TEMPORARY FIX OR LASTING SOLUTION?

Investigating the Longitudinal Impact of Teacher Professional Development on K–2 Science Instruction

Judith Haymore Sandholtz
UNIVERSITY OF CALIFORNIA IRVINE

Cathy Ringstaff
Bryan Matlen
WESTED

ABSTRACT
This study investigated the extent to which a state-funded teacher professional development program designed to improve K–2 science education led to changes that persisted beyond the funding period. The study used a longitudinal, mixed-methods approach and examined persistence of changes in teachers' content knowledge, self-efficacy, instructional time, and instructional practices in science. It also examined the extent to which school contexts and resources provided ongoing support for teachers to implement what they learned in the professional development. Data sources, collected over a 5-year period, included a teacher survey, a self-efficacy assessment, content knowledge tests, interviews, and classroom observations. Findings indicated a beginning pattern of decline during the 2 years after the program ended, but outcomes remained higher than before the professional development. Contextual factors varied widely across schools and influenced, in particular, the amount of time teachers devoted to science and their decisions about instructional strategies.

THE ELEMENTARY SCHOOL JOURNAL
Volume 117, Number 2. Published online October 28, 2016
 0013-5984/2016/11702-0002$10.00

SCIENCE plays a significant role in everyday life; consequently, a solid foundation in science is essential for students of all ages (National Research Council [NRC], 2011). Despite the established need for science education, national and state reports indicate that elementary students spend too little time studying science and that teachers are not well prepared to teach it (California Council on Science and Technology [CCST], 2010; McMurrer, 2007, 2008; Olson & Labov, 2009). Without teachers who are prepared and confident to teach science, elementary students are unlikely to encounter high-quality science instruction (Dorph, Shields, Tiffany-Morales, Hartry, & McCaffrey, 2011). Consequently, teacher professional development is an important method for improving science education in elementary schools. Professional development has the capacity to build teachers' content and pedagogical knowledge and thereby promote changes in classroom instruction (National Staff Development Council, 2001; Sparks, 2002; Stigler & Hiebert, 1999). The potential of professional development to prompt changes in science instruction is promising, but the majority of studies examine short-term impact. There is a paucity of research examining persistence of teacher change after programs end.

This study investigated the extent to which a state-funded, 3-year teacher professional development program designed to improve K–2 science education led to outcomes that persisted beyond the funding period. Earlier research on the impact of this program found significant changes in teachers' science content knowledge, self-efficacy related to science teaching, instructional time devoted to science, and instructional practices in science after just 1 year of professional development (Sandholtz & Ringstaff, 2011). These significant changes early in the 3-year program were unexpected because implementation of new instructional practices tends to be a gradual process and shifts often do not occur for several years (Guskey, 2002; Hawley & Valli, 1999; Johnson, 2007). Subsequent research found that these changes were largely sustained in the second and third program years (Sandholtz & Ringstaff, 2013a, 2013b). The current research, funded by the National Science Foundation, examined the sustainability of those changes over time. Specific research questions included: (*a*) To what extent did changes resulting from teacher participation in the professional development persist after the program ended? (*b*) To what extent did school and district resources and networks continue to provide adequate support for teachers to implement what they learned in the professional development?

Conceptual Framework

A key strategy for improving science education in elementary schools is professional development because it can strengthen teachers' preparation to teach science. State and national reports over the last decade have indicated that elementary teachers continue to feel less prepared to teach science than mathematics and language arts (Banilower et al., 2013; CCST, 2010; Dorph et al., 2011; Fulp, 2002; Olson & Labov, 2009; Weiss, Banilower, McMahon, & Smith, 2001). In the 2012 National Survey of Science and Mathematics Education, only 39% of elementary teachers reported feeling very well prepared to teach science; and within the field of science, they indicated feeling better prepared to teach life science and earth science than physical science (Banilower et al., 2013). Their perceptions of their preparedness reflect limited backgrounds in science. Elementary teachers have less extensive backgrounds in science than middle and high school teachers and do not meet the recommendations of the National Science Teachers Association (Banilower et al., 2013; Olson & Labov, 2009; Weiss et al., 2001). Teachers' science content knowledge is important because it influences classroom practice (Math and Science Partnership, 2008). With stronger content knowledge of science, elementary teachers are more inclined to use subject-specific strategies aimed at developing students' conceptual understanding of science. Increased preparation in science content has been strongly linked to elementary teachers' ability to construct inquiry-based lessons (Luera, Moyer, & Everett, 2005) and their increased use of inquiry-based classroom practices (Supovitz & Turner, 2000). Researchers also link changes in teachers' beliefs with changes in instructional practice (Fishman, Marx, Best, & Tal, 2003; Knapp, 2003; Richardson, 1996; Tschannen-Moran, Hoy, & Hoy, 1998). Teachers who have more confidence in their abilities tend to use more student-centered than textbook-driven approaches (de Laat & Watter, 1995). They also are more likely to try new ideas and to implement innovative and challenging instructional methods in the classroom (Ross, 1998; Tschannen-Moran et al., 1998).

Desimone (2009) proposes an operational theory of how professional development influences teachers, their instructional practice, and student learning. The core theory of action includes four main steps. First, teachers participate in effective professional development. Second, their participation increases their knowledge and skills and/or changes their attitudes and beliefs. Third, given their new knowledge and skills (or attitudes and beliefs), teachers improve their instruction through changes in content, pedagogy, or both. Fourth, the changes in their instructional practices promote student learning. In this model, context is considered a key mediating influence, and important contextual factors include, for example, student characteristics, teacher characteristics, curriculum, school leadership, policies at multiple levels, and classroom, school, and district environments.

In keeping with this operational theory, researchers have documented changes in teachers' attitudes as well as their science instruction over the course of professional development programs. Studies of elementary and middle school teachers showed gains in teachers' confidence in teaching science and their use of standards-based instructional materials and practices throughout their participation in professional development (Basista & Mathews, 2002; Lakshmanan, Heath, Perlmutter, & Elder, 2011). Reports from the Local Systemic Change (LSC) initiative linked the extent of elementary teachers' participation in professional development with increased use of standards-based instructional materials and increased instructional time spent on science (Heck & Crawford, 2004; Heck, Rosenberg, & Crawford, 2006). Similarly, Supovitz and Turner (2000) examined data from the LSC initiative and found that the quantity of professional development was strongly linked to inquiry-based teaching practice and an investigative classroom culture.

Although professional development has been shown to enhance teachers' attitudes toward science and their use of inquiry-based strategies, the extent to which it leads to changes in science instruction over the long term is unclear. The majority of studies about the effects of professional development on science instruction focus on the time period directly following the program activities. Few studies examine the longevity of professional development outcomes in science. In a small longitudinal study of six middle school science teachers who participated in a whole-school reform program, researchers found that changes in teachers' classroom practices were sustained over a 3-year period following the program (Johnson, Onwuegbuzie, & Turner, 2007). In a survey-based study of K–12 teachers who participated in a large-scale, statewide mathematics and science initiative, researchers found that teachers who received professional development in science showed sustained gains in their attitudes and preparation but significant declines in their inquiry-based teaching practices during the second and third years after the professional development (Supovitz, Mayer, & Kahle, 2000). These declines contrasted with sustained gains for mathematics teachers.

This study addresses the need for more research on persistence of teacher outcomes and focuses on a group less studied in science education—K–2 teachers. The research covers a 5-year time span and investigates changes in teachers' knowledge, attitudes, and instructional practices in science. The study also specifically examines contextual factors that may influence longevity of teacher outcomes and that may shift after professional development ends.

