Abigail Jurist Levy, Yueming Jia, Lisa Marco-Bujosa,
Julie Gess-Newsome, and Marian Pasquale

Science Specialists or Classroom Teachers: Who Should Teach Elementary Science?

Abstract

This study examined science programs, instruction, and student outcomes at 30 elementary schools in a large, urban district in the northeast United States in an effort to understand whether there were meaningful differences in the quality, quantity and cost of science education when provided by a science specialist or a classroom teacher. Student performance on the state's mandated science achievement test and student engagement in science lessons were used as student outcome measures. A conceptual framework of the elementary science experience guided the study, and data were collected on all components of schools' science instruction and science programs, including their costs, through interviews, observations, surveys, and school and district records. The data suggest that there is no single answer to the question. While poorly resourced school science programs produced poor student outcomes, not all well-resourced programs produced positive student outcomes. Students in schools where there was a high school-wide value placed on science—in both science specialist and classroom teacher models—achieved the best student outcomes. Those most effective science specialist schools had significantly lower per classroom costs than the most effective schools where classroom teachers taught science; and they also had the greatest commitment to science.

Introduction

Elementary science is important. Evidence indicates that students who do not have a solid exposure to science in the early years rarely make learning gains equal to those that did when they reach the secondary levels of schooling (Nelson & Landel, 2007). Additionally, the Framework for K-12 Science Education (National Research Council, 2012) and the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) lay out a path for science learning in which students' knowledge, understanding and skills are built on a foundation that is established in the elementary grades and continues across 13 years of schooling. Yet despite the importance of science in the early years, less and less time has been dedicated to science instruction. This trend has been documented by Sandler (2003) and McMurrer (2007) who found that 44% of districts across the country had cut the amount of instructional time for science in elementary schools in response to mounting pressures from mandatory testing in reading and mathematics (McMurrer, 2008). More recently, 46% of elementary teachers express concern about the limited amount of time available to teach science and 61% of schools report that a supportive context for science teaching does not exist (Banilower, Smith, Weiss, Malzahn, & Campbell, 2013).

There are a number of hypotheses and research findings to explain why science is not taught. Elementary teachers often possess inadequate content knowledge, have inadequate materials and facilities, are caught by competing curricular priorities, lack time and school/administrative support, and exhibit a minimal sense of self-efficacy to teach science, (Gess-Newsome, 1999; Rhoton, Field, & Prather, 1992; Schwartz, Lederman, & Abd-El-Khalick, 2000). These constraints fall into two broad categories: school-level support and the capacity of teachers.

Keywords: elementary teachers, science education, science specialist, principal leadership, cost effectiveness

Role of Elementary Science Specialists

Given the importance of implementing the NGSS at the elementary level and the many significant difficulties associated with changing classroom teachers' (CT) science instruction, one approach has been to turn to the role of science specialist (SS) as a school's primary source of science leadership, teaching, and/or support. Content specialists are commonly utilized in elementary schools to provide instruction in science and technology, the arts, and physical education, both in the United States and abroad (Ardzejewska et al., 2010; Gerretson, Bosnick, & Schofield, 2008). In the United States, approximately 16% of elementary students receive science instruction from a SS in addition to their regular teachers, and another 10% receive science instruction from a SS instead of their regular CTs (Banilower, et al., 2013). These percentages have held fairly constant for over a decade in the United States (Weiss, Banilower, McMahon & Smith, 2001). Unfortunately, little research exists that indicates what effect this approach will have on science instruction and student achievement (Schwartz, et al., 2000). As a consequence, school and district leaders are investing scarce resources in a strategy about which little is known and upon which much depends.

Goals of the Elementary Science Specialist Study

Levy, Pasquale and Marco's research agenda (2008), outlined a set of questions about the role and effectiveness of the SS model that warrant investigation, and the study we report on here responds to that agenda. In order to understand whether one model offered definitive advantages over the other, we constructed a study to explore the following research questions: (1) Are there differences in the quality, quantity, or cost of elementary science teaching when it is provided by a science specialist or a classroom teacher? (2) To what extent are differences associated with student achievement, as measured by students' scores on the state science achievement tests and students' engagement in science lessons?

In designing this study, we created a framework from our hypothesized influences on the quantity and quality of science instruction as delivered by SSs or CTs. This model evolved recursively during the study. Its final version is presented in Figure 1. In this framework, student outcomes are a product of the quality of students' science experiences, which are in turn products of the *science programs* within their schools, and the *science instruction* they receive. By science program, we mean those features that provide the context and supports necessary for a high quality science experience. These features include the value that is placed on science school wide, the support for science teaching and learning that principals provide, and the science-related resources available to teachers and students.

Each of the three features of the science program includes a number of components, displayed in the left column, followed in parentheses by the number of items that provided data on each component via interviews, observations, surveys, or kit inventories. (The details of data collection and analysis are provided in the section that follows.) For example, the *value of science* refers to the collective value that a school places on science as an aggregate of formal and informal, individual and school-wide policies and practices that indicate the importance of science as a subject, and the importance of the quality of science teachers' preparation and ability. The degree to which science learning is considered a school wide, shared responsibility, for example, is an indicator of the importance of science, and evidence of qualifications and/or ability to teach science that a principal looks for in those teachers with the responsibility to teach science is an indicator of the importance of teacher quality.

Principal support for science refers to the instructional support the building administration provides teachers by, for example, observing science lessons; institutional supports refer to such things as a school schedule that allows time for teachers to meet and discuss science during the school day. The *core science program* refers to the instructional materials, scope and sequence, assessments, and professional development endorsed and provided by the district's science department. Support for that program includes principals' roles in ensuring that it will be implemented at the highest level of quality possible. *Ancillary science activities* refers to those experiences, such as field trips or school visitors that provide additional science experiences for students beyond the district's kit-based program. In recognition of the value of a coherent science experience for students, only those ancillary activities that were purposefully aligned with the district's program were considered.

Resources incudes the space and materials available for science teaching and learning, which are closely linked. Teaching science requires a variety of equipment, materials, and supplies, often live specimens, and the science kits themselves, all of which need sufficient space to use in classrooms, organize, and maintain. The number and variety of ancillary science activities that are aligned to the district's core science program is also an indicator of the resources made available to teachers and students that enrich a school's science program.

Although a school's science program is distinct from the science instruction that students receive, the two are related and, in some ways, inter-dependent. It is very challenging for an effective and dedicated teacher of science to continue to provide a rich science experience for her students in the absence of the supports provided by a rich science program in her school. Similarly, a strong science program cannot compensate for poor science instruction. Therefore, we consider science instruction separately, and represent it as a composite of *teachers' roles*, the *instruction* they provide, and the *quantity of science instruction* their students receive.

