Research Article

Effectiveness of a Curricular and Professional Development Intervention at Improving Elementary Teachers' Science Content Knowledge and Student Achievement Outcomes: Year 1 Results

Brandon S. Diamond,[1] Jaime Maerten-Rivera,[2] Rose Elizabeth Rohrer,[3] and Okhee Lee[4]

[1]*Department of Biology, University of Miami, Cox Science Center, 1301 Memorial Drive, Coral Gables, Florida 33146*
[2]*School of Education, University of Miami, Coral Gables, Florida*
[3]*Department of Sociology, University of New Mexico, Albuquerque, New Mexico*
[4]*Steinhardt School of Culture, Education, and Human Development, New York University, New York, New York*

Received 15 August 2013; Accepted 24 February 2014

Abstract: Teacher knowledge of science content is an important but under-studied construct. A curricular and professional development intervention consisting of a fifth grade science curriculum, teacher workshops, and school site support was studied to determine its effect on teachers' science content knowledge as measured by a science knowledge test, a questionnaire, and classroom observations. These three measures, along with college science courses taken, were then used to examine the effect of teachers' science content knowledge on student achievement outcomes. The intervention had a significant effect on the treatment group teachers' science knowledge test scores and questionnaire responses compared to the control group, but not on the classroom observation ratings. Teachers' scores on the science knowledge test were found to be the largest significant teacher-level predictor of student achievement outcomes regardless of participation in the intervention. © 2014 Wiley Periodicals, Inc. J Res Sci Teach 51: 635–658, 2014
Keywords: science content knowledge; teacher knowledge; professional development; elementary science; cross-sectional multilevel modeling; science achievement; teacher effect on student

As the cyclical nature of education reform returns to a mindset reminiscent of the Sputnik era, bringing the training of new scientists and engineers to a high priority status, researchers and policymakers are trying to determine how to best improve science education for all students. Surprisingly, very little research examines how to improve teacher knowledge of science content (Fleer, 2009; Heller, Daeler, Wong, Shinohara, & Miratrix, 2012; Shallcross, Spink, Stephenson, & Warwick, 2002), how to measure teacher knowledge of science content, or its impact on classroom practice and student achievement (Chinnappan & Lawson, 2005). In fact, on the rare occasions that teachers' science content knowledge (SCK) is addressed, it is usually studied in preservice teachers rather than those currently teaching (Ball, Lubienski, & Mewborn, 2001). There have also been few studies that focus on the content of professional development (PD) and how much content knowledge (CK) increases during PD opportunities (Garet, Porter, Desimone, Birman, & Yoon, 2001).

Contract grant sponsor: Institute of Education Sciences, U.S. Department of Education; Contract grant number: R305A090281.
Correspondence to: Brandon S. Diamond; E-mail: b.diamond@bio.miami.edu
DOI 10.1002/tea.21148
Published online 26 March 2014 in Wiley Online Library (wileyonlinelibrary.com).

This study used multiple measures to examine SCK of elementary school teachers over the 1-year duration of a curricular and PD intervention. The project was created as a collaboration between a university and a large urban school district to implement a new fifth grade science curriculum intended to maximize inquiry-based learning and understanding of science concepts by all students, especially English language learners (ELLs). This study evaluated the effectiveness of the intervention at improving SCK and included a control group of teachers who did not receive the intervention. The study also evaluated the relationship between teacher SCK and student achievement outcomes on a high-stakes science test. Specifically, the study examined two research questions, each with sub-questions tested by multiple measures:

(1) What effect did the intervention have on teacher SCK in the treatment group, compared to the control group, over the course of the school year, as SCK was measured by (a) a teacher science knowledge test, (b) a self-reported science knowledge questionnaire, and (c) classroom observations?
(2) What effect did teacher SCK have on student achievement outcomes on a high-stakes science test when SCK was measured by (a) a teacher science knowledge posttest, (b) a self-reported post-intervention science knowledge questionnaire, (c) classroom observations, and (d) college science courses taken?

Literature Review

Until recently, there was very little research available on practicing teacher SCK, and how it relates to student achievement. This is in spite of the fact that lack of SCK is often cited as a primary cause of the inability of teachers to teach science effectively (Fleer, 2009). In fact, variations of teachers' science knowledge and understanding have been identified as a main factor responsible for the differences in the quality of elementary science teaching (Shallcross et al., 2002).