Method

Professional Development Program

This study focused on K–2 teachers who had completed a 3-year professional development program that provided science and technology assistance for teachers in rural districts. The professional development included three key components: (*a*) intensive adult-level science content instruction, (*b*) pedagogical training focused on science instruction and how to connect science to language arts and mathematics, and (*c*) training and support designed to facilitate teacher collaboration. The program provided teachers with over 100 contact hours each year and included intensive summer institutes, regional meetings, and school site sessions. The content instruction during the summer institutes focused on a different branch of science each year (physical, earth, and life sciences) and was based on topics included in the California state science standards. A team that included a university professor with expertise in science and advanced mathematics, an elementary teacher with expertise in research-based instructional strategies and science inquiry, and an English language learning specialist led the 6-day summer institutes. Following each summer institute, teachers participated in regional meetings as well as sessions at their schools during the academic year. The authors of this paper were not involved in designing or providing the professional development.

The pedagogical component, which emphasized scientific inquiry, was intertwined with the adult-level content instruction. Teachers learned science through research-based instructional strategies that included hands-on experiments and investigations. The team of instructors focused on scientific inquiry as a way for teachers to learn science content and to experience pedagogical strategies that they could use in their own classrooms. After experiencing instructional approaches as learners, teachers had a model for implementing these strategies. The program introduced teachers to scaffolded guided inquiry (SGI). The central aim of the guided inquiry process is to promote learning through student investigation. The inquiry-based instruction mirrors scientific methods and engages students in higher-level thinking and science process skills such as making predictions, summarizing knowledge, analyzing data, and evaluating their findings. SGI lessons are meant to take place over several days and include nine steps: (*a*) display the big idea, (*b*) gather needed materials, (*c*) discuss engaging scenario, (*d*) identify a focus question, (*e*) make a prediction, (*f*) collect data, (*g*) make claims based on evidence, (*h*) draw a conclusion, and (*i*) reflect. As part of the inquiry process, students are supposed to write in science notebooks and record their focus question, prediction, data, claims and evidence, conclusions, and reflections. The program also helped teachers develop inquiry-based science units through the use of a curriculum-mapping tool. The tool provides a process for documenting curriculum, planning for implementation of standards, and matching assessment with instruction. As part of the pedagogical component, teachers also learned instructional strategies shown to be effective for English Language Learners and approaches for integrating science instruction with mathematics and language arts.

To enhance collaboration among teachers in these rural school districts, the program created opportunities for teachers to work together. For example, during summer institutes, teachers worked in teams to develop curriculum maps for various grade levels. With the aim of creating professional learning communities, the program organized teachers into cluster groups according to the geographic proximity of their schools. During the school year, teachers participated in regional "cluster" meetings where they could reunite to discuss implementation, share strategies, and plan events. The program also hosted a website for teachers to communicate, share lesson plans, and access instructional resources.

Participants

The professional development program began with 44 K–2 teachers from 16 schools in 16 districts in four counties in northern California. Half of the districts were one-school districts in which a particular grade level may have only one teacher. Student enrollment prior to the program ranged from 148 to 5,087, and the poverty level ranged from 11% to 30% of families. Student performance on standardized tests indicated low academic achievement. Similar to national demographics for elementary teachers (National Center for Education Statistics, 2011), the teachers were predominantly white females with varying years of teaching experience and university coursework in science. Due to changes in teaching assignments, relocations, death, and attrition, there were 34 participating teachers by the end of the 3-year program. The follow-up research project included 30 of those teachers, representing 14 schools and 13 districts. Their teaching assignments, in the second academic year after the program ended, were: five in kindergarten, seven in grade 1, 12 in grade 2, two in combination grades (K–1–2, 2–3), and four in upper grades (4, 5, 6).

Data Sources

Data sources for the study included: (*a*) a teacher survey, (*b*) a science teaching self-efficacy assessment, (*c*) a content knowledge test, (*d*) interviews, and (*e*) classroom observations. Teachers completed the teacher survey and the efficacy assessment at five points: before participating in the program, in the spring of each academic year during the 3-year program, and in the second academic year after funding ended. The teacher survey, developed by Horizon Research (2000) for national studies of science teaching at the elementary level, focused on teachers' opinions about science and science instruction, their preparedness, and their instructional practices. Reliabilities are .80 for each composite variable of the instrument (Germuth, Banilower, & Shimkus, 2003). The self-efficacy assessment, the Science Teaching Efficacy Beliefs Instrument (STEBI) developed by Riggs and Enochs (1990) specifically for use with elementary teachers, focused on teachers' beliefs about their effectiveness in teaching science. The instrument includes two subscales—the Personal Science Teaching Efficacy Belief Scale (PSTE) and the Science Teaching Outcome Expectancy Scale (STOE). The PSTE subscale measures teachers' beliefs in their own abilities to teach science and the STOE subscale measures teachers' beliefs about the extent to which student learning depends on effective teaching. Reliabilities for the two subscales are .91 for the PSTE and .73 for the STOE (Riggs & Enochs, 1990).

Teachers' content knowledge was measured at the end of the program and again 2 years later. The content knowledge test was a cumulative content test that included a total of 68 items related to physical science (25 items), earth science (25 items), and life science (18 items). The cumulative test consisted of multiple-choice questions taken from the original content tests that teachers completed over the course of the project. The original test items were developed by the university professors who assisted with instruction at the summer institutes, with feedback from assessment experts at WestEd.

For this follow-up study, researchers interviewed 28 of the 30 participating teachers during the second academic year after funding ended. Two teachers were unavailable for interviews due to extenuating personal circumstances. Researchers conducted interviews in person or via telephone. Each interview lasted approximately one hour and was audio-recorded and subsequently transcribed. The semistructured interviews focused on instructional time in science, confidence in teaching science, content knowledge in science, instructional and curricular choices, integration of science into other subjects, and support and resources for teaching science. The questions probed for changes during the time after the professional development ended and reasons for any reported changes. Classroom observations of a strategic sample of teachers provided additional data about science instruction and pedagogical strategies. From those teachers who had been observed during the 3-year program, researchers selected 10 teachers who represented a range of schools and grade-level assignments for classroom observations during the follow-up study. Researchers took notes, collected relevant documents, and used a rubric to rate lessons on instructional strategies promoted in the professional development such as facilitated exploration, inquiry, literacy strategies, and content integration.

Data Analysis

The study employed a mixed-methods design to take advantage of the complementary strengths of quantitative and qualitative approaches (Johnson & Turner, 2003). We adopted a quantitative dominant concurrent triangulation strategy (Creswell, 2003), in which the quantitative surveys and qualitative interviews and observations occurred in the same phases of the research. The primary data sources were the survey, self-efficacy assessment, and content tests. The data from interviews and observations were used for both elaboration of participants' perceptions and classroom instruction and for triangulation of findings.

For all survey subscales, responses were converted to a numeric value and negatively worded questions were reverse coded. Ratings for each question were then averaged within each teacher and year. On the self-efficacy assessment, teacher responses were converted to numeric values on a 5-point scale (i.e., 1 = "strongly disagree," 2 = "disagree," 3 = "uncertain," 4 = "agree," 5 = "strongly agree"). Data were analyzed across the measure's two subscales (Riggs & Enochs, 1990): the Personal Science Teaching Efficacy Belief Scale and the Science Teaching Outcome Expectancy Scale. To examine the time teachers spent teaching science, we used questions from the teacher survey (Horizon Research, 2000) that focus on science instructional time (see Table 1). Since each question consisted of different response categories, we developed a composite score across the five questions. The composite was derived by quantifying each response category in each question such that they aligned to a 0–1 scale. These scores were then averaged across all questions to determine the composite. To examine teachers' use of instructional strategies in science and student activities during science instruction, we used relevant questions from the teacher survey (see Table 1). Teacher ratings were converted to numeric values on a 5-point scale (i.e., 1 = "never," 2 = "rarely," 3 = "sometimes," 4 = "often," 5 = "all or almost all"). To examine teachers' perceptions of principal support, we examined a series of questions from the teacher survey (see Table 1). Teacher ratings were converted to numeric values on a 5-point scale (i.e., 1 = "strongly disagree," 2 = "disagree," 3 = "no opinion," 4 = "agree," 5 = "strongly agree").