The *teacher role* encompasses the leadership functions a teacher may perform in supporting the district's science program over and above teaching the kits. This may include raising funds for additional science activities or purchases, working with external science partners, participating in a science committee, or providing support to other teachers of science. Teachers' responsibilities for the core science program include such tasks as preparing science lessons, grading student work, or talking with parents about their child's progress in science. *Self-efficacy* refers to teachers' personal sense of comfort with science in general and with implementing the science curriculum, including their understanding of the strengths and weaknesses of the kits they teach.

Science instruction, as understood by observing lessons and interviewing teachers about their instruction, comprises teachers' general knowledge of pedagogy, and particularly their knowledge of classroom management, which is critically important when dealing with the plethora of science materials. *Pedagogical content knowledge* refers to such things as teachers' clarity, their responses to students' misconceptions, and their ability to link science content to other topics, lessons, or science domains. Rather than assume teachers' content knowledge based on their qualifications, content knowledge was measured during each observed lesson and in interviews based on teachers' understanding of the major learning goals of kits and activities, and their accuracy in portraying concepts and using instructional materials. *Integration* refers to the nature and extent to which teachers made use of mathematics and/or English language arts skills and knowledge during science lessons. *Access* refers to factors that may limit some or privilege other students with regard to the amount of science learning they experience relative to others in their grade, e.g., students who are assigned non-science work during science lessons, particular classes that are scheduled for more or fewer science lessons than their grade-level counterparts, or the number of students

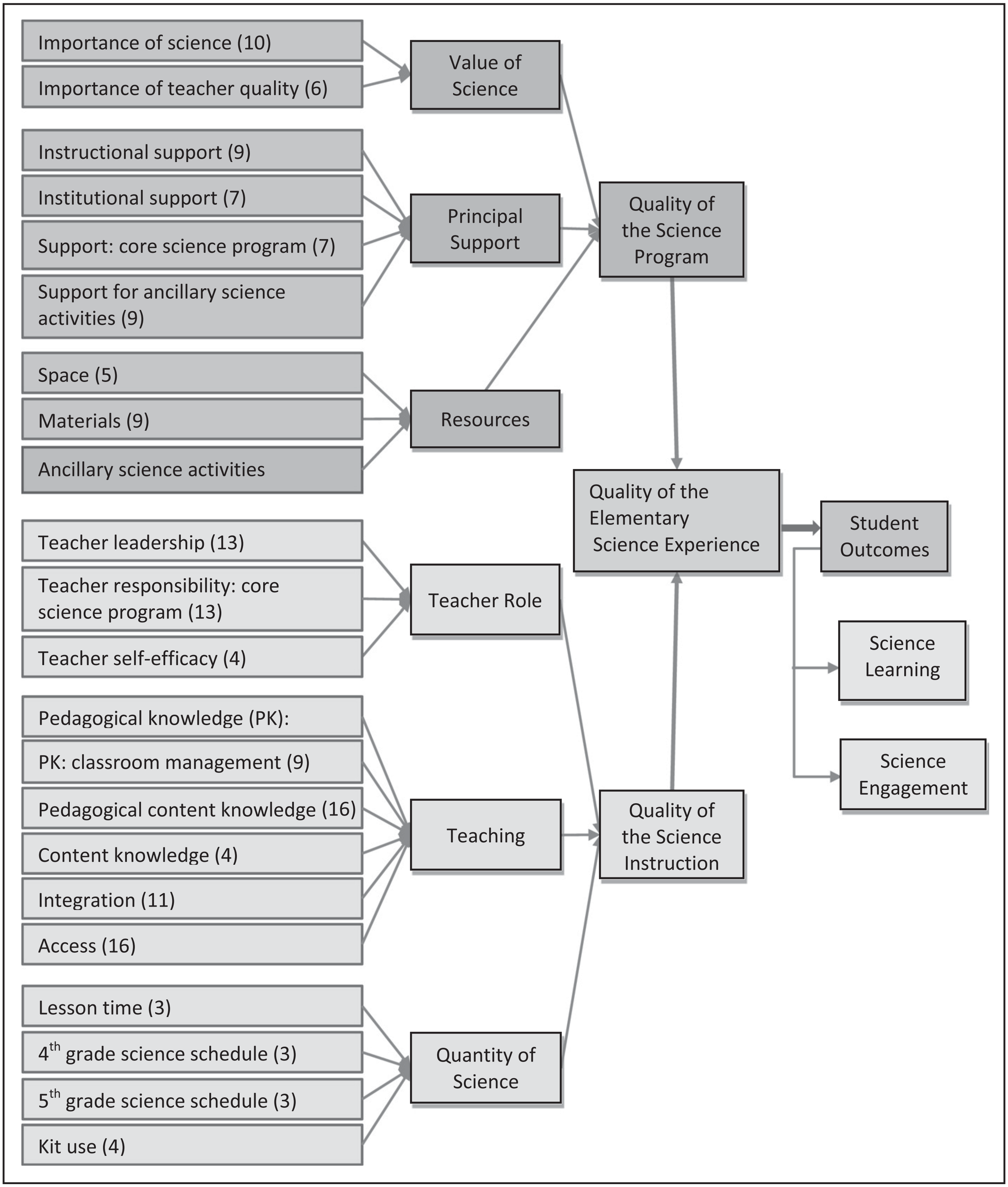

Figure 1. Conceptual Framework of the Elementary Science.

who are not present for science lessons in their entirety.

Finally, the *quantity of science* is measured by the number of minutes per year of science teaching that students receive, and the nature and extent to which each science kit has been used.

Methodology

The Elementary Science Specialist study was conducted in Mansfield[1], a large, urban district in the northeast U.S. between 2008 and 2012, and focused on the fourth and fifth grades. This district's elementary science program uses a selection of commercially-produced science kits, and the study centered its data collection and analysis on the use of two: Magnetism and Electricity in grade 4 and Levers and Pulleys in grade 5 (Delta Education, 2014).

Study Sample

Of the district's 75 schools that included fourth and fifth grades, 63 used a SS model. Classroom teachers taught science to their students in the remaining 12 schools. Schools were selected through a recruitment process involving the district's central office and science department staff. In order to be eligible, schools must use the district's science program, and teachers of science (SSs or CTs who teach science) must be at least in their second year of using the district's science kits.[2]

Thirty schools participated in our study (see Table 1); SSs were responsible for science in 26 schools (with two exceptions discussed below), and CTs taught science to their own students in four schools. The number of CT schools[3] is low district wide with our sample representing one third of those schools.