In 1986, Lee Shulman introduced the concept of pedagogical content knowledge (PCK) as a way of understanding effective instructional practices for teaching specific subject matter in ways that students can understand (Ben-Peretz, 2011; Kaya, 2009; Shulman, 1986). Since then, PCK has been studied much more than CK, even though strong CK is necessary for strong PCK (Kaya, 2009; Van Driel, De Jong, & Verloop, 2002). The prominence of research on PCK may be because PCK is often described as being more directly related to teaching than CK. However, there is still a lack of literature describing the impact of teacher SCK on classroom practice and student achievement (Chinnappan & Lawson, 2005), so it is possible that teacher SCK may be as important as teacher science PCK for student learning.

Mathematics CK has been studied more extensively than SCK. Teacher knowledge of mathematics was one of the first variables of teaching effectiveness investigated in the 1960s (Ball et al., 2001). Adequate CK is necessary for "interpreting reform ideas, managing the challenges of change, using new curriculum materials, enacting new practices, and teaching new content" (Ball et al., 2001, p. 437). A study of PD showed that SCK was a major predictor of teachers' use of inquiry-based science teaching in the classroom (Supovitz & Turner, 2000). Yet there is still insufficient understanding of the CK it takes to teach effectively. Without knowing what kind of CK is necessary for effective teaching, it is difficult to help teachers develop the knowledge they need.

While there is much more literature examining the SCK of preservice teachers than the SCK of practicing teachers, there are some recent studies that examine teacher SCK in various contexts (Diamond, Maerten-Rivera, Rohrer, & Lee, 2013; Heller et al., 2012; Jüttner, Boone, Park, &

Neuhaus, 2013; Nowicki, Sullivan-Watts, Shim, Young, & Pockalny, 2013). The general trend seems to be that elementary school teachers tend to have major gaps in their SCK, and that these gaps are a major obstacle to effective teaching (Nowicki et al., 2013). This is, in part, due to poor science preparation in preservice teacher programs.

Professional development is important to improving the quality of U.S. schools, the effectiveness of policy changes for teachers and teaching practice, and gains in student achievement (Desimone, 2009). The impact of PD has typically been measured by either teacher outcomes or student outcomes. However, first, the impact of PD on teacher outcomes needs to be measured, and then the relationship between teacher change and student achievement outcomes should be examined for a more complete model to emerge (see Figure 1).

This section addresses three issues: (a) effect of PD on teacher SCK (Research Question 1), (b) effect of teacher SCK on student achievement outcomes (Research Question 2), (c) predictors of teacher SCK and student science achievement, and (d) measures of SCK.

Effect of Professional Development on Teacher Content Knowledge

The effects of PD on various teacher constructs and teaching practices have been examined extensively across subject areas, including science. The effect on SCK specifically, however, has rarely been addressed. The effectiveness of PD on teacher outcomes was studied using data from a Teacher Activity Survey conducted to evaluate the Eisenhower Professional Development Program (Garet et al., 2001). The study examined three structural features: the form of the activity (e.g., workshop vs. study group), the duration of the activity, and collective participation of teachers. The study also examined three core features: content focus, active learning, and coherence in the teachers' overall PD. The outcomes were all measured via teacher self-report. The results indicated that, except for the form of the activity, all of the other structural and core features of PD had positive effects on enhanced teacher CK and pedagogy.

PD courses integrating SCK and pedagogy have been found to increase teachers' confidence in teaching science (Shallcross et al., 2002). Cox and Carpenter (1989) created a continuing education course for practicing elementary school teachers to develop science teaching skills through hands-on inquiry lessons. The researchers specifically decreased the amount of science content in favor of teaching methods, nature of science, and process skills. Rather than end-of-unit or end-of-semester tests, the teachers were only given daily mini-quizzes, so there were no data for their resulting cumulative CK. However, surveys demonstrated that the teachers *felt* significantly more comfortable with the content and better prepared to teach it.

Effect of Teacher Content Knowledge on Student Achievement Outcomes

Teacher knowledge and perceptions have been shown to have a direct effect on student achievement outcomes. However, the lack of direct measures of teacher SCK, such as tests, surveys, or observations, forces the use of other measures that are supposedly correlated with SCK, such as the science content of PD and the college education of the teachers.