To examine the rate of change across the study years for each outcome measure, we used piecewise linear models (O'Connell & McCoach, 2008; Raudenbush & Bryk, 2002). These models can directly assess the rate of change during and after the program (i.e., before or after 2011) by allowing the linear slopes to vary between these two time periods. For these analyses, we included data from 28 of the 30 teachers. The other two teachers did not submit completed surveys and assessments for at least three of the five administrations. The piecewise linear modeling analyses were conducted using the lme4 package in R (Bates, Mächler, & Bolker, 2011). The modeling technique utilized maximum-likelihood estimation, which allows for missing data at random (which ranged from 4%–11% across each outcome measure).

In all piecewise models, changes in the outcome measures were modeled using two fixed-effects terms: (*a*) the program term, which represents the linear change across the preprogram and program years (i.e., 2008–2011), and (*b*) the postprogram term, which represents the linear change across the end of the program to 2 years beyond the program (i.e., 2011–2013). The piecewise models included teacher random effects on all program terms. To evaluate these piecewise models, we compared them to linear mixed-effects models that included a single fixed-effect term for year and teacher random effects on year. The latter model assumes the same linear change across both the program and postprogram years. To correct for mul-

Table 1. Items From the Teacher Survey

Category	Survey Item
Instructional time in science	Number of science lessons typically taught per week
	Length of a typical science lesson
	Number of science units taught
	Length of typical science units (number of weeks)
	Number of days out of last five that science was taught
Instructional strategies in science	Introduces content through formal presentations
	Demonstrates a science-related principle or phenomenon
	Teaches science using real-world concepts
	Arranges seating to facilitate student discussion
	Uses open-ended questions
	Requires students to support their claims with evidence
	Encourages students to explain concepts to one another
	Encourages students to consider alternative explanations
	Allows students to work at their own pace
	Helps students see connections between science and other disciplines
	Uses assessment to gauge what students know before or during a unit
	Embeds assessment in regular class activities
	Assigns science homework
	Reads and comments on students' reflections
Student activities during science instruction	Participates in discussions with teacher
	Works in cooperative groups
	Works on solving a real-world problem
	Shares ideas or solves problems in small groups
	Engages in hands-on activities
	Designs or implements their own investigation
	Works on extended science investigations
	Records, represents, and/or analyzes data
	Writes reflections in notebook
	Uses mathematics to solve problems
Principal support for science	Encourages selection of science content and strategies that address individual learning styles
	Encourages implementation of national standards in science education
	Encourages innovative instructional practices
	Provides needed materials and equipment
	Provides time for teachers to meet and share ideas
	Encourages making connections across disciplines
	Acts as a buffer between teachers and external pressures

tiple comparisons, we applied the Benjamin-Hochberg correction across all model comparisons.

To examine changes in teachers' science content knowledge, we examined the results of the cumulative test administered at two time points: at the end of the program and again 2 years later. One question in the life science portion was dropped from analysis because it was inconsistent across the 2 years. We analyzed results for all 30 teachers. Teachers' accuracy on the content tests were converted to a proportion between 0 and 1. We used mixed-effects linear regression to analyze the content knowledge scores by topic across the 2 years.

Analysis of the interview data followed qualitative research procedures including coding and data displays (Bogdan & Biklen, 1998). All verbatim interview transcriptions from the follow-up study were compiled in an electronic database and coded using qualitative analysis software. The transcriptions were coded according to a system of a priori codes generated from the conceptual framework and the protocol for teacher interviews after the program ended. During the coding process, emergent subcategories were added. The software allowed for multiple codes to be assigned to interview excerpts, which aided in searching for and retrieving data across categories. For example, teachers' descriptions of science instruction in their classrooms may have included information related to multiple codes such as "strategies used," "scaffolded guided inquiry," "changes in strategies," and "factors influencing strategies." The code "lack of" could be assigned as a double code to any of the categories. To establish reliability in coding, we used a system of multiple coders for each transcript. All coders received training in the coding system and demonstrated competence in applying the codes on a sample transcript. Two coders then independently coded each interview transcript. Any discrepancies were noted, discussed, and resolved. A third coder subsequently reviewed all transcripts for consistency with the system.

As a secondary data source, the interviews provided a means to explicate the survey findings and to provide a more elaborated understanding of teachers' instruction (Johnson et al., 2007). To achieve these purposes, researchers created data displays by searching across the database for specific codes or combinations of codes. Researchers analyzed the data in these displays to determine the fit of teachers' responses with the survey results and discern the range of teachers' perspectives and reasoning. Teachers' interview responses offered information about teachers' reasons for curricular choices and instructional strategies in science. Researchers focused on teachers' descriptions of changes in the years after the program ended and contextual factors that influenced their decisions. Representative excerpts from the interviews are included in the findings section to highlight both differences from, and similarities with, the statistical analyses as well as the range of teachers' responses. For example, the interview responses indicate that teachers held differing perspectives about their retention of content after the program ended but also indicate that they did not identify subject matter knowledge as a reason not to teach science.

Researchers analyzed classroom observations as another secondary data source. Classroom observations provided information about the way in which teachers taught science, integrated science with other subjects, and implemented strategies such as inquiry-based lessons and investigations. For example, in interviews, teachers talked about adapting the scaffolded guided inquiry model due to time constraints. Data from classroom observations illustrated the specific ways in which teachers made adaptations. Given the scheduled and bounded nature of the observations, they did not provide information about aspects such as the extent to which teachers used these instructional strategies over the school year.

Results

This longitudinal study reports on the extent to which the teacher professional development program led to changes that persisted beyond the 3-year program period. In the following sections, we examine the persistence of four categories of outcomes: content knowledge, self-efficacy, instructional time, and instructional strategies. We then discuss contextual factors that supported or hindered the sustainability of teacher change.

Persistence of Outcomes

Content knowledge. Comparison of teachers' scores on the cumulative test indicate that teachers' content knowledge of science remained steady between the end of the program ($M = .70$, $SD = .11$) and 2 years later ($M = .69$, $SD = .11$). The full model included year and topic as fixed effects, with teacher random effects on year and topic (log-likelihood = 119.55). This model was statistically different from an unconditional model that included only the random effect of teacher ($\chi^2 = 38.05$, $p < .001$; log-likelihood = –100.5, respectively) and was not statistically different from a model that included an interaction between the fixed effects in the full model ($\chi^2 = 1.78$, *ns*; log-likelihood = –120.44). The fixed-effects estimates revealed that there was no effect of year (*Estimate* = –.004, $SE = .008$, $p = .63$), but that overall, teachers performed statistically better on earth science relative to life science (*Estimate* = –.09, $SE = .02$, $p < .001$) and physical science (*Estimate* = –.06, $SE = .02$, $p = .004$). Data from our prior research indicate that teachers' content knowledge in each branch of science was highest directly following each summer institute, declined during the subsequent academic year, but remained higher than before teachers' participation in the program (Sandholtz & Ringstaff, 2013b).