Table 1. Total Study Participants

	Schools	Principals	Science Specialists	Classroom Teachers who Teach Science	Classroom Teachers
SS Schools	26	26	26	4	22
CT Schools	4	4	0	8	0
Total	**30**	**30**	**25**	**12**	**22**

Schools in our sample serve an average of 340 students with the smallest school averaging an enrollment of 120 students and the largest 700. The study schools are similar to the other 45 schools in the district that have fourth and fifth grades with regard to poverty rates, percent students of color, and performance on the state's standardized tests in math, science, and English language arts. In comparison to the district, however, an ANOVA showed the sample schools have a marginally lower percent of limited English proficient students (n=30) (M= 20.63, SD=9.32) than non-sample schools (n=45) (M= 25.3, SD=10.31), but this difference is not statistically significant, $F(1,73)=3.62$; $p=.06$.

Table 1 displays the participants in the study. In all, 59 teachers and 30 principals participated in the study. Of the 59 teachers, 37 were teachers of science—25 SSs and 12 CTs.[4] The remaining 22 CTs did not teach science; their students received science instruction from a SS who participated in the study. Nine of the 25 SSs had a license to teach science to grades K-6, 5-8 or both, or their license application was pending approval; one of the 12 CTs who taught science had a license to teach science grades K-8. The 25 SSs had taught science for an average of 6 years (SD=4.3).

Measurement and Data Collection

Evidence of the quality of science programs and science instruction, as well as the cost of providing each school's overall science experience was gathered through interviews, surveys, observations, and kit-use inventories. For the purposes of this study, student outcomes were measured by class scores on the state's science test administered to 5th grade students, and by student engagement in science lessons, which were observed by members of the research team.

Teachers of science were interviewed twice, first about how science "works" in their school and second about their use of one of the two science kits of interest. In addition, three of their lessons using that kit were observed. The teachers also completed a survey that provided additional information about science in their schools and detailed information about the tasks they undertake (and the time required) related to their role as a science teacher in their school and in the district. CTs whose students were observed during their science lessons with their SS were also interviewed about their role with regard to science teaching and learning, and how science is managed in their schools. Principals were interviewed twice, first about science in their schools and in the district, and second about their roles with regard to science, the roles of others in their schools [e.g., volunteers, parents, or other teachers (*not* teachers of science)] with regard to the core science program, the tasks they each undertake, and the time and resources these tasks require. Detailed field notes were taken during observations and each teacher's instruction was recorded [5]; all interviews were recorded and these audio files were transcribed. Table 2 displays the total data collected over the course of the study, with the exception of kit use data. In order to document the extent to which the two science kits of interest were used by CTs and SSs, science department staff completed inventories of the two kits

[1] A pseudonym.

[2] Some exceptions were made for teachers with less experience but strong interest in participating in the study.

[3] For the remainder of the article we use *SS schools* to refer to schools where science specialists are responsible for teaching science; we use *CT schools* to refer to schools where classroom teachers are responsible for teaching science, and *CTWTS* to refer to classroom teachers who teach science, whether they are located in a SS or CT school.

[4] The models for science teaching are discussed in the Findings section.

[5] Teachers were fitted with a lapel microphone and a digital recorder; students were not purposefully recorded.

upon their return from all schools in the district to the science resource center for refurbishment. Inventories were collected on returned kits from 76 schools[6] and 226 teachers over four kit rotations for the fourth grade, and 5 kit rotations for the fifth grade. In all, a total of 525 kit inventories were collected. A total of 153 interviews were conducted with principals, SSs, CTs who teach science (CTWTS), and CTs whose students receive science instruction from their SS.

All data collection instruments were developed for this study. A sample of CTs and SSs were involved in the development and refinement process for the interview protocols; first in a series of pilot-tests and field tests, and then in a round of cognitive interviews to strengthen the validity and reliability of the items (Desimone & Le Floch, 2004). The development of the classroom observation protocol was based on the district's guidance for effective instruction. Guidance is communicated in two district policy documents that describe the dimensions of effective instruction, as well as specific expectations for and evidence of effective instruction, and school and classroom resources (Mansfield Public Schools, 2006; Mansfield Public Schools, 2007). In order to measure the presence of these attributes, the work of Hapkiewicz (1992); Rezba, Auldridge, and Rhea (1999); Tushnet, Millsap, Abdullah-Welsh, Brigham, Cooley, Elliot et al. (2000); Minner and DeLisi (2012); and Piburn, Sawada, Falconer, Turley, Benford and Bloom, (2000) was consulted. The observation protocol ultimately incorporated and adapted items from these instruments, and also included items developed specifically for this study. The protocol underwent a revision process that included field tests with multiple observers, concluding when observer teams reached an agreement rate of 80 percent on the coding for several science lessons.

[6] One school provided kit-use data in year 1 of the study. It was closed in the second year of the study, leaving the 75 schools as the total referred to throughout this article.

Table 2. Data Collection

	Interviews				Surveys		Observations	
	Principal	*SS*	*CTWTS*[a]	*CT*	*SS*	*CTWTS*	*SS*	*CTWTS*
SS Schools	51	48	8	22	23	4	70	12
CT Schools	8	0	16	0	0	8	0	22
Total	59	48	24	22	23	12	70	34
Total	153				35		104	

[a] Classroom teacher who teaches science

Analysis

The school was the unit of analysis, and so interview, survey, observation, and kit use data were reduced and summarized in order to arrive at a set of scores that represented the composite and overall quantity, quality, and cost of the science experience provided by each school. Multiple data sources from each school allowed for triangulation, resulting in a school portrait that was multi-dimensional.

A codebook was developed, tested, and revised by the research team to guide the data coding and analysis process. Data for each school was then assembled and consensually coded by pairs of researchers to ensure suitable levels of inter-rater agreement. The entire research team assigned weights to each item to reflect its value relative to other items within the same category. Raw scores were then converted to weighted scores. Each school's weighted scores were summed by category, and the percentage of a perfect score for each category was then derived. In cases of missing data, whether unreported or not applicable, the total possible score was adjusted down so a final rating for a given category would be fairly derived. If half of the data within a category were missing, that category was treated as missing.

The cost of schools' science programs was calculated using the ingredients method (Levin & McEwan, 2001). Interview and survey items probed the time spent by teachers, building administrators and others on tasks associated with each component of the elementary science model, as well as materials, supplies, books, other resources, and space. Time was monetized based on the district's average teacher or building administrator salary. In the case of science department staff, the average teacher, central office administrator, or lab technician salary was used as appropriate. Where parents or community members volunteered, the hourly rate of science department lab technicians was used to monetize their time. Where outside organizations partnered with a school in a formal arrangement, the average teacher salary was used. Where possible, the actual dollar value of materials, supplies, etc. was gathered from school records. The value of space dedicated to science, such as a science classroom, was calculated based on the cost per square foot charged by the district for the rental of excess school buildings and the average classroom size.