PD that includes science content has been shown to have a significant positive effect on student science test scores regardless of the form of PD, partly due to the PD's effect on teacher

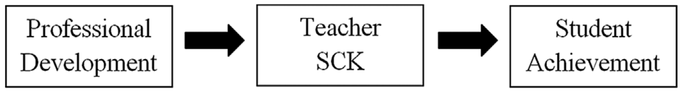

Figure 1. Model for impact of PD on teacher SCK and student achievement.

SCK (Heller et al., 2012). Heller et al. (2012) offered a PD opportunity in which teachers learned about circuits through hands-on inquiry-based investigations. After the PD, teachers and their students were both given tests of SCK. PD resulted in a test score gain of 22% for treatment teachers, compared to 2.4% for control teachers, with an effect size of about 1.8 *SD*s. In addition, treatment students showed an improvement of about 7% over the control students. This suggests that PD that improves teacher CK also improves students' content knowledge.

The Longitudinal Study of American Youth (LSAY) found that 90% of variation in student achievement in reading and mathematics was attributable to differences in teacher qualifications (Ball et al., 2001; Monk, 1994). Monk (1994) analyzed the LSAY for the number of mathematics and science courses taken in college by high school teachers and found that up to five mathematics courses showed positive effects of a 1.2% gain in student test score per course, but then the positive effects leveled off to a 0.2% increase per course beginning with the sixth course. The largest effects on student achievement were found in courses on mathematics pedagogy taken by teachers, rather than those on mathematics content.

Monk (1994) also used the LSAY to analyze high school teacher SCK, separating classes taken into biology, chemistry, physics, earth science, elementary science education, and secondary science education. For sophomores, teacher preparation in life sciences had no effect on student performance. For juniors, life sciences preparation showed a significant negative effect. For both sophomores and juniors, teacher preparation in physical sciences showed a significant positive relationship with student performance. Monk postulated that the negative effects on life sciences were probably a function of selection bias as far as which teachers chose to take physical sciences and the fact that juniors were usually taking physical sciences, so a teacher with only life sciences experience might not be prepared to teach the students.

The university and specific program in which a teacher was trained had a significant effect on student math achievement (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009). Program features that are most positively linked to student achievement include how closely the preparation follows what teachers need to do in their first year of teaching and how much time is required for field experience. The results of these studies of teacher CK suggest that improving teacher knowledge and perceptions can also improve student achievement.

Predictors of Teacher SCK and Student Science Achievement

Our previous research provided data on teacher variables (Lee & Maerten-Rivera, 2012) and student variables (Maerten-Rivera, Myers, Lee, & Penfield, 2010) that may affect their respective SCK. Theoretical considerations of these variables, discussed here, guided the development of statistical models of predictors for this study.

Predictors of Teacher SCK. Lee and Maerten-Rivera (2012) found that the number of science courses taken in college was a significant predictor of the change in teachers' self-reported science knowledge after participating in an intervention, accounting for 13% of the variation. This result was consistent with other literature that has shown that science courses taken in college lead to higher self-reported SCK (Diamond et al., 2013; Nowicki et al., 2013; Shallcross et al., 2002).

The number of years of teaching experience has been shown to affect student achievement. Two meta-analyses (Rice, 2003; Wayne & Youngs, 2003) found that number of years teaching generally had a positive effect on student achievement. Hanushek and Rivkin (2010) found that more effective teachers are more likely to stay in teaching than less effective teachers, which may partially explain the positive effect.

The highest degree a teacher has earned has shown mixed results as a predictor of student achievement in previous studies (Wayne & Youngs, 2003). A meta-analysis showed that highest degree was both a significant positive and a significant negative predictor of student achievement, depending on the study. Rice (2003) found that advanced degrees have a positive effect on student science and mathematics achievement when the degree is in the respective field.

Predictors of Student Science Achievement. There are several student background factors that have been found to influence science achievement. Boys have repeatedly been shown to have higher science achievement than girls (Bacharach, Baumeister, & Furr, 2003; Maerten-Rivera et al., 2010; Mau & Lynn, 2000; O'Reilly & McNamara, 2007; Reis & Park, 2001). Racial and ethnic minorities have consistently been shown to have lower science achievement (Bacharach et al., 2003; Chapin, 2006; Maerten-Rivera et al., 2010) and lower academic achievement in general (Horn, 2003; Lee, 2002; Madaus & Clarke, 2001). Childhood poverty, which can be accounted for by the receipt of free or reduced price lunch, also has a negative impact on student achievement (Duncan, Yeung, Brooks-Gunn, & Smith, 1998; Guo, 1998; Lee, 2007; Maerten-Rivera et al., 2010). When standardized tests are used as the measure of student achievement, English language learners (Abedi, 2002; Abedi & Lord, 2001; Abedi, Lord, Hofstetter, & Baker, 2000; Horn, 2003; Maerten-Rivera et al., 2010) and special education students (Horn, 2003; Koretz & Hamilton, 1999, 2001; Maerten-Rivera et al., 2010) frequently underperform compared to the general population.

