiction readings designed to present scientific ideas in a unique way to expand student understanding. Unlike the other control programs, this program is organized in a magazine style arrangement with stories and articles relating to the overall topic. The program includes links to other subject areas such as language arts and history and class investigative activities. Lessons also include visual learning activities and built in learning checks for review. Teachers at School D used this program as their main science curriculum, although they supplemented regularly with teacher-made, commercial, and online resources.

The majority of the teachers paced their classes (treatment and control) based on the curriculum map, 18-week planner and student needs with two exceptions during year one. In treatment classes, teachers were observed somewhat following the Houghton Mifflin Science Fusion program and did not adhere to the implementation guidelines with fidelity, although one teacher used the Houghton Mifflin Science Fusion program exclusively with a moderately high level of implementation.

Instructional Practices and Strategies: Science instruction occurred throughout the day depending on the teacher. Classes lasted for either 80-minute blocks or 45-minute classes. Classes occurred every day during the same time for 45-minute classes and every other day for block classes. All students had sufficient copies of science resources (e.g., student textbooks or worksheets) and the school's technology resources were able to support online instruction in most classrooms, although teachers did not necessarily utilize.

Science instruction in control classrooms was relatively consistent. Overall teachers would begin instruction with a brief review of the previous day's lesson or accept questions related to homework. This was followed by an introduction to the new lesson or a brief review if the lesson would be a continuation of the previous day's lesson. There was one exception to this; one control teacher started each lesson with a "share and review" type activity where students would pair up and trade answers on homework or share their thoughts on a science related question the teacher would pose. Science lessons usually included a combination of lecture, labs or other hands on activities and review. One teacher had students work on two long-term projects over the course of the year and time was allotted each day for them to work on some aspect of the project. All control classes included the opportunity for students to work in pairs or small groups on a regular basis, although one teacher focused slightly more on individual work than the other teachers. Most teachers allowed students to self-select partners and groups, while on occasion assigned partners and group members based on abilities or for behavioral reason.

In the treatment classrooms lessons were somewhat similar to one another, in part because to some degree, most participating teachers were attempting to follow the implementation guidelines and prescribed pacing and therefore the structure of the treatment lessons were somewhat similar. However, because treatment teachers at the site implemented with less fidelity than teachers at other study sites, treatment teachers tended to fall back into the structure of the program they used previously and therefore treatment and control lessons were similar to some degree. Lessons started with the teacher going over the previous day's homework/lesson and answering related questions and/or introducing the new lesson. Most teachers posed an "essential question" to hook students and get them engaged, though they did not always utilize the Science Fusion material to do this. This was followed by a lecture or introduction if the lesson was new and included time for discussion and student questions. If the lesson was a continuation of the previous day, the teacher would briefly review the concept and then have students continue the previous day's reading, activity, lab etc. or they would have students jump right into continuing lab work if it hadn't been finished the previous day, with one exception. One teacher tended to spend at least several minutes each day reviewing, regardless of whether or not it was a new lesson or a continuation of the previous day. Most treatment teachers allowed students to use laptops to access the internet for research purposes to compliment or support the concepts they were learning.

***Homework**:* There was a great deal of consistency in terms of the amount of homework that treatment and control teachers assigned; however the types of homework differed. Teachers assigned homework 3-4 nights per week or assigned a packet at the beginning of the week that included the equivalent of 3-4 nights per week of homework. About half the control teachers and the majority of treatment teachers assigned homework that included reinforcement worksheets and science vocabulary. There was one control teacher that did not include the amount of vocabulary related work as the other teachers. Also, more control than treatment teachers assigned homework that included research related to ongoing science projects. Across all classes, teachers did not assign homework on Fridays and considered review or test preparation as the homework the night prior to a test.

Assessment: In terms of assessment practices there was minimal disparity between control and treatment classes. Informal assessment (i.e. observation, checking homework, discussion and other informal monitoring activities) and unit tests occurred with equal regularity and in similar ways in all classes. Assessments included a mix of matching/multiple-choice, short answer, and essay questions. On Fridays some form of assessment was given in the form of short quizzes, a unit test or short answer responses related to the concepts covered during the week. Control teachers tended to use work produced during labs as a form of assessment more often than their treatment counterparts.