In interviews, teachers readily acknowledged that the professional development had built their science knowledge, but they had differing views about their retention since the program ended. Whereas some teachers declared that their content knowledge had not changed at all or that they "remembered a lot of what I learned," others stated that they "probably have forgotten some" or that it definitely "declined, which kind of scares me." Some teachers suggested they had retained more content knowledge in some branches of science than others. For example, a second-grade teacher stated that she knew the physical science content, but had "lost so much of the life science." In contrast, a kindergarten teacher felt she remembered less physical science because it had been taught during the first year of the program. Another teacher commented, "If it is a concept that I don't use in the classroom, it probably did leave me. Especially physical science, which is always difficult for me anyway. . . . If you don't use it, you lose it!"

Although teachers held differing perspectives about their retention of the content, they did not refer to lack of subject matter knowledge as a reason to not teach science. Teachers noted that the professional development had given them a valuable foundation of content knowledge and a "good framework and understanding of key scientific concepts." They indicated that, with this background knowledge, they now were able to answer students' questions better and could review their notes before teaching a unit. As one teacher reflected, "I think that is part of teaching. You are not going to remember all the content you have learned over the years. So we research before we teach the lesson again. . . . I have to go back and refresh my memory. I have to do that with other subjects too."

Another teacher pointed out that she could not remember some specific information "like the thickness of the Earth's crust" but that she could readily find that type of information. The teachers' perspectives about their science knowledge appeared to matter more than their actual scores on the content knowledge test. A kindergarten teacher summarized the impact of the professional development on teachers' science knowledge in this way: "I think that you know your subject matter a little more than you did before, or at least you think you know it more than before. Before I would think, 'I don't know anything about physical science.' . . . I used to think, 'Oh man, this is over my head and over their head and there's nothing I can do about it.' But you get to know the three areas a bit more and you realize there are a lot of things I can do, even in physical science, with young children and not just think, 'Oh, they'll get it when they get to fourth grade.'" Teachers, particularly those with limited college and high school coursework in science, credited the program with building their knowledge of science.

Self-efficacy. Teachers' self-efficacy in teaching science generally declined after the professional development ended but remained higher than preprogram levels. For example, the percentage of teachers in the follow-up study who indicated they understood science concepts well enough to teach elementary science effectively was 71% preprogram, 89% end-of-program, and 86% 2 years later. The percentage of teachers who reported knowing the steps necessary to teach science effectively shifted from 38% preprogram to 100% end-of-program to 86% 2 years later. Before the program, only 8% of the teachers felt very well prepared to lead a class of students using investigative strategies. This percentage increased to 80% by the end of the program but decreased to 63% 2 years later. Similarly, the percentage of teachers who felt very well prepared to encourage students' interest in science shifted from 20% preprogram to 85% end-of-program but declined to 70% 2 years later. We found a similar trend in teachers' overall sense of preparedness to teach science. In contrast to 71% preprogram, no teachers in the follow-up study reported feeling inadequately or only somewhat prepared to teach science at end-of-program or 2 years later. However, the percentage of teachers who felt very well prepared to teach science dropped from 89% end-of-program to 38% 2 years later. One area in which teachers' sense of preparedness remained stable after the professional development ended was managing a class of students engaged in hands-on/project-based work. The percentages shifted from 16% preprogram to 80% end-of-program and 78% 2 years later.

For both personal science teaching efficacy and science teaching outcome expectancy, the data were fit using the piecewise linear model and the linear model. For personal science teaching efficacy, the piecewise linear model produced a better fit than the linear model ($\chi^2 = 37.02$, $p < .001$; log-likelihoods = –24.19 and –42.70, respectively). The piecewise model indicated a significant positive rate of change across the program years ($Estimate_{program} = .21$, $SE_{program} = .03$, $p < .001$). However, there was a negative rate of change across the postprogram years ($Estimate_{postprogram} = -.078$, $SE_{postprogram} = .03$, $p = .01$). The estimates suggest that teachers' personal science teaching efficacy ratings decline each year after the program at approximately a third of the rate that they increase each year during the program (see Fig. 1).

For science teaching outcome expectancy, the piecewise linear model produced a better fit than the linear model ($\chi^2 = 18.93$, $p < .001$; log-likelihoods = –33.04 and –42.50, respectively). The piecewise model indicated a significant positive rate of change across the program years ($Estimate_{program} = .14$, $SE_{program} = .03$, $p < .001$).

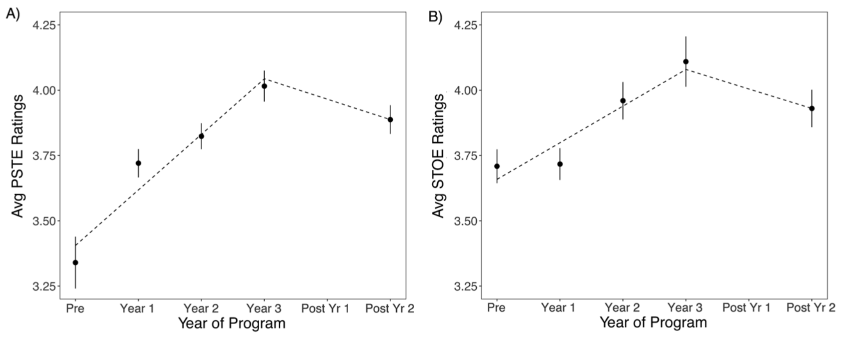

Figure 1. Mean (*A*) personal science teaching efficacy (PSTE) ratings and (*B*) science teaching outcome expectancy (STOE) ratings across each study year, fit with piecewise linear models. Error bars represent the standard error of the means.

However, the model indicated a significant negative rate of change across the postprogram years ($Estimate_{postprogram} = -.07$, $SE_{postprogram} = .03$, $p = .03$). The estimates suggest that teachers' science teaching outcome expectancy ratings decline each year after the program at approximately half the rate that they increase each year during the program. The model is displayed against the data in Figure 1.

In interviews for this follow-up study, teachers described substantial gains in their confidence in teaching science because of the professional development, and did not mention any decline after the program ended. Teachers indicated that their confidence had grown from the knowledge they gained, the strategies they learned, and from teaching science lessons over the course of the project. One teacher pointed out that "it was tough at first" to learn and implement the scaffolded guided inquiry approach, but "once you got it down, it made it so much easier." Teachers explained that their confidence prompted a willingness to experiment more in science. For example, a second-grade teacher, who previously was inclined to "just follow a science workbook," described "being more confident to jump into science and not worry if an experiment doesn't work or it doesn't go as planned." Another teacher said that, much to her surprise, other teachers now think of her as "the science lady." Although they consistently referred to increased confidence in teaching science, many teachers admitted still feeling more comfortable teaching language arts and mathematics. As one teacher noted, "I have taught those subjects for 20-something years and I have only formally done science for 4 years or so."