We report both significance and effect size in the analyses. Following the norm of the field, we use .05 as the cutoff point for a significant effect; however, p values less than .08 will also be provided as an indication of borderline significance. Because significance (p value) is very sensitive to sample size, we also report effect size. An effect size larger than .30 is considered an indicator of meaningful difference between groups (Shymansky, Hedges, & Woodworth, 1990).

Findings

Models of Science Specialist Deployment

SSs can be deployed in a variety of ways (Levy, Pasquale, & Marco, 2008), but the 26 SS schools in our sample were consistent in their approach, with two exceptions. Of the 26 schools, 22 had one SS who was the sole provider of science lessons to all students in the school, the remaining 4 SS schools had two specialists who divided that responsibility by grades.[7] In two SS schools, CTs taught science with their SS. These CTs and their specialist colleagues made unique arrangements to co-teach science. This was done either by teaching all science lessons together or by conducting science lessons independently and on different days. Of the 26 SS schools, 13 had a science classroom, and 13 used a cart or other conveyance to carry their materials and supplies from one classroom to another. This typically involved negotiating stairs, carrying large containers of water, live animal specimens, or other materials. Using Century, Rudnick and Freeman's (2008) categorization of the critical components of SSs' work, the specialists in this district were responsible for some or all of a variety of tasks. Those SS tasks that correspond to the responsibilities for the core science program as specified in our conceptual framework include organizing, managing, and providing materials; providing science instruction alone or co-teaching science with a CT; and fulfilling other responsibilities related to teaching such as preparing lessons, reviewing student work and preparing report cards, and attending professional development. Tasks that are associated with leadership for science in the school include outreach to parents and the community, leading professional development, participating in science committees, raising funds for science, and facilitating school-wide science-related events.

Differences between SS and CT Schools' Science Programs

Comparing the science programs points out what we believe to be a unique pre-disposition to science in these four CT schools in contrast to the 26 SS schools, that are more varied overall. This science inclination can be seen in CT schools' higher scores for the *value of science* and *principal support.* Differences in the other categories are far less obvious.

With regard to the overall *value of science* (Table 3), CT schools have higher scores for the *importance of teacher quality*. ANOVA results showed that the difference is statistically significant ($F(1,29)=19.73$; $p=.00$), with a large effect size ($d=0.45$). In these CT schools, for example, principals explicitly valued CTs' prior experience with science teaching when making hiring decisions, whereas principals in SS schools did not necessarily require special knowledge of or experience teaching science when hiring a SS. They often reported that they filled that position for a variety of reasons including a desire to retain a CT in the face of declining enrollment and class reduction. Principals in CT schools were more apt to encourage science-related professional development, and to provide substitutes for school-day workshops than principals in SS schools.

CT schools also had higher scores for the *importance of science*. The difference is marginally significant ($F(1,29)=3.63$, $p=.07$); however, the effect size is large ($d=1.49$). In this case, principals and teachers in CT schools were more likely to express a stronger sense of shared responsibility for science teaching and learning, whereas principals and teachers in SS schools more often reported that the SS had the sole responsibility for students' science learning and achievement.

Scores for *principal support* for science were higher in CT schools than SS schools, and this was evident in all four sub-categories (Table 4). The results of analyses showed that CT schools have higher scores in *support for ancillary science activities* that are statistically significant ($F(1,29)=5.50$, $p=.03$), and have an effect size of 1.04. For example, principals in CT schools more often directly engaged with external science partners, wrote grants to support science activities or purchases for their school, and/or provided support for school field trips. CT schools have higher scores in *instructional support* that are marginally significant ($F(1,28)=3.28$, $p=.08$), and have a large effect size of .61. These principals would, for example, conduct non-essential observations of science lessons and provide teachers with feedback on their science teaching, they would communicate with teachers about science and actively support collaboration among teachers about science-related matters. In addition, CT schools have slightly higher scores in *institutional support* and support for the core science program than SS schools. By institutional support we mean creating school structures and systems with the explicit intention of enabling teachers to attend to science teaching and learning. Principals of CT schools did such things as creating a school schedule that allowed time for teachers to collaborate specifically around science during the school day, creating an instructional schedule that prioritized science lessons, and/or creating a science committee. CT principals provided more *support for the core science program* by dealing with issues around space, materials, professional development over and above the district's basic kit training that would enhance teachers' science instruction. Although these differences do not reach a level of statistical significance, the effect sizes for the two variables indicate meaningful differences between the SS and CT schools ($d=0.36$; $d=0.61$ respectively).

Table 3. Overall Value of Science

	Importance of Teacher Quality	Importance of Science
	Mean (SD)	*Mean (SD)*
CT	.84 (.19)	.39 (.20)
SS	.23 (.26)	.22 (.17)
Total	.31 (.33)	.24 (.18)

The differences in *resources* for CT and SS schools are less stark (Table 5). CT schools have slightly lower scores in *space* (with the exception of one CT school that had a SS and classroom for the younger grades—they do not have science classrooms) and *materials*, and higher scores in *ancillary science activities* than SS schools, but these differences do not reach a level of significance, and the effect sizes are less than .30.

These data suggest that the CT schools participating in the study shared a high commitment to science, which was greater overall than that of the SS schools.

7 In these schools, only one SS served the 4th and 5th grades and our study focused on that SS alone.

Table 4. Principal Support

	Instructional Support	Institutional support	Support: Core Science	Support: Ancillary Science
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
CT	.84 (.17)	.51 (.14)	.44 (.10)	.67 (.10)
SS	.61 (.25)	.41 (.14)	.27 (.24)	.35 (.27)
Total	.64 (.25)	.42 (.14)	.30 (.23)	.39 (.28)

Principals and teachers from CT schools distinguished themselves by their shared sense of responsibility for providing science to their students, and by their principals' investment of time and resources in supporting their teachers' science instruction. Although this level of commitment did not characterize the group of SS schools as a whole, there were several schools among them that also were highly committed to science. This range can be seen in the larger standard deviations displayed in tables above, in the scores of particularly effective schools discussed later in this article, and in their teachers' and principals' comments and actions.