***Comparability**:* In terms of comparability, both the Houghton Mifflin Harcourt Science Fusion and the control classrooms were similar overall, especially since implementation in treatment classes was low at this site. For example, science vocabulary, review and labs were relatively equally emphasized in both types of classes. In addition, both types of classes incorporated time for group work and discussion. However in treatment classrooms, vocabulary and short labs were emphasized on a more regular basis.

Appendix E:

Key Features and Resources for Treatment and Control Programs

Table D1. Program Features and Pedagogy of Treatment and Control Programs

	Science Fusion	Control Program 1: (2005, 2006)	Control Program 2: (2001)	Control Program 3: (2005)	Control Program 4: (2002) Year 1 Only
Key Program Features and Pedagogy	▪ Enduring Understanding framework ▪ "Big Ideas" serve as the overarching concept for each unit. ▪ Essential Questions identify the conceptual focus of the lesson. ▪ Lesson plans with 5E-based learning experiences organized by schema theory. ▪ STEM lessons, activities and Video Based Projects that incorporate and develop 21st Century Skills. ▪ Lab Activities integrated into each ▪ lesson ▪ Organized into 11 modules covering Life, Earth, Physical and Science and Technology ▪ Modules contain 2-5 units broken down into anywhere from 2 to 7 lessons. ▪ Lessons typically consist of the following elements: ○ Options for Instruction for the lesson ○ Engaging lesson opener to assess prior knowledge. ○ Probing Questions to build inquiry skills ○ Active Reading prompts teach students how to analyze and interact with content ○ Vocabulary development ○ Visualize It! make abstract concepts more concrete ○ Do the Math! ○ Think Outside the Book extends learning ○ Variety of lab activities that address a variety of student levels, inquiry levels — directed, guided, and independent ○ Ongoing Formative Assessment strategies to check comprehension	▪ Standards-based instruction ▪ Research based instructional strategies, including a focus on big ideas ▪ Connects the Big Idea to the real world through engaging photos and questions ▪ Lessons typically consist of the following elements: ○ Engaging section opener that connects new learning to prior knowledge ○ Information-rich visuals that connect to the text and support student learning. ○ Hands on learning activities that reinforce key concepts ○ Vocabulary Strategies ○ Built in assessment opportunities with student self-checks for comprehension	▪ Lessons designed around a connection to other areas of science. ▪ Built in assessments and lab activities ▪ Feature articles following every chapter emphasizing Science and technology in the real world. ▪ Lessons typically consist of the following elements: ○ Introduction to lesson key topics and terms to learn. ○ Math break activity ○ Science Connection ○ Lab activity ○ Built in Review questions	▪ Lessons organized around hands on science inquiry. ▪ Engaging graphics that reinforce concepts ▪ Lessons open with engaging investigative projects ▪ Integrated math skill activities ▪ Skill and Key concept activities ▪ Lessons typically consist of the following elements: ○ Introduction to key lesson topics ○ Engaging introductory activities ○ Reading guide ○ Lab activities ○ Highlighted vocabulary terms ○ Connections to other academic subjects ○ Learning checks ○ Built in review	▪ Introduces related new ideas that expand students' understanding ▪ Links science activities to other subject areas such as language arts, history, and social studies. ▪ Enagaging nonfiction stories. ▪ Lessons typically consist of the following elements: ○ Class Investigative activity ○ Visual learning activities ○ Summary and self-assessment

Table D2. Program Resources of Treatment and Control Programs

	Science Fusion	Control Program 1: (2005, 2006)	Control Program 2: (2001)	Control Program 3: (2005)	Control Program 4: (2002) Year 1 Only
Program Resources	**Student Resources** ▪ Write in Student Edition **Teacher Resources** ▪ Teacher's Edition ▪ Teacher's Lab Resource ▪ Assessment Guide **Digital Resources** ▪ Student & Teacher Edition ▪ Digital Lessons ▪ Virtual Labs ▪ Video Based Projects ▪ Teacher Resource Bank ▪ Online Assessment ▪ Power Notes Presentations ▪ Extra Support for Vocabulary ▪ NSTA *SciLinks* ▪ History Channel Videos ▪ ExamView Test Bank	▪ Teacher's Guide ▪ Student Edition ▪ Unit Resource Book ▪ Unit Assessment Book ▪ Unit Transparency Book ▪ Unit Lab Manual ▪ Audio CDs ▪ Lab Generator CD-ROM ▪ Test Generator CD-ROM ▪ eEdition CD-ROM ▪ Easy Planner CD-ROM ▪ Content Review CD-ROM ▪ Power Presentations CD-ROM ▪ Online Resources	▪ Teachers Editions ▪ Student Edition ▪ Lab Book	▪ Student Edition ▪ Teacher's Edition ▪ Skills Handbook ▪ Discovery Channel School Video ▪ Go Online Web Links ▪ Interactive Textbook	▪ Student Activity Book ▪ Student Reader