Student reading and mathematics achievement have both been shown to impact student science achievement (Maerten-Rivera et al., 2010). Students with higher reading ability have higher comprehension and make more inferences than students with lower reading ability (Yuill & Oakhill, 1991). Reading level has also been directly connected to science achievement scores with elementary school students (Maerten-Rivera et al., 2010) and middle and high school students (Gustin & Corazza, 1994; O'Reilly & McNamara, 2007). Mathematics level has also been connected to science achievement, but the effect is smaller than that of reading level (Gustin & Corazza, 1994; Maerten-Rivera et al., 2010).

Measures of Teacher Content Knowledge

Teacher CK has typically been assessed using self-reports and classroom observations. However, tests of the overall CK of practicing teachers are rare. Each of the measures presents strengths and limitations.

While few studies have administered tests to teachers as a measure of SCK, U.S. teachers have been given written tests as part of the teacher certification process for about a century (D'Agostino & Powers, 2009). These certification tests vary by state and, over time, have varied in terms of which states require them. In lieu of tests, GPA has been used by researchers as a measure of teacher performance while in college. A meta-analysis of the predictive value of teachers' certification test scores and college GPA as predictors of teacher performance showed that while the median correlation between certification test scores and college GPA was 0.55, the test scores showed a correlation with supervisor and principal ratings of only 0.05 for preservice teachers and 0.11 for in-service teachers. There was also a low correlation between college GPA and these supervisor ratings (0.07), although the correlation was higher in the first year of teaching (0.25). According to this analysis, it would seem that GPA is a better indicator of teacher performance than certification tests, but *only in the first year of teaching*. These results demonstrate that certification tests and GPA are both poor predictors of teacher performance in general, suggesting the need for measures with more predictive value.

Surveys are viewed by researchers as the easiest to implement method of data collection, especially in large data sets (Desimone, 2009; Supovitz & Turner, 2000). However, they are most likely to contain bias favoring socially desirable responses. A validity test of the Teachers' Sense of Self-Efficacy Scale (TSES) spanning five countries (Canada, Cyprus, Korea, Singapore, and United States) that examined measurements of invariance and explored the relationship between the scale and job satisfaction found the model was shown to be good both within and between countries (Klassen et al., 2009). Another study found a good fit for a structural equation model of teacher self-reports of their own learning behaviors (Gordon, Dembo, & Hocevar, 2007). Teacher self-reported enthusiasm for teaching mathematics is strongly correlated with both self-reported and student-reported quality of instructional behavior, but teacher self-reported enthusiasm for mathematics as a subject is only correlated with self-reported quality of instructional behavior, not student-reported (Kunter et al., 2008). Teacher self-report scales often show evidence of validity, but they are still not considered by the research community to provide information as good as that provided by classroom observations.

Most researchers view observation as the most unbiased, if most difficult, form of data collection (Desimone, 2009). It removes self-report bias (although observer bias comes into play) and allows researchers to directly witness what happens during PD activities and in classrooms. Observations allow researchers to make finer distinctions in teacher practice than are afforded with surveys. For observations to be considered valid and reliable measures of teachers' overall instruction, "three observations are required for one stable observation, and at least three stable observations over an extended period of time are required" (Desimone, 2009, p. 189). The need for so many observations is proposed in response to the trend for observations to be performed once or twice per teacher for as short as 20 minutes each. The drawback to observation is that it is "the most time-consuming and expensive method of measuring [PD] and teaching" (Desimone, 2009, p. 188).

Many early studies from the 1960s and 1970s showed low correlations between classroom observations and teacher self-reports, but most of these studies did not use methods that stand up to scrutiny today (Desimone, 2009). The observations tended to be short, infrequent, and overgeneralized. Later studies that used more rigorous research methods showed that classroom observations and teacher self-reports tended to have moderate to high correlations (Desimone, 2009; Supovitz, Mayer, & Kahle, 2000).