It is reasonable to ask why principals of schools where the commitment to science was less robust were willing to participate in this study and reveal the status of science in their schools. One hypothesis could be that the promise of confidentiality reassured even those principals of SS schools where the profile of science was quite low, and they wanted their willing teachers to be able to contribute to the study and benefit from the participant stipends. Or, possibly principals were sensitive to the need to pay more attention to science and wanted the study to include schools where science was not a high priority so that the findings would be useful to them and their colleagues. Perhaps principals believed that they were doing the best that could be done given the constraints they faced and were proud of what they had accomplished with meager resources. Whatever their motivations, none of these principals explicitly dismissed science as irrelevant in light of the compelling pressure to focus on math and English language arts – a characterization we see in the literature (Blank, 2012; Traphagen, 2011). Therefore, we contend that this study presents a picture of science programs in schools where the value of science teaching and learning is at least minimally acknowledged by principals; and the fate of science in non-participating schools may be more precarious.

Table 5. Resources

	Space	Materials	Ancillary Science Activities
	Mean (SD)	*Mean (SD)*	*Mean (SD)*
CT	.48 (.22)	.61 (.11)	.75 (.14)
SS	.58 (.25)	.62 (.10)	.67 (.25)

Differences between SS and CT Schools' Science Instruction

Overall, the differences in science instruction between the two groups of schools are less striking than the differences in their science programs. With regard to *teacher roles*, SSs take on a greater *leadership* role for science in their schools than their CT counterparts, such as doing more fundraising, engaging with science-related partners, and participating in science-related committees, while teachers in CT schools take on more *responsibility for the core science program*, i.e., they report spending more time correcting science homework, communicating with parents about science, integrating science into other subjects, preparing report cards, or conducting assessments of students' science learning. These differences, while not statistically significant, had meaningful effect sizes (d=0.35 and d=0.59 respectively). There were virtually no differences between the groups with regard to *self-efficacy*.

The quality of the *science teaching* that was observed was also fairly similar across the two groups (Table 6). Although CTs scored consistently higher on all six measures, the differences were never statistically significant, nor did they have meaningful effect sizes with the exception of pedagogical knowledge specific to classroom management. In this regard, for example, CTs had more effective strategies for dealing with behavior problems during the lessons, managed materials more efficiently, and their instructions were more often clear and more easily understood by their students. Where the difference did not reach statistical significance, it had a large effect size (d=0.48).

Plausible explanations for this difference may be that the SSs' professional preparation for teaching science was not strong overall; only nine of the 25 SSs had a license to teach science. Additionally, SSs who travel to their students' classrooms are a "guest" when they teach. They have not set the expectations for student behavior, they do not know their students as deeply as the resident CT, and the CT often leaves the room when the SS begins her lesson in order to plan, meet with other CTs, or perform other responsibilities. These conditions, especially when combined, make managing student behavior difficult.

Figure 2 displays a summary of the score differences between the SS and CT programs and instruction. Only the differences between SS and CT schools with respect to the overall value of science and principal support reached a point of statistical significance and/or a meaningful effect size.

Finally, there are some meaningful differences in the *quantity of science* that students receive (Table 7). Students in CT schools have longer science lessons than students in SS schools. That difference is statistically significant ($F(1,25)=4.68$; $p=.04$), with an effect size of 0.41. CT schools also provide more science time per week for 4th and 5th graders. These differences do not reach a level of statistical significance, but the effect size for 4th grade science time (d=0.34), indicates a meaningful difference between the two.

All of these differences make intuitive sense. SSs have a school-wide focus whereas CTs' attend to their own

Table 6. Quality of Science Instruction

	PK[a]					
	General	Classroom Mngmt	PCK[b]	CK[c]	Integration	Access
	Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
CT	.83 (.02)	.89 (.09)	.83 (.14)	.88 (.13)	.50 (.22)	.09 (.47)
SS	.81 (.11)	.79 (.15)	.79 (.11)	.83 (.12)	.39 (.13)	.09 (.35)
Total	.81 (.10)	.80 (.14)	.79 (.12)	.83 (.12)	.41 (.12)	.09 (.36)

[a] Pedagogical knowledge
[b] Pedagogical content knowledge
[c] Content knowledge

classroom, and so the differences in the scopes of their roles with regard to science are not surprising. Similarly, CTs' comparatively better record of managing classroom behavior can be explained by the same reasoning. In contrast to half of the SSs in the study who did not have their own classrooms, CTs know their students well and teach in their own classrooms where they have established and can enforce the norms and expectations for their students. Similarly, CTs have more flexibility to adjust the time they spend on science lessons with their students, whereas SSs follow a rigid school-wide schedule. Finally, given the lack of science teaching backgrounds for many SSs, it is not surprising that their instruction is quite similar to CTs.

Differences in Student Outcomes

State science test. Over the past five years and including all 75 schools in the district that serve grade 5, SS schools had a slightly higher percentage of students passing the state mandated science test (M=66.16, SD=16.03) than CT schools (M=62.00, SD=14.58), however the difference was not significant, controlling for limited English proficiency composition, low income student composition, and percentage of students who passed MCAS math. The magnitude of the difference was also small (d=0.26).

The similarity in student performance within our sample schools was consistent with the district as a whole, where an average 67.9 percent of students in our SS schools passed the science test (SD=15.73) compared to 68.6 percent of students in our CT schools (SD=3.30). However, more compelling was the finding that principal support—particularly *institutional support*—accounted for 10% of the variance in the pass rate, over and above that explained by other demographic factors. This suggests that the efforts principals make to provide visible, concrete support for science, and especially structural supports, e.g., schedules that explicitly allow time for collaboration and communication about science during the school day, make a positive difference to student achievement.

Teacher characteristics—particularly teachers' sense of self-efficacy—also accounted for 10% of the variance in the percentage of students passing the science test. Although confidence in one's abilities is not always warranted, the NGSS will level the playing field for all science teachers. Teachers will need time and support in order to feel comfortable with the new demands the standards make on them and their students, and confident in their own abilities to meet those demands effectively. This finding suggests that providing teachers with such support will make a difference to their students' performance.

Finally, *science instruction* accounted for only 7% of the variance in the percentage of students passing the state science test. We believe the relatively low impact of teaching on student achievement reflects several factors, the most obvious being that good preparation for this test is not necessarily good instruction, which was evident in our observations. Second, the test is cumulative, covering science topics taught in grades 3-5; however the district's high student and teacher mobility rates prevent students from experiencing the curriculum sequence as intended, thereby contributing to low pass rates. Third, this test is administered in the early spring, when only two thirds of the 5th grade curriculum has been taught.

Of equal interest is the finding that the other components of science programs, such as the value of science, resources, or the quantity of science that students received as measured by minutes of science per week or kit coverage had significant associations with students' pass rate on the science test.