Instructional time. In the years after the program ended, teachers continued to devote more instructional time to science than prior to the program. The instructional time data were fit using the piecewise linear model and the mixed-effects linear model. The piecewise linear model did not produce a better fit than the linear model ($\chi^2 = 5.65$, $p = .23$; log-likelihoods = 102.68 and 99.96, respectively). However, a visualization of the data (see Fig. 2) suggested that instructional time increases after the first year of the program and levels off thereafter. To examine this possibility, we reknotted the piecewise model such that the first term represented the rate of change from preprogram to the first year (i.e., first-year), and the second term represented the rate of change from the first year of the program

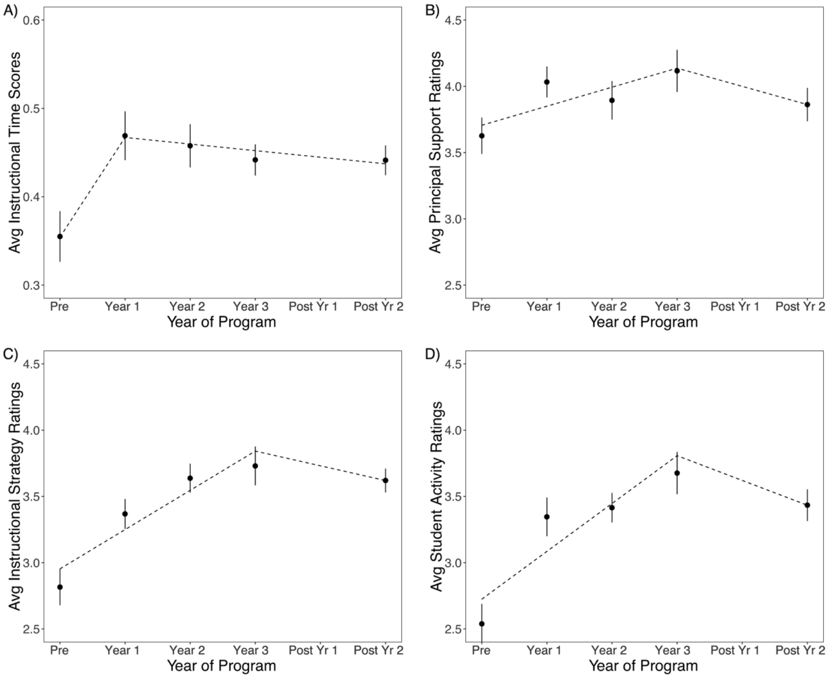

Figure 2. Mean score on (*A*) instructional time composite and mean teacher ratings of (*B*) principal support, (*C*) instructional strategies, and (*D*) student activities across each study year, fit with piecewise linear models. Error bars represent the standard error of the means.

to the second year after the program ended (i.e., post-first-year). This modified piecewise model produced a better model fit than both the original piecewise and the linear models ($\chi^2 = 28.07$, $p < .001$; log-likelihood$_{\text{modified-piecewise}} = 116.72$). The modified piecewise model indicated a significant positive rate of change across the pre- and first program years ($Estimate_{first\text{-}year} = .12$, $SE_{first\text{-}year} = .03$, $p = .002$), but the model indicated no significant changes after the first year of the program ($Estimate_{post\text{-}first\text{-}year} = -.01$, $SE_{post\text{-}first\text{-}year} = .01$, $p = .24$). Thus, the model suggests that instructional time in science stays relatively stable after the first program year with only slight declines.

In interviews after the program ended, teachers commented that they could not teach science as often as they preferred. In schools with designated time frames for teaching mathematics and language arts, teachers struggled to include science in their instructional day and tried various approaches. For example, teachers would "squeeze it in at the end of the day," or try to make "a designated day like Friday every week," or teach "social studies for 2 weeks and then science for 2 weeks," or focus on science for "2 months in the fall and 2 months in the spring." Even when teachers tried to assign a specific slot for science, they often found that science instruction depended "on how much time we had that day" or "how we get through the other activities." Teachers described challenges in not only fitting science into the instructional day but also implementing fully some of the hands-on and investigative teaching strategies they learned in the professional development.

Instructional strategies. The instructional strategies data were fit using the piecewise linear model and the linear model. The piecewise linear model produced a better fit than the linear model ($\chi^2 = 23.97$, $p < .001$; log-likelihoods = –123.94 and –135.92, respectively). The piecewise model indicated a significant positive rate of change across the program years ($Estimate_{program} = .30$, $SE_{program} = .05$, $p < .001$). The model indicated a negative slope in the years after the program; however, this estimate was marginally significant ($Estimate_{postprogram} = -.11$, $SE_{postprogram} = .06$, $p = .06$). Thus, after the professional development ended, teachers' use of reform-based instructional strategies in science declined slightly but remained higher than before the program. As the estimate was marginally significant, data from future time points will increase the confidence in this conclusion. The model is displayed in Figure 2.

After the program ended, teachers continued to use a range of student-centered activities in science, but the reported frequency declined. The data about student activities were fit using the piecewise linear model and the linear model. The piecewise linear model produced a better fit than the linear model ($\chi^2 = 23.97$, $p < .001$; log-likelihoods = –123.94 and –135.92, respectively). The piecewise model indicated a significant positive rate of change across the program years ($Estimate_{program} = .36$, $SE_{program} = .06$, $p < .001$). However, the model indicated a significant negative rate of change across the postprogram years ($Estimate_{postprogram} = -.19$, $SE_{postprogram} = .07$, $p = .01$). The estimates suggest that the frequency of student-centered activities during science instruction decline each year after the program at about half the rate that they increase each year during the program. The model is displayed in Figure 2.

In contrast to the survey results, teachers talked extensively about using student-centered activities and elements of the scaffolded guided inquiry (SGI) model when describing their science instruction in interviews after the program ended. Although teachers noted that it took practice and repeated efforts for them to become comfortable using the model, their comments highlighted benefits for students:

> I think that SGI is an excellent way to prepare kids to understand the science hypotheses procedures that they will get at later grades. They are going to hear it over and over and this is a great way at this level to introduce it.

> It works, especially with the vocabulary with the students. They use a lot of higher-level knowledge with having to collect the data themselves and then using the data. There is a lot of comprehension going on too, and then they are able to reflect [on questions such as] "I wonder what would happen if . . ."

> They discuss science and vocabulary that they use; they are not just regurgitating. They are synthesizing the information and making predictions. It is not really all about observation; it's about making predictions based on the observations. . . . It has a huge impact on their higher-order thinking. . . . I think it really motivates them and helps them to be more inquiring in other areas.

In contrast to one teacher who stated that she used "100% all of it," the majority of teachers described adapting the SGI model due to the amount of instructional time needed to complete all of the components. Their adaptations varied according to their perceptions of the balance between benefits and challenges. For example, one teacher sometimes would "leave off the engaging scenario, but we definitely do the reflections because I think that is where most of the learning would be." In contrast, others described the reflection pieces as "really challenging to do with first graders" and "not always getting to the reflections." Teachers found that kindergarten students often did not have the writing or drawing skills, particularly early in the year, to "do a journal where they are writing freely to come up with their own ideas . . . even their drawings look like a blob on a piece of paper and you are not sure what they are trying to get at." Consequently, they adapted the task or waited until later in the year to incorporate student notebooks. Teachers at every grade agreed on the value of having students make predictions, but whereas some teachers continued to have students formally write down their predictions, others had students share orally instead.

During classroom observations, researchers noted teachers adapting the SGI model. Some teachers gave students handouts that helped to scaffold the information that students needed to provide in each category: focus question, prediction, data collection, claims and evidence, conclusion, and reflection. For example, for claims and evidence, one teacher had students complete the sentences: "I claim that the ______. I claim this because______," and instructed them to "include a picture to match your sentence." For the conclusion and reflection categories, another teacher included the following on a student handout:

> Conclusion:
> The problem we had to investigate was: What is found in soil?
> We predicted that the soil contains: ______________.
> We found that soil contains: ______________.
> Reflection:
> I wonder ______________.

In classrooms where teachers continued to use the full model, students included a kit inventory in their notebooks in which they listed all of the items needed for the investigation and wrote about the procedures and findings without any fill-in-the-blank scaffolding.