The number of courses taken in college is a common proxy for teacher CK in both science and mathematics. When researchers have counted the number of mathematics courses taken as a measure of teacher mathematics CK, the effects have typically been small (Ball et al., 2001; Boyd et al., 2009). According to Ball et al. (2001) and Begle (1979) found that taking mathematics courses led to positive effects on student achievement in 10% of cases, negative effects in 8%, and no effect in 82%, while majoring or minoring in mathematics yielded positive main effects in 9% of cases and negative main effects in 4%. On the other hand, the same study found that the number of math methods courses taken accounted for 23% of positive student performance and 5% of negative. Later studies have also shown that more mathematics content courses do not lead to improved student achievement (Boyd et al., 2009).

In summary, teacher SCK is an under-studied construct, in spite of the importance that has been attached to it in regards to quality of science instruction. PD has been shown to improve teachers' pedagogy and confidence, but its effect on SCK has rarely been addressed. There have also been few studies that examine the effect of teacher SCK on student achievement outcomes, although student achievement is generally seen as a primary goal of teaching. When teacher CK is studied, indirect methods, such as surveys, classroom observations, and college science courses taken, are used much more often than more direct measures, such as test scores. This study uses test scores, surveys, college science courses taken, and classroom observations of a subsample to measure teacher SCK, the effect of an intervention including PD on teacher SCK, and the effect of teacher SCK on student achievement.

Teacher Professional Development Intervention

While the intervention continues for three years, this study involves the first year only. The teacher PD intervention is comprised of (a) curriculum materials such as a student book, teachers' guide, science supplies, and online Supplementary materials; (b) teacher workshops throughout the school year; and (c) school site support for curriculum implementation. The intervention components were designed to complement and reinforce one another for the improvement of teacher knowledge and practices in science instruction based on state science content standards, along with English language development of ELLs, in urban elementary schools. Control schools and teachers did not receive the intervention.

Curriculum Materials

A comprehensive stand-alone science curriculum for grade 5 was developed from previous projects. Alignment between the curriculum and the benchmarks tested by state science assessment at grade 5 was ensured through consulting the state science content standards. The curriculum was developed to address all of the benchmarks in the state science standards in place for the 2010–2011 school year (see Appendix A). It is partly through the in-depth exploration of these concepts that teachers in the treatment group were expected to increase their SCK.

The curriculum used an Inquiry Framework to guide students through the inquiry process that was designed to encourage students to ask questions, to develop their own understandings, and to share their understandings with others. Suggestions to promote students' engagement in science inquiry and their understanding of science concepts were presented throughout the teachers' guide. Because inquiry-based learning is largely directed by students and the teacher does not completely control what aspects of a topic will be brought up in class, it is necessary for teachers to have stronger SCK when teaching science through inquiry (Basista & Mathews, 2002; Kanter & Konstantopoulos, 2010).

The curriculum was designed to promote science learning in combination with literacy development and English proficiency for students from diverse languages and cultures. Although the curriculum was designed to meet the needs of ELLs, it also worked well with native English speakers.

At the beginning of their participation in the intervention, the teachers in the treatment schools were provided with complete class sets of curriculum materials. The control schools implemented the district-adopted science curriculum.

Teacher Workshops

As a school-wide initiative, all fifth grade science teachers in the treatment schools were invited to attend 5 days of workshops throughout the school year. The first 3 days of the teacher workshops occurred during the summer of 2010 and focused on the curriculum that would be covered through December. In mid-January, teachers attended an additional full-day workshop that was similar in structure to the summer workshop but focused on the portion of the curriculum that would be covered from January until the state science assessments in mid-April. Teachers attended the year-end workshop in May 2011 to offer their feedback on the intervention, to plan for the next year, and to participate in data collection activities. Teachers received stipends for attending the summer workshop, and schools received payments for substitute teachers during the school year.