Student engagement. We chose engagement in science lessons as a student outcome because it was observable by researchers in the classroom, and

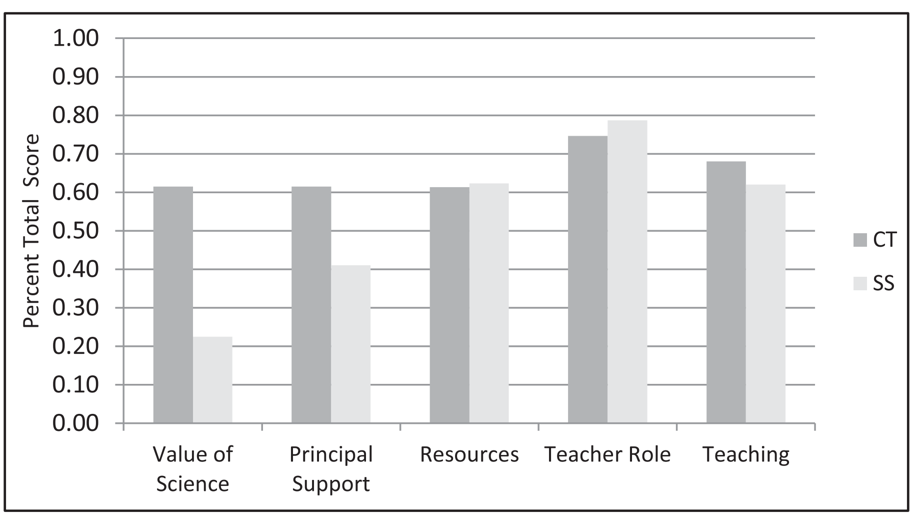

Figure 2. Summary Comparison of CT and SS programs and Instruction.

Table 7. Quantity of Science

	4th Grade Min/Week	5th Grade Min/Week	Lesson Time
	Mean (SD)	*Mean (SD)*	*Mean (SD)*
CT	148.75 (61.29)	122 (22.52)	54.03 (12.17)
SS	109.06 (47.59)	115.69 (45.21)	45.04 (6.76)
Total	114.35 (50.29)	116.34 (43.18)	46.42 (8.19)

although it is not a direct measure of student learning, it is without question a necessary prerequisite. Students in CT schools were more engaged in their science lessons (*M*=.73, *SD*=.27) than students in SS schools (*M*=.11, *SD*=.62). ANOVA results showed the difference between the two groups is marginally statistically significant ($F(1,25)=3.85$, $p=.06$); however the magnitude of the difference is large ($d=1.03$).

In addition to estimating student engagement by CT and SS model alone, we also estimated it for SS schools with and without a classroom (Table 8).

The standard deviations suggest that there was more variance among the SS classrooms than CT classrooms. This could be explained by the teachers themselves, and the quality of their instruction; however, another program component had a strong association with high levels of student engagement: the *value of science.* The *value of science* accounted for 15% of the variance in students' engagement, and within that category, the importance placed on teacher quality was most prominent. This suggests that when hiring either CTs who will be responsible for teaching science or SSs, the value that principals place on candidates' science knowledge and science teaching abilities when making hiring decisions pays off for student engagement. Similarly, the effort and resources that principals invest in making science professional development available and accessible to their teachers makes a difference in the degree to which teachers are able to engage their students in science learning. In contrast to student achievement, *science instruction* explained 43% of the variance in student engagement. Moreover, all components of instruction with the exception of integration of science with English language arts and/or mathematics, contributed to the explanation of the variance.

Table 8. Student Engagement in Science Lessons for CT and SSs with and without a Classroom

	Mean (SD)
CT	.85 *(.07)*
SS – 0 Classroom	.01 *(.59)*
SS – 1 Classroom	.16 *(.63)*
Total	.20 *(.62)*

These two features—*value of science* and *science instruction*—to the exclusion of all others including the school-wide importance placed on science, principals' support for the core science program, teachers' leadership roles for science in their school or the amount of science students receive, explain students' engagement in science lessons. These findings suggest that by hiring good science teachers and providing them with professional development and other supports, principals will enable students to receive good science instruction that effectively engages them in science learning.

Differences in Science Program Costs

Table 9 presents a summary of the per classroom costs of the two science program models. It was difficult to assign costs to institutional support; therefore principal support is the sum of support for the core science program, instructional support, and support for the ancillary science program. Worth noting is the fact that although CT schools scored significantly higher on the value of science, this was not reflected in higher costs because many features of valuing science could not be monetized. These features included a school-wide, shared responsibility for science; principals' clear and explicit goals for science; a school-wide emphasis on science writing; and the seriousness with which students take their science lessons as observed and reported by SSs and CTs. A notable difference between the program costs for CT and SS schools is that CT principals invested twice as much in providing support for the core science program as SS principals, and ten times the investment in instructional support, whereas SS principals more than doubled the amount they spent on providing support for ancillary science activities.

Because the cost of a dedicated science classroom was considerable, the two right-hand columns break out the costs for SS schools that have a classroom and those that do not. It is clear from these data that an SS model without a classroom is the least expensive; about 30% less than both an SS program that provides a classroom and a CT program.

Table 10 displays the ratio of model costs to student engagement. Although the cost per classroom for the CT model is the highest, it is only 3% greater than the SS model with a classroom ($2,343 and $2,269 respectively). Moreover, the mean level of student engagement achieved by CTs is more than five times that of SSs with their own classroom. In comparison, the SS model, inclusive of those with and without classrooms, is about two thirds the cost of the CT model, but the student outcome is poor. Put another way, while the SS model may cost 44% less than the CT model, more than 87% of student engagement is lost while having limited impact on student test scores (which might not be an accurate measure of student learning). Is that outcome for students worth the savings? All of this suggests the CT model is the most cost effective, while raising the possibility that the SS model may only be even moderately effective when a dedicated classroom is provided.