In describing their instructional strategies in science, several teachers specifically stated that they had "not gone back to the textbook," and others described using the textbook as one resource among others. Teachers described hands-on activities and inquiry-based investigations as not only more engaging for students but also better for student learning. Only one teacher discussed reverting back to textbook-based instruction after the program ended because "it was easier for me." She stated, "I went back to just looking at my teachers' manual and using the resources that were there. So pencil and paper and let's read the book. . . . I tried to put in a few hands-on experiments, but they weren't really linked to the big idea." After determining that the students "weren't learning anything," she again implemented strategies learned in the program.

During classroom observations conducted after the professional development ended, researchers observed the ways in which teachers used inquiry-based investigations during science instruction. For example, in a kindergarten lesson on weather, the teacher began by showing students a book jacket and asking various "what" and "why" questions that uncovered students' inferences about the weather as well as their reasoning. Throughout the lesson, the teacher questioned students and pushed them to justify their answers. Later the students went outside to gather information about the weather to include on their weather charts and temperature graphs. In a first-grade class, students conducted an experiment about states of matter in which they made predictions and recorded their observations each day. One day, after taking the "pops" from the freezer, students noticed that one was frozen and hard, but another one was still liquid. The teacher commented, "We got something we didn't expect. Scientists love it when they get something they didn't expect." After exploring different ideas and options, they decided to put the "pops" into a larger freezer in the cafeteria and check them again in 2 days. In notebooks, students revised their procedures section accordingly. In a second-grade class, students wrote predictions about what they would find in soil. Using magnifying glasses, they examined soil samples, drew pictures of what they found, and labeled their drawings. As students reported their findings, the teacher started to make a word bank on the board. Later in the week, the students determined how plants, animals, and humans obtain necessary minerals. Across classrooms, teachers implemented inquiry-based instruction in science, but time constraints limited how often and the extent to which they used the complete SGI model.

Contextual Factors

Contextual factors varied across schools and both supported and impeded teachers' actions related to science instruction after the program ended. The most important factors that influenced teachers were resources, curricular demands, administrative support, and collegial support. In terms of resources, time was the most critical resource that influenced teachers' science instruction, but teachers also indicated an ongoing need for materials and supplies. The percentage of teachers who reported that their schools were well supplied with materials for investigative science instruction increased from 38% preprogram to 72% end-of-program but declined to 56% 2 years later. The availability of equipment and supplies varied across schools. For example, a teacher at one school described having "pretty much whatever we want to purchase or have access to," while a teacher at another school stated, "There aren't any [resources]. Everything I do, I go out and make on my own, buy on my own, collect on my own." Teachers did not need expensive equipment but rather replenishment of materials such as eye droppers or cups. As one teacher noted, "We have them in our kits but as you keep using the kits every year, things tend to disappear or you use them up."

Teachers needed time for preparation of lessons, collaboration with other teachers, and classroom instruction of science. On the survey, they consistently disagreed or strongly disagreed with the statement about having time during the school week to work with peers on science curriculum and instruction. The percentage shifted only slightly over a 5-year period, from 71% preprogram to 72% end-of-program to 73% 2 years later. Similarly, when asked in interviews during the follow-up study about teaching science or integrating English language strategies into science or using student notebooks, teachers repeatedly mentioned time as the major barrier. As one teacher phrased it, "Time, as always." Beyond the time needed for planning and preparation, teachers often struggled to find time to teach science during the school day. Curricular demands, particularly in mathematics and language arts in the early grades, inhibited instructional time for science. Across schools, teachers felt the press to focus on mathematics and language arts, particularly since those subject areas are emphasized on the standardized tests. As one teacher put it, "They don't tell us we can't teach science, but the definite emphasis is you better get the others done first." In schools designated by the state as program improvement sites, which meant not meeting statewide proficiency goals for 2 consecutive years and being subject to improvement and corrective action measures, teachers encountered strict daily time schedules that lessened time for science:

> Our schedule says that we can do social studies, science, or PE in a 20-minute slot at the end of the day. Those are the three things we can teach in that slot. We're supposed to squeeze it in. So we have 20 minutes to teach those three things each day. We are a program improvement school, so I'm sure that's part of why the pacing is so strict.

> Now everything is structured to such minute detail, as far as time and subject and curriculum and actual materials to use, that you are only allowed a certain amount of time to do anything else.

In schools without the rigid schedules, teachers noted curricular demands from specialized programs as well as the emphasis on mathematics and language arts. As one teacher described, "The only barrier is time. Because reading and math is so consuming and you have to read in math. And then there are all of the other pull-out programs and computer labs. There are lots of things that we offer which are wonderful but also make it difficult to find blocks of time." Teachers at some schools, but not others, could integrate science with language arts and use designated English Language Development (ELD) time. A kindergarten teacher stated, "I actually cut out my journaling time to add extra science time because I figured they were writing anyway, so that's how I made it happen. Whenever we were doing our science notebooks, I would cut out writing in the morning because I figured they were writing in the afternoon." In contrast, teachers at another school discovered that the district would not allow them to deviate from the established ELD curriculum: "A lot of us said, 'Why can't we just teach science during ELD? These kids need science. They need language development. Why can't we do it at the same time?' But we couldn't. . . . I still had to have separate ELD."

After the professional development ended, teachers particularly missed the collegial support and opportunities to collaborate with other participants. They commented on the benefits of "getting together with the teachers and bouncing ideas off of each other," "creating lessons and talking to other teachers," and "brainstorming about the news things coming." At some sites, teachers enthusiastically described working in teams, but at others, teachers simply longed to "be able to talk to somebody about science." Although the program aimed to sustain collaboration with other participants via its website, the collaboration did not materialize.

Support from school principals also varied across schools, but generally lessened after the professional development ended. At some sites, newly assigned principals made it "more challenging to include science" because science was not a priority. Teachers stated that the administrators' "focus is on CSTs [California Standards Tests], language arts, and math with little or no emphasis on science." For instance, one principal decided to switch the family science night to a family math night. Some principals were very supportive of science instruction but had no funding for science equipment and supplies. In contrast, other principals would provide funds for materials or "whatever I want to go to that has to do with science." At one school, during discussions about lengthening the school day, teachers highlighted the need for more science instruction, and the principal responded by hiring a science resource teacher.

The survey data about principal support were fit using the piecewise linear model and the linear model. The piecewise linear model produced a better fit than the linear model ($\chi^2 = 23.97$, $p < .001$; log-likelihoods = –123.94 and –135.92, respectively). The piecewise model indicated a significant positive rate of change across the program years ($Estimate_{program} = .14$, $SE_{program} = .06$, $p = .03$). The model indicated a negative rate of change across the postprogram years; however, this estimate was only marginally significant ($Estimate_{postprogram} = -.14$, $SE_{postprogram} = .08$, $p < .10$). Although the rate of decrease after the program is similar to the rate of increase during the program, the negative rate of change postprogram is only marginally significant. This difference in significance may be due in part to increased variability after the program ended. Teachers' interview responses also indicate that principal support varied considerably across school sites. Data on future time-points will inform the rate of decrease and the level of significance. The model is displayed in Figure 2.

Discussion

The results of this study highlight both the potential of professional development to enhance science instruction in the early elementary grades and the need to address issues of sustainability after professional development ends. The design of the 3-year program was consistent with core recommendations for effective programs (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Garet, Porter, Desimone, Birman, & Yoon, 2001; Hawley & Valli, 1999). The professional development was sustained and intensive, connected to district and school goals, focused on curriculum content rather than abstract principles, based on active learning by participants, and aimed to cultivate collegiality. In addition, the program developers surveyed participating teachers and considered their specific needs. Previous research on this program's impact showed substantial changes in teachers' content knowledge, self-efficacy, and instructional practices in science after just one year of professional development and found that these changes were largely sustained in the second and third years (Sandholtz & Ringstaff, 2011, 2013a, 2013b). However, for a lasting impact on K–2 science instruction, gains from professional development need to be sustained over an extended period of time.