The purpose of the workshops was to familiarize the teachers with the science content and hands-on activities, as well as the state science content standards and high-stakes science assessment. Teacher workshops focused on CK and how students learn that content, along with opportunities for teachers to engage in active learning, two PD methods that have been found to have positive effects on teacher self-reported enhanced knowledge (Garet et al., 2001). The summer and winter workshops provided teachers with the opportunity to experience the inquiry-based labs of the curriculum as students would, a PD method that has been found to improve teachers' comfort (Cox & Carpenter, 1989) and confidence (Shallcross et al., 2002) with science content, as well as improve their SCK (Appleton, 2008) more than learning it didactically. By performing the labs as learners, teachers had opportunities to improve their own knowledge of science content (see Appendix A). They were also able to ask questions and clarify misconceptions with the workshop facilitators. Most labs were carried out during the workshops, and the few that could not be carried out due to time restraints or other logistical reasons (e.g., requiring time to carry out over a few weeks) were discussed as a group. While teachers in the control group were free to participate in district PD opportunities, they were not included in the workshops provided through this intervention.

School Site Support

Three members of the research team were each assigned to a group of treatment schools and on average visited each school every 4–6 weeks, for a total of four to six times. The school site support team paid particular attention to high-needs schools and teachers on the basis of classroom visits or in response to requests from teachers or school administrators. Assistance included planning lessons, co-teaching, suggestions for additional materials (e.g., home learning, lab management, supplementary assessment materials), delivery of supplies, and on-site PD. The on-site PD included further explanations of science content and guidance on implementing inquiry activities, which could be expected to contribute to teacher SCK. While conducting the visits with all of the teachers, the research team members also met with school administrators to address concerns. Control teachers did not receive school site support as an intervention.

Method

Research Setting

The research was conducted in a large urban school district in the Southeast United States with a linguistically and culturally diverse student and teacher population. The ethnic makeup of the student population in the district was 65.1% Hispanic, 25.2% Black Non-Hispanic, 7.9% White Non-Hispanic, and 1.8% Other. In addition, 77.8% of elementary students were eligible for free or reduced price lunch programs, 12.7% were designated as ELLs and were enrolled in English to Speakers of Other Languages (ESOL) programs, and 11% were enrolled in exceptional student education (ESE) programs.

Research Design

The study used a cluster randomized trial design. First, 64 schools were randomly selected from a pool of 206 available schools (not including 23 due to district monitoring and nine that participated in our previous project). Second, the 64 schools were randomly assigned to 32 experimental and 32 control schools. As a school-wide initiative, the study involved all fifth grade science teachers from the 64 schools. The student and teacher samples of the treatment and control schools were comparable.

Teacher Participants

Out of the 260 eligible teachers from the 64 schools, a total of 37 were not included in the study because they were not teaching fifth grade science for the majority of the school year

(20 cases), they chose not to participate in the study (13 cases), or data from SCK measures were completely missing (four cases). Therefore 227 teachers participated in the study, but 223 provided adequate data to be included in this analysis. The number of teachers who provided each data point is presented in Table 1. The demographic information for the 223 participating teachers who provided at least one SCK measure is shown in Table 2. The teachers came from diverse backgrounds in terms of ethnicity, language, and highest degree earned. The teacher sample in the study was representative of the teacher population in the school district. In addition, the diversity of the teacher sample was consistent with the diversity of the student sample.

Student Participants

Table 3 presents student demographics at the treatment and control schools during the first year of the intervention. The students came from diverse backgrounds in terms of ethnicity, language, socioeconomic status, ESOL status, and ESE status. The student sample in the study was representative of the student population in the school district.

Instruments, Data Collection, and Coding

At the beginning of the summer workshop, teachers from the treatment schools were provided with informed consent and then asked to complete the background information form, questionnaire, and science knowledge test. At the year-end workshop, the questionnaire and science knowledge test were administered again. Teachers from control schools were given the same instruments at the school sites in the first quarter of the school year, and the questionnaire and science knowledge test were administered again at the end of the school year. Control group teachers were given a small stipend for their participation. Classroom observations were conducted with one randomly selected teacher from each of the 64 schools three times from September to January. Detailed validity and reliability information for each measure can be found in Maerten-Rivera, Adamson, and Ahn (2012) and Maerten-Rivera, Huggins, and Adamson (2013).