*Note that while these are the program materials listed with the control program, it us unknown whether control teachers had access to all of these resources whether because they were not purchased initially or because items have been transferred from teacher to teacher and lost over time, etc. In general, however, control teachers had access to the Student Editions, Teacher Edition, and lab book. In addition, control teachers may have incorporated other program materials (other than the primary program).

Table D3. Science Topics in Treatment and Control Programs

	Science Fusion	Control Program 1: (2005, 2006)	Control Program 2: (2001)	Control Program 3: (2005)	Control Program 4: (2002) Year 1 Only
Science Topics Covered	**Module A – Cells & Heredity** (includes cells, cell processes and energy, genetics and heredity, DNA, and genetic technology)	**X**	**X**		
	Module B – Diversity of Living Things (includes evolution, history of life, classifying life, viruses, bacteria, protists, fungi, plants, plant processes, animals, and animal behavior)	**X (does not include bacteria, protists, fungi, animal reproduction and behavior)**	**X**		**X (only includes plants)**
	Module C – The Human Body (includes the human body, skeletal, muscular, circulatory, respiratory, digestive, excretory, nervous, endocrine and reproductive systems, the immune system and infectious diseases)	**X**	**X**		
	Module D – Ecology & the Environment (includes populations and communities, ecosystems and biomes, earth's resources and living things, and human impact on the environment)	**X (does not include populations and communities)**	**X**	**X**	
	Module E – The Dynamic Earth (includes earth's surface, weathering and erosion, earth's history, minerals and rocks, plate tectonics, earthquakes, and volcanoes)	**X**		**X**	**X (only includes earth's surface and history)**
	Module F – Earth's Water & Atmosphere (includes earth's water, oceans, the atmosphere, weather and climate)	**X**		**X**	
	Module G – Space Science (includes earth, moon and sun, space, the solar system, stars, galaxies and the universe)	**X**		**X**	
	Module H – Matter and Energy (includes the properties of matter, energy, thermal energy and heat, atoms, periodic table, chemical reactions, acids bases and solutions)	**X**	**X**	**X**	

	Science Fusion	Control Program 1: (2005, 2006)	Control Program 2: (2001)	Control Program 3: (2005)	Control Program 4: (2002) Year 1
	Module I – Motion, Forces & Energy (includes motion, forces, work and machines, energy, electricity and magnetism and electromagnetism)	**X**	**X**	**X**	**X (only includes electricity and magnetism)**
	Module J – Sound & Light (includes properties of waves, sound, sound technology, light and electromagnetic spectrum)	**X**	**X**	**X**	
	Module K – Intro to Science & Technology (includes nature of science, scientific inquiry, measurement and data, engineering, technology and science)			**X (includes scientific inquiry only)**	

*Note that the above crosswalk reflects what content is available from each respective control program. Actual content taught is discussed in the main report, on page 29-30.

Appendix F:

Use of HMH Science Fusion Resources

Table F1. Percent of Usage of Science Fusion Program Components: Lesson Activities