Our first research question in this study focused on the persistence of these changes beyond the 3-year period of professional development. Our findings indicated a beginning pattern of decline, but across categories the outcomes remained higher than before the professional development. Teachers' scores on a cumulative content knowledge test remained stable during the 2 years after the program ended; the primary decline in content knowledge occurred in the academic year following each summer institute. Teachers' self-efficacy in teaching science declined after the program ended but continued to be higher than preprogram. Instructional time in science was highest at the end of the first program year and then remained fairly stable with only slight declines. Teachers continued to use a broader range of instructional strategies in science than preprogram but their reported frequency declined. Teachers' use of student-centered activities had the most significant decline in frequency after the program ended.

Our second research question focused on the extent to which school and district resources and networks provided adequate supports for teachers to continue to implement what they learned in the professional development. Our findings indicate wide variations across schools that influenced, in particular, the amount of time teachers devoted to science and their decisions about instructional strategies in science. The professional development bolstered teachers' confidence and interest in teaching science, and teachers saw benefits from implementing the inquiry-based instructional models they learned. But without the ongoing support that the program provided, teachers found it more difficult to devote the time they wanted to science.

Beyond offering teachers instruction in science content and pedagogy, the professional development provided three types of support not afforded by most of the schools. First, teachers regularly engaged in planning science instruction over the 3 years of the program. This designated and guided time for planning as part of the professional development established the foundation for a cycle of planning, implementation, and assessment. In grade-level teams, teachers developed lessons during the summer institutes that they subsequently used in their classrooms and then assessed and revised in subsequent meetings. But that planning time was not available to most of the teachers at their school sites throughout the professional development or in the years afterward. Consequently, when the professional development ended, a particularly valuable resource ended as well. Second, teachers felt a sense of accountability that prompted them to prioritize teaching science. As participants in a program focused on K–2 science instruction, teachers recognized at least an informal expectation that they would implement what they learned. When teachers collaboratively planned lessons during the summer institutes, they anticipated that they would discuss their experiences at later sessions. As one teacher commented, "Being a part of a program like [this one] and knowing I have to do this is really motivating." Another teacher talked about the process of selecting and submitting five students' science notebooks at the end of each year: "I knew that day was coming, so I was driven to make sure that I got to science and it was a priority for me. If it's not a priority for you, if it's not tested, if your administration is like, 'That's nice but we want to see the reading scores go up,' it's not quite as important." After the professional development ended, many of the principals offered little support for incorporating science into the K–2 curriculum and certainly did not hold the teachers accountable to teach science. Without formal or informal expectations of teaching science, teachers at some schools found it increasingly challenging to make science a consistent part of the instructional day. Their participation in the professional development had essentially given them a license to make science a priority. Third, the professional development involved collaboration with other teachers, a form of support that the teachers particularly valued. Half of the districts were one-school districts, often with only one teacher assigned to a grade level. During the professional development, teachers connected with participants from other schools and districts during summer institutes and regional meetings, and they collaborated in grade-level teams. In addition to sharing ideas, materials, and lessons, they shared a commitment to learn about and teach science. After the program ended, networking among participants generally ended as well. Program developers created a website with the aim of sustaining collaboration, but without some face-to-face interactions and meetings, electronic communication diminished. At individual schools, collegial support varied widely.

In interviews after the program ended, teachers talked about resources they needed to continue to teach science and to implement the instructional strategies they learned. They frequently recommended some type of ongoing professional development, not as extensive as the 3-year program but enough to fulfill needs related to planning and teaching science:

> I think just, like some refreshers, kind of something that goes back and like revisiting a strategy or something, just to keep it kind of fresh in your mind.

> It would be nice to get together at least once a year. "Hey are you still doing this? What have you done this year?" I think that is the one thing we got out of every year, creating that SGI lesson and talking to teachers. "What did you do? Oh, that's a great idea. How did you get the resource for that?"

> Some kind of seminar or class just to review things again, or even just to share lesson plans or ideas. It's always nice to have somebody else to talk to, so you don't get stuck in a rut of teaching the same thing the same way every year.

Teachers also frequently recommended more opportunities for collegial interaction and collaboration. Their interactions with other participants served as a valuable source of ideas and also enhanced their motivation to teach science. As one teacher stated, "It would be nice to have collaboration on a few things. At least one other teacher that wanted to work together." Another teacher related that "having those other people that were excited about it really kept me motivated. . . . When you have someone else that was excited about teaching science, then you could share your ideas and share your lesson plans." Teachers also pointed out the value of having someone with expertise in science, perhaps from the county office, who could help guide their efforts. One teacher requested help from someone who could "offer ways to modify the SGI model" to fit within a shorter time period.

Over 3 years of professional development, the teachers increased their science content knowledge, demonstrated more confidence in teaching science, spent more instructional time on science, and incorporated more student-centered, inquiry-based instructional strategies (Sandholtz & Ringstaff, 2013b). They perceived benefits of including science in the K–2 curriculum and using different teaching approaches. But 2 years after the professional development ended, teacher outcomes were beginning to decline. Depending on their district and school contexts, teachers often lacked the ongoing supports to sustain the instructional changes they had made in science. Differences in school-level support were evident during the program period but became much more influential after the support from the program ended. Without modest interventions, the changes teachers made appear likely to decrease over time, particularly in program improvement schools that are under substantial pressure to increase student scores in mathematics and language arts.

Conclusion

This study highlights three key issues related to the sustainability of professional development outcomes in science. First, the findings of this study underscore the need to provide ongoing support for teachers in order to maximize the longevity of professional development outcomes. Even when professional development was well designed, extended over multiple years, was targeted to participants' needs, and led to significant changes, instructional changes in science began to decline after the program ended. Professional development is a key strategy for improving the status of science education in elementary schools (Banilower et al., 2013; CCST, 2010; NRC, 2011). Given the time and resources invested in high-quality professional development, it makes sense to look for ways to extend the positive outcomes. Similar to regular tune-ups for automobiles, the investment needed for sustainability may be minor in comparison to initial costs but pay dividends in terms of long-term function.

Second, this research demonstrates the importance of contextual factors in promoting and inhibiting long-term changes in instructional practices. Variations and shifts in district- and school-level support appear to exert substantial influence on whether or not instructional changes in science persist over time. Changes in school principals and their curricular priorities directly impact teachers' instructional decisions, particularly when teachers' external support from the professional development has ended. Similarly, changes in a school's state-level designation, such as program improvement status, can increase the emphasis placed on mathematics and language arts as well as the rigidity of daily and weekly time schedules. Beyond instructional time and schedules, administrative support affects teachers' choice of instructional strategies. The science lessons of teachers who perceive greater support from principals tend to be of higher quality and incorporate active student participation and problem-solving approaches (Bowes & Banilower, 2004). The majority of the teachers in this study continued to use the instructional strategies they learned in the program but not as frequently and with modifications due to time constraints. Over time, without school-level support from administrators, the teachers may find it even more challenging to teach science on a regular basis and use inquiry-based strategies consistently. Moreover, teachers in schools without a core group of program participants, or a teacher leader who takes an active role, may have insufficient collegial support to sustain instructional changes. Collegial support and opportunities for collaboration are key factors for teachers to not only try new instructional strategies but also to sustain them (Appleton & Kindt, 1999; Franke, Carpenter, Levi, & Fennema 2001). In rural settings and one-school districts, teacher collaboration is particularly vital to reform yet often difficult to maintain (Boyer, 2006; Harmon, Gordanier, Henry, & George, 2007).