Teacher Science Knowledge Test (TEST). The teacher science knowledge test consists of 24 multiple-choice and six short response items. Of the 30 items, 24 were taken from Trends in International Mathematics and Science Study (TIMSS) or National Assessment of Educational

Table 1
Total number of teachers to provide data for each measure

Measure	Treatment *n*	Control *n*	Total *N*
Total teachers in study	122	105	227
At least 1 SCK measure	120	103	223
All SCK measures	97	78	175
Science knowledge pretest (PRETEST)	106	90	196
Science knowledge posttest (POSTTEST)	114	98	212
Both science knowledge tests	100	85	185
Science knowledge scale (PRESKS)	107	92	199
Science knowledge scale (POSTSKS)	114	99	213
Both SKS	100	87	187
Observations	32	30	62
College science courses (COURSES)	119	99	218
Advanced degree (DEGREE)	119	103	222
Years teaching (YEARS)	117	103	220

Table 2
Teacher demographics (N = 223)

Variables	Demographic Groups	*n*	%
Gender	Male	41	18.4
	Female	182	81.6
Ethnicity	Hispanic	122	54.7
	Black Non-Hispanic	47	21.1
	White Non-Hispanic	42	18.8
	Haitian	5	2.2
	Asian	1	0.4
	Other	2	0.9
Native language(s)[a]	English	180	80.7
	Spanish	101	45.3
	Haitian Creole	8	3.6
	French	4	1.8
	Other	4	1.8
Other fluent language(s)[a]	English	42	18.8
	Spanish	26	11.7
	Haitian Creole	3	1.3
	French	1	0.4
	Other	5	2.2
ESOL training[a]	Bachelor's or master's degree in ESOL	19	8.5
	ESOL endorsement through college coursework	78	35.0
	ESOL endorsement through school district	114	51.1
	Grandfathered in through teaching	12	5.4
	No preparation for ESOL	18	8.1
Degrees	Bachelor's	108	48.4
	Master's	90	40.9
	Multiple Master's Degrees	7	3.1
	Specialist	11	4.9
	Doctorate	2	0.9
	Other	2	0.9

[a]Multiple categories could be selected.

Table 3
Descriptive statistics of student-level variables (N = 5,784)

Variable	Abbreviation	Values	*P*/*M*	*SD*
Female	FEMALE	0 = Male, 1 = Female	0.49	0.50
Ethnicity Hispanic	HISP	0 = Non-Hispanic, 1 = Hispanic	0.67	0.47
Ethnicity Black	BLACK	0 = Non-Black, 1 = Black	0.23	0.42
Free/reduced price lunch status	LUNCH	0 = Not on FRL, 1 = FRL	0.77	0.42
English as second or other language levels 1–4	ESOL	0 = Non-ESOL 1-4, 1 = ESOL 1–4	0.12	0.33
English as second or other language level 5 (exited)	ESOL5	0 = Non-ESOL5, 1 = ESOL5	0.41	0.49
Exceptional student education	ESE	0 = Non- ESE, 1 = ESE	0.11	0.32
Reading score	READ	100–500	304.88	58.25
Science score	SCI	100–500	317.84	61.44

Programs (NAEP) and six items were project-developed. Eleven of the NAEP and TIMSS items were administered to fourth grade and were rated as difficult content and high cognitive complexity for this grade level, while 12 items were administered to eighth grade. One NAEP item was for an unknown grade level. The items were selected to reflect state science content standards in life sciences, physical sciences, earth sciences, and nature of science at intermediate grades 3–5.

Of the 227 teachers in the study sample, 42 teachers missing either PRETEST or POSTTEST were removed from this analysis, resulting in 185 teachers. When checking a scale for internal consistency, social science research considers an alpha of 0.70 as the lowest acceptable and 0.80 to be indicative of high internal consistency reliability (Corn, 2010). PRETEST was shown to have a Cronbach α of 0.77 for the treatment group and 0.82 for the control group, while POSTTEST had an alpha of 0.64 for treatment and 0.76 for control, indicating good reliability overall. The alphas for the POSTTEST were expected to be lower than for the PRETEST, because the test is less able to discriminate as more teachers approach the maximum score.

TEST was scored out of 38 possible points. For each multiple-choice question, one point was awarded for each correct answer. For short response questions, teachers were awarded the number of points, 0–3, recommended by the original source of each question, with partial credit usually available. Treatment and control teachers both scored a mean of approximately 30.50 points ($SD = 5.09$ and $SD = 5.61$, respectively) on PRETEST (see Table 4). Treatment teachers scored a mean of 32.71 points ($SD = 3.46$) on POSTTEST, for an increase of 2.12 points, whereas control teachers scored a mean of 31.13 points ($SD = 4.83$), for an increase of 0.64 points.