Grouping the sample by key characteristics enables a variety of comparisons to be made. Table 11 provides a profile of eight groups of CT and SS schools, displayed in the table columns (the ten schools with the highest and lowest student engagement scores[8]), all CT and

[8] Three of the four CT schools were in the highest SE score group

Table 9. Per Classroom Cost of Science Programs: CT and SS Schools

	CT Schools (n=4)	SS Schools (n=26)	SS Schools 0-Classroom (n=13)	SS Schools 1-Classroom (n=13)
Value of Science	**$ 23**	**$ 207**	**$ 214**	**$ 201**
Principal Support	**$ 1,101**	**$ 520**	**$ 231**	**$ 767**
Support: Core Science Program	$ 983	$ 463	$ 140	$ 740
Instructional Support	$ 99	$ 10	$ 7	$ 12
Support: Ancillary Science	$ 20	$ 47	$ 84	$ 16
Resources	**$ 1,219**	**$ 827**	**$ 276**	**$ 1,300**
Total Program Cost per Classroom	**$ 2,343**	**$ 1,554**	**$ 720**	**$ 2,269**

all SS schools, the SS schools sorted by those with a dedicated science classroom and those without, and the two most effective CT and SS schools based on student engagement scores. The table rows provide summary data for each group including the mean total program costs and the percent of those costs invested in the value of science, principal support, and resources; the returns on those investments in the form of mean student engagement and teacher leadership scores[9]; summary mean instructional quality score; and the mean program quality scores. We include teacher leadership scores as an outcome because it represents a potential return on an investment in teachers of science beyond the instruction they provide. We hypothesize that SSs are more likely than CTs to take on a leadership role for science in their schools, and that principals may consider that role as an added advantage when weighing the costs and benefits of the SS versus the CT model.

Table 11 shows that total program costs vary greatly, from $550 to $2,898, and provide some evidence that money matters. The two least expensive school groups produced by far the lowest student engagement scores—the ten schools with the lowest student engagement scores (-.43 at a per classroom cost of $552) and the SS schools with no dedicated science classrooms (.01 at a per classroom cost of $720). Moreover, both these groups (all of which are SS schools) also had the lowest overall program quality scores (.40 and .37 respectively). Although these SSs may not have been the strongest teachers, they were also swimming against the tide, teaching science in schools where it was not valued highly, and where few resources were invested.

At the same time, investments in science cannot explain everything. Two school groups with similarly high per classroom program costs—SS schools with a dedicated science classroom ($2,269) and the two most effective CT schools ($2,294)—achieved vastly different student outcomes (.16 and .88 respectively). Teachers in these two CT schools achieved their success with the support of significant program investments, over 90% of which came in the form of principal support, and in schools where the quality of their science programs was relatively high (.55). In contrast, SS schools with a dedicated science classroom had science programs that were relatively weak (.44), and although these SSs took on markedly greater science leadership roles, their student outcomes were relatively poor (.16). It may be that while a classroom doesn't guarantee strong student outcomes, it relieves SSs of the logistical burdens they would otherwise face, and frees them up to take on other responsibilities. A classroom may also give a principal and others in the building the sense that the SS has been given all that is needed, and their attention can turn to other matters. An obvious question is whether these SSs could achieve better student outcomes *and* maintain the profile of science in their schools if the science programs were stronger?

The schools that stand out among the 30 in the sample are the two most effective SS schools. These schools invested 42% less than the costliest program ($1,660 per classroom), and achieved nearly the best student engagement score (.87). These SSs may have been stellar teachers (their summary instruction scores were not as high as the two most effective CTs, but higher than the SS mean score), but they also worked in schools with the strongest science programs (.65), and particularly the highest score in the value of science (.69). Moreover, unlike the two most effective CT schools, these SS schools distributed their investments more equitably between principal support and resources. Rather than swimming against the tide, these teachers were working in schools where their efforts were amplified by a culture that valued science and provided concrete support to enable it to flourish in the experiences of their students. These conditions suggest that an elementary school's strong, supportive science culture might enhance the effectiveness of a science specialist, enabling students to have more positive learning outcomes than their teacher could achieve on her own.

Conclusions and Implications

A simple comparison of the CT and SS models in this sample suggests that the science *programs* in CT schools were

[9] Tasks a teacher may perform to support the district's science program beyond teaching the kits, e.g., fundraising, working with external science partners, leading a science fair or a science committee.

Table 10. Ratio of Costs to Student Engagement

Model	Cost/Student Engagement	% Difference from Highest Cost	% Difference from Highest Level of Student Engagement
CT (n=4)	$ 2,343/.85	0	0
SS – All (n=26)	$ 1,554/.11	-44%	-87%
SS – 0 Classroom (n=13)	$ 720/.01	-69%	-99%
SS – 1 Classroom (n=13)	$ 2,269/.16	-3%	-81%

Table 11. Distribution of Program Costs, Student Outcomes, P rogram Component Scores

	10 Schools w/Highest SE Score	10 Schools w/Lowest SE Score	CT Schools (n=4)	SS Schools (n=26)	SS Schools w/CL (n=13)	SS Schools w/no CL (n=13)	2 Most Effective CT Schools	2 Most Effective SS Schools
Value of Science (% of total cost)	**.09**	**.01**	**.01**	**.13**	**.09**	**.30**	**.02**	**.03**
Principal Support	**.57**	**.05**	**.47**	**.33**	**.34**	**.32**	**.91**	**.55**
Support: Core Science Program	.90	.64	.89	.89	.96	.61	.89	.85
Instructional Support	.03	.21	.09	.02	.02	.03	.09	.05
Support: Ancillary Science	.07	.14	.02	.09	.02	.36	.02	.10
Resources	**.34**	**.94**	**.52**	**.53**	**.57**	**.38**	**.07**	**.43**
Total Program Cost	**$2,898**	**$551**	**$2,343**	**$1,554**	**$2,269**	**$720**	**$2,294**	**$1,660**
Mean SE Score *(SD)*	.80 *(.16)*	-.43 *(.42)*	.73 *(.27)*	.11 *(.62)*	.16 *(.63)*	.01 *(.59)*	.88 *(.11)*	.87 *(.15)*
Mean Teacher Leadership Score	.53 *(.24)*	.66 *(.19)*	.50 *(.22)*	.65 *(.20)*	.71 *(.21)*	.60 *(.18)*	.35 *(.13)*	.44 *(.37)*
Mean Instructional Score	.71 *(.33)*	.57 *(.34)*	.67 *(.32)*	.61 *(.31)*	.62 *(.34)*	.62 *(.31)*	.76 *(.22)*	.69 *(.26)*
Mean Value of Science Score	.41 *(.28)*	.21 *(.19)*	.62 *(.19)*	.22 *(.22)*	.23 *(.25)*	.23 *(.19)*	.39 *(.10)*	.69 *(.20)*
Mean Principal Support Score	.48 *(.24)*	.42 *(.20)*	.62 *(.13)*	.41 *(.22)*	.40 *(.25)*	.42 *(.19)*	.64 *(.16)*	.64 *(.10)*
Mean Resources Score	.63 *(.17)*	.57 *(.15)*	.55 *(.16)*	.60 *(.17)*	.70 *(.11)*	.48 *(.11)*	.62 *(.12)*	.62 *(.17)*
Mean Overall Program Score	.50 *(.40)*	.40 *(.18)*	.59 *(.16)*	.41 *(.20)*	.44 *(.20)*	.37 *(.16)*	.55 *(.12)*	.65 *(.15)*

of higher quality overall and the difference was statistically significant; however there were no meaningful differences in the quality of the *instruction* that CTs and SSs provided. That said, among these 30 schools and across the two science program models, there were significant differences in the level of student engagement in their science lessons, and although some teachers provided better science instruction than others—and instructional quality accounted for 43% of the variance in student engagement—the difference in quality could not be explained by a teacher's CT or SS status, but rather to individual teachers within each group.