Third, this study's findings and limitations highlight the need for more research about the sustainability of professional development. During their participation in the program, teachers in this study may have been inclined to overestimate instructional time on science and their use of instructional strategies in the self-report data. This inclination may have lessened after the professional development ended and influenced our findings. Classroom observations indicated teachers could implement inquiry-based approaches but did not include all teachers in the study and could not corroborate teachers' frequency of use over time. In addition, the small sample of teachers and the small, rural school settings may limit the generalizability of the findings to other populations of teachers. These limitations, as well as a beginning pattern of decline after the professional development ended, suggest a need for longitudinal studies in other settings. To design follow-up support, we need to know more about differences in school contexts and their specific impact on teachers' decisions about instructional time and strategies in science. Modest but targeted forms of follow-up support may provide enough reinforcement to sustain instructional outcomes over time, but we need research that examines this proposition. The main goal of professional development is to improve instruction and thereby foster student learning. But unless teachers are teaching science and using student-centered and inquiry-based approaches on a consistent basis, we cannot anticipate improved student learning in science.

Note

This material is based on work supported by the National Science Foundation under grant no. DRL-1119589. 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. We would like to thank Jonathan Nakamato for statistical consultation.
Judith Haymore Sandholtz is a professor in the School of Education, University of California Irvine, 3200 Education Building, Irvine, CA 92697-5500; e-mail: judith.sandholtz@uci.edu. Cathy Ringstaff is a senior research associate, and Bryan Matlen is a research associate, in the Science, Technology, Engineering, and Mathematics (STEM) program at WestEd in Redwood, CA.

References

Appleton, K., & Kindt, I. (1999). Why teach primary science? Influences on beginning teachers' practices. *International Journal of Science Education*, **21**(2), 155–168.

Banilower, E. R., Smith, S., Weiss, I. R., Malzahn, K. J., Campbell, K. H., & Weis, A. H. (2013). *Report of the 2012 National Survey of Science and Mathematics Education*. Chapel Hill, NC: Horizon Research.

Basista, B., & Mathews, S. (2002). Integrated science and mathematics professional development programs. *School Science and Mathematics*, **102**(7), 359–370.

Bates, D., Mächler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42.

Bogdan, R. C., & Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods* (3rd ed.). Boston: Allyn & Bacon.

Bowes, A. S., & Banilower, E. R. (2004). *LSC observational study: An analysis of data collected between 1998 and 2003*. Chapel Hill, NC: Horizon Research.

Boyer, P. (2006). *Building community: Reforming math and science education in rural schools*. Fairbanks, AK: Alaska Native Knowledge Network.

California Council on Science and Technology (CCST). (2010). *The preparation of elementary school teachers to teach science in California*. Sacramento, CA: Author.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Oxford, OH: National Staff Development Council.

de Laat, J., & Watter, J. J. (1995). Science teaching self-efficacy in a primary school: A case study. *Research in Science Education*, **25**(4), 453–464.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, **38**(3), 181–199.

Dorph, R., Shields, P., Tiffany-Morales, J., Hartry, A., & McCaffrey, T. (2011). *High hopes—few opportunities: The status of elementary science education in California*. Sacramento, CA: Center for the Future of Teaching and Learning at WestEd.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, **19**(6), 643–658.

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, **38**(3), 653–689.

Fulp, S. L. (2002). *Status of elementary school science teaching*. Chapel Hill, NC: Horizon Research.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, **38**(4), 915–945.

Germuth, A., Banilower, E., & Shimkus, E. (2003). *Test-retest reliability of the Local Systemic Change teacher questionnaire*. Chapel Hill, NC: Horizon Research.

Guskey, T. R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, **59**(6), 45–51.

Harmon, H., Gordanier, J., Henry, L., & George, A. (2007). Changing teaching practices in rural schools. *Rural Educator*, **28**(2), 8–12.

Hawley, W. D., & Valli, L. (1999). The essentials of effective professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 127–150). San Francisco: Jossey-Bass.

Heck, D. J., & Crawford, R. A. (2004). *LSC teacher questionnaire study: A longitudinal analysis of data collected between 1997 and 2003*. Chapel Hill, NC: Horizon Research.

Heck, D. J., Rosenberg, S. L., & Crawford, R. A. (2006). *LSC teacher questionnaire study: A longitudinal analysis of data collected between 1997 and 2006*. Chapel Hill, NC: Horizon Research.

Horizon Research, Inc. (2000). *Local systemic change through teacher enhancement science teacher questionnaire*. Chapel Hill, NC: Horizon Research.

Johnson, C. C. (2007). Whole-school collaborative sustained professional development and science teacher change. *Journal of Science Teacher Education*, **18**(4), 629–661.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, **1**(2), 112–133.

Johnson, R. B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 297–319). Thousand Oaks, CA: Sage.

Knapp, M. S. (2003). Professional development as a policy pathway. *Review of Research in Education*, **27**, 109–157.

Lakshmanan, A., Heath, B. P., Perlmutter, A., & Elder, M. (2011). The impact of science content and professional learning communities on science teaching efficacy and standards-based instruction. *Journal of Research in Science Teaching*, **48**(5), 534–551.

Luera, G. R., Moyer, R. H., & Everett, S. A. (2005). What type and level of science content knowledge of elementary education students affect their ability to construct an inquiry-based science lesson? *Journal of Elementary Science Education*, **17**(1), 12–25.

Math and Science Partnership Knowledge Management and Dissemination (KMD) Project. (2008). *Why teachers' science content knowledge matters: A summary of studies*. Retrieved from http://mspkmd.net/pdfs/blast18/3b1.pdf

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

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

National Center for Education Statistics. (2011). *Fast facts: Teacher trends*. Washington, DC: U.S. Department of Education.

National Research Council (NRC). (2011). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

National Staff Development Council. (2001). *Standards for staff development (revised)*. Oxford, OH: National Staff Development Council (NSDC).

O'Connell, A. A., & McCoach, D. B. (Eds.). (2008). *Multilevel modeling of educational data*. Greenwich, CT: Information Age.

Olson, S., & Labov, J. (2009). *Nurturing and sustaining effective programs in science education for grades K–8*. Washington, DC: National Academies Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (pp. 102–119). New York: Simon & Schuster Macmillan.

Riggs, I., & Enochs, L. (1990). Development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, **74**, 625–637.

Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. In J. Brophy (Ed.), *Advances in Research on Teaching* (Vol. **7**, pp. 49–74). Greenwich, CT: JAI.

Sandholtz, J. H., & Ringstaff, C. (2011). Reversing the downward spiral of science instruction in K–2 classrooms. *Journal of Science Teacher Education*, **22**(6), 513–533.

Sandholtz, J. H., & Ringstaff, C. (2013a). Assessing the impact of teacher professional development on science instruction in the early elementary grades in rural U.S. schools. *Professional Development in Education*, **39**(5), 678–697.

Sandholtz, J. H., & Ringstaff, C. (2013b, May). *Does professional development make a difference? Results of a 3-year study of K–2 science instruction and student learning*. Paper presented at annual meeting of the American Educational Research Association, San Francisco.

Sparks, D. (2002). *Designing powerful professional development for teachers and principals*. Oxford, OH: National Staff Development Council.

Stigler, J., & Hiebert, J. (1999). *The teaching gap*. New York: Free Press.

Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice. *Educational Policy*, **14**(3), 331–356.

Supovitz, J., & Turner, H. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, **37**(9), 963–980.

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, **68**(2), 202–248.

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.