	Never (did not do for any lessons)	Rarely (used for only a few lessons taught)	Some (used for about 50% of all lessons)	Often (used for about 75% of all lessons)	Used for almost all lessons
Lesson Vocabulary	0.0%	0.0%	0.0%	0.0%	100.0%
Active Reading	3.2%	0.0%	3.2%	12.9%	80.6%
Reinforce and Review	3.2%	3.2%	6.5%	6.5%	80.6%
Lesson Review	0.0%	3.2%	6.5%	12.9%	77.4%
Predict/Infer/Identify	6.5%	0.0%	3.2%	16.1%	74.2%
Essential Question	0.0%	3.2%	9.7%	22.6%	64.5%
Engage Your Brain	0.0%	6.5%	12.9%	19.4%	61.3%
Visualize It!	3.2%	3.2%	19.4%	12.9%	61.3%
Activities & Discussion	0.0%	3.2%	22.6%	19.4%	54.8%
Lesson Quiz	19.4%	16.1%	9.7%	3.2%	51.6%
Differentiated Instruction Activities as noted in the TE	9.7%	19.4%	6.5%	25.8%	38.7%
Going Further	10.0%	20.0%	13.3%	20.0%	36.7%
Think Outside the Book	3.6%	21.4%	25.0%	21.4%	28.6%
Think Science	10.0%	20.0%	23.3%	26.7%	20.0%
Do the Math	6.9%	6.9%	31.0%	37.9%	17.2%
STEM Engineering and Technology	22.6%	25.8%	25.8%	16.1%	9.7%

*% reflects reported percent of teachers who reported using the listed program components as noted.

Table F2. Percent of Usage of Science Fusion Program Components: Chapter Activities

	Did not do for any chapter	Occurred for some but not all chapters	Occurred for every chapter
Introduced the Big Idea	0.0%	25.8%	74.2%
Advance Planning	3.2%	35.5%	61.3%
Unit Test	16.1%	16.1%	67.7%
Citizen Science Unit Project	73.1%	26.9%	0.0%

*% reflects reported percent of teachers reporting using the listed program components as noted.

Table F3. Percent of Usage of Science Fusion Program Components: Lab Activities

	Every lesson	Every week	Every unit	Once per month
Science Fusion Virtual Lab	10.0%	5.0%	55.0%	30.0%
Science Fusion Daily Demo	0.0%	23.5%	35.3%	41.2%
Science Fusion Quick Lab	0.0%	13.0%	56.5%	30.4%
Science Fusion Exploration Lab	0.0%	7.1%	35.7%	57.1%

*% reflects reported percent of teachers reporting using the listed program components as noted.

Table F4. Percent of Usage of Science Fusion Ancillary Components

	Never (did not do for any lessons)	Rarely (used for only a few lessons taught)	Some (used for about 50% of all lessons)	Often (used for about 75% of all lessons)	Used for almost all lessons
Science Fusion Glossary	4.0%	12.0%	4.0%	16.0%	64.0%
Teacher Resource Bank	26.1%	8.7%	4.3%	4.3%	56.5%
Electronic Teacher Edition (eTE)	13.6%	4.5%	4.5%	22.7%	54.5%
ExamView Test Banks	33.3%	4.2%	4.2%	4.2%	54.2%
Power Notes Presentations	14.3%	23.8%	9.5%	4.8%	47.6%
Assessment Guide	24.0%	16.0%	4.0%	12.0%	44.0%
Teacher Access	13.6%	4.5%	13.6%	27.3%	40.9%
Unit Quiz	48.0%	4.0%	8.0%	0.0%	40.0%
Electronic Student Edition (eSE)	24.0%	20.0%	24.0%	0.0%	32.0%
Lab Manuals	25.0%	0.0%	33.3%	16.7%	25.0%
Student Access English	62.5%	4.2%	8.3%	4.2%	20.8%
Media Gallery Module A-K	40.0%	20.0%	12.0%	8.0%	20.0%
Common Core - Language Arts and Math	28.0%	28.0%	12.0%	16.0%	16.0%
Audio Selection Screen	32.0%	32.0%	12.0%	12.0%	12.0%
History Channel Videos	41.7%	37.5%	16.7%	0.0%	4.2%
People in Science	24.0%	28.0%	28.0%	16.0%	4.0%
Video Based Projects	32.0%	36.0%	12.0%	16.0%	4.0%
NSTA Learning Center and Science Links	36.0%	20.0%	32.0%	8.0%	4.0%
Online Assessments	76.0%	16.0%	4.0%	0.0%	4.0%
School Home Connection (Family Letters)	48.0%	24.0%	20.0%	8.0%	0.0%
Take It Home	40.0%	20.0%	36.0%	4.0%	0.0%
Multi-Language Glossary	80.0%	12.0%	8.0%	0.0%	0.0%

*% reflects reported percent of teachers reporting using the listed program components as noted.

PRES Associates, Inc.

PO Box 10730
Jackson, WY 83002
Phone: 307-733-3255, Fax: 307-222-0312
info@presassociates.com