Similarly, program investments were not consistently associated with student outcomes. Although poorly supported programs produced poor student outcomes, well-funded programs were not always associated with strong student outcomes; in fact, sometimes the opposite was true. In SS schools that had a dedicated classroom, for example, student outcomes were bleak despite significant investments.

These data suggest, then, that there is no simple answer to the question, who should teach elementary science, CTs or SSs? The answer is: it depends. Employing an SS model guarantees that students will have some science instruction because lesson time is built into schools' schedules, but they may not necessarily have *high quality* science instruction. And although there are heroic examples of SSs who are effective without a classroom, the data show that this piece of real estate has some benefit for students and for the school, albeit at a cost. The same can be said of the CTs in our sample. Their principals invested heavily in both time and resources to provide them with a variety of supports, and these CTs delivered very positive student outcomes. One might ask if, in the absence of this level of commitment and support, and in light of competing pressures, would these CTs provide the same quality and quantity of science instruction? Regardless of the model, the data suggest a delicate balance between investing in science and valuing science. They are not the same thing, they are both critical, and whereas each one by itself can only do so much to advance student learning, it appears that both together can accomplish more than one would expect.

References

Ardzejewska, K., McMaugh, A., & Coutts, P. (2010). Delivering the primary curriculum: The use of subject specialist and generalist teachers in NSW. *Issues in Educational Research*, *20*(3), 203-219.

Blank, R. K. (2012). *What is the impact of decline in science instructional time in elementary school*. Noyce Foundation. Available at www. csss-science. org/downloads/NAEPElemScienceData.pdf.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc.

Century, J., Rudnick, M., & Freeman, C. (2008). Accumulating knowledge on elementary science specialists: A strategy for building conceptual clarity and sharing findings. *Science Educator*, *17*(2), 31-44.

Delta Education. *Full option science system*. (2005). Nashua, NH.

Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in educational research. *Educational Evaluation and Policy Analysis*. *(26)*1, 1-22.

Gerretson, H., Bosnick, J., & Schofield, K. (2008). A case for content specialists as the elementary classroom teacher. *The Teacher Educator*, *43*(4), 302-314.

Gess-Newsome, J. (1999). Delivery models for elementary science instruction: A call for research. *Electronic Journal of Science Education*, *3*(3).

Hapkiewicz, A. (1992). Finding a list of science misconceptions. *MSTA Newsletter*, *38*, 11-14.

Levin, H. M. & P. J. McEwan. (2001). *Cost-effectiveness analysis: Methods and applications*. Thousand Oaks, CA: Sage Publications.

Levy, A. J., Pasquale, M. M., & Marco, L. (2008). Models of providing science instruction in the elementary grades: A research agenda to inform decision makers. *Science Educator*, *17*(2), 1-18.

Mansfield Public Schools. (2006). *Dimensions of effective teaching*. Mansfield, MA. Retrieved from http://mansfieldpublicschools.org/dept

Mansfield Public Schools. (2007). *Seven essentials of whole school improvement*. Mansfield, MA. Retrieved from http://mansfieldpublicschools.org/dept

McMurrer, J. (2007, December). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington, D.C.: Center on Education Policy.

McMurrer, J. (2008). *Instructional time in elementary schools: A closer look at changes for specific subjects.* Washington, D.C.: Center on Education Policy.

Minner, D., & DeLisi, J. (2012). *Inquiring into science instruction observation protocol* (ISIOP). Newton, MA: Education Development Center, Inc.

National Research Council. (2012). *A Framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: The National Academies Press.

Nelson, G. D., & Landel, C. C. (2007). A collaborative approach for elementary science. *Educational Leadership, 64*(4), 72-75.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). *Reformed teaching observation protocol (RTOP)*. Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Rezba, R.J., T. Auldridge, & L. Rhea. (1999). *Teaching & learning the basic science skills*. Retrieved from www.pen.k12.va.us/VDOE/instruction/TLBSSGuide.doc

Rhoton, J., Field, M., & Prather, J. (1992). An alternative to the elementary school science specialist. *Journal of Elementary Science Education*, *4*(1), 14-25.

Sandler, J. O. (2003). Lest science be left behind. *Education Week*, *22* (29), 247-254.

Schwartz, R. S., Lederman, N. G., & Abd-El-Khalick, F. (2000). Achieving the reforms vision: The effectiveness of a specialists-led elementary science program. *School Science and Mathematics*, 181-193.

Shymansky, J.A., Hedges, L.V. & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching*, *27*, 127-144.

Traphagen, K. (2011). Strengthening Science Education: The Power of More Time to Deepen Inquiry and Engagement. Washington, DC: National Center on Time and Learning. Retrieved from http://www.timeandlearning.org/files/StrenghtheningScienceEducation.pdf.

Tushnet, N.C., Millsap, M.A., Abdullah-Welsh, N., Brigham, N., Cooley, E., & Elliot, J., et al. (2000). Final report on the evaluation of the National Science Foundation's Instructional Materials Development Program. Washington, D.C.: National Science Foundation, Directorate for Education and Human Resources, Division of Research, Education, and Communication.http://www.nsf.gov/pubs/2000/nsf0071/nsf0071.pdf, accessed on 3/3/2008 Instrument: http://www.nsf.gov/pubs/2000/nsf0071/nsf0071_13.pdf, accessed on 3/3/2008

Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). Report of the 2000 national survey of science and mathematics education. Chapel Hill, NC: Horizon Research, Inc.

Abigail Jurist Levy, PhD, Principal Research Scientist, Education Development Center, Inc. Correspondence concerning this article should be sent to: Abigail Jurist Levy, Education Development Center, Inc., 43 Foundry Avenue, Waltham, Massachusetts 02453-8313. Email: alevy@edc.org

Yueming Jia, PhD, Senior Research Associate, Education Development Center, Inc.

Lisa Marco-Bujosa, **MEd,** Research Associate, Education Development Center, Inc.

Julie Gess-Newsome, PhD, Associate Dean, College of Education, Oregon State University-Cascade

Marian Pasquale. MEd, Senior Research Scientist, Education Development Center, Inc.

Acknowledgement: This material is based upon work supported by the National Science Foundation under Grant No. 0611794. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.