Teacher Self-Reported Science Knowledge (SKS). The science knowledge scale (SKS) is a section of the questionnaire that consisted of four Likert-type questions asking the teachers to indicate how knowledgeable they felt about teaching the nature of science, physical science, earth/space science, and life science at their grade level (see Appendix B). The questionnaire items used a 4-point rating system that ranged from 1 (*not knowledgeable*) to 4 (*very knowledgeable*).

Teachers missing either PRESKS or POSTSKS were removed from this analysis, resulting in the removal of 40 cases. Cronbach α for PRESKS was 0.95 for the treatment group and 0.91 for the control group, indicating strong reliability. Cronbach α for POSTSKS was 0.88 for the treatment group and 0.94 for the control group, indicating strong reliability.

SKS for each teacher was calculated as the average score (1–4) the teachers gave themselves for how knowledgeable they felt about the four science topics surveyed. A score of a 3 on SKS means that teachers felt generally knowledgeable about fifth grade science (see Appendix B). Treatment teachers scored a mean of 2.74 ($SD = 0.71$) on PRESKS and control teachers scored a mean of 2.92 ($SD = 0.59$). Treatment teachers scored a mean of 3.24 ($SD = 0.49$) on POSTSKS, for an increase of about 0.50. Control teachers scored a mean of 3.16 ($SD = 0.61$) on POSTSKS, for an increase of about 0.24.

Classroom Observations (OBS). Classroom observations were conducted using the project-developed classroom observation guideline, which provides detailed descriptions of the scales and criteria for 1–5 ratings on teachers' observed teaching practices during particular lessons. Specifically, the observation scale on SCK (Teacher Knowledge of Science Content) measures the extent to which the teacher has an accurate and comprehensive grasp of the science content of the lesson (see Appendix C). Only the SCK portion of the observation instrument was used for this study.

Three members of the research team conducted observations. The observers were trained together during a 1 week, 4 hours per day, training program led by experienced former observers.

Table 4
Descriptive statistics of teacher-level variables included in ANOVA

Measure	Treatment				Control			
	n	*M*	*SD*	α	*n*	*M*	*SD*	α
PRETEST	100	30.59	5.09	0.77	85	30.49	5.61	0.82
POSTTEST	100	32.71	3.46	0.64	85	31.13	4.83	0.76
PRESKS	100	2.74	0.71	0.95	87	2.92	0.59	0.91
POSTSKS	100	3.24	0.49	0.88	87	3.16	0.61	0.94
OBS	32	3.13	0.71	0.53	30	3.16	1.04	0.89
COURSES	118	4.77	4.55	—	96	5.11	4.99	—
DEGREE	119	0.52	0.50	—	103	0.49	0.50	—
YEARS	117	12.69	7.71	—	103	13.66	9.55	—

Note: For each measure involving a pre and post, all teachers who filled out both were included in the analysis. For each measure involving 1 data point, all teachers who provided that information were included in the analysis.

The training objective was to ensure consistent scoring by learning how the OBS scales were meant to be interpreted, practicing the coding of observation notes from our previous research, and practicing observations using videotaped lessons.

One randomly selected teacher from each of the 62 schools was observed three times between September and January. Therefore, the observed teachers were a random subset of the total sample. Each round of observations was conducted for both treatment and control teachers during the same part of the year, so the district pacing guide allowed us to see comparable lessons across both groups. After the observation was completed, there was no discussion between the observer and teacher, as the observations were conducted to collect data on classroom instruction, not to offer PD.

Throughout the data collection period, 10% of the observations were conducted with pairs of observers and scores were compared to ensure observer agreements. Scores that were within one point were considered to be in agreement. Based on this criterion, there was 95% agreement of scores between observers. Any discrepancies were discussed to make the scoring more consistent in subsequent observations.

Four teachers were not available for all three observations, leading to a total of 182 observations. The observation score for each teacher was calculated as the mean of the three observation scores (1–5). When there were fewer than three observations done, the mean of the performed observations was used. Treatment teachers scored a mean OBS score of 3.13 ($SD = 0.71$) and control teachers scored a mean OBS score of 3.16 ($SD = 1.04$). An OBS score of 3 indicates that teacher knowledge of science content during observed lessons was generally accurate within the bounds of the lesson content, without adding significant correct or incorrect information to the provided content (see Appendix C).

Cronbach α for the OBS scores of the treatment group was 0.53, indicating poor reliability, whereas Cronbach α for the OBS scores of the control group was 0.89, indicating strong reliability. These internal consistency estimates suggest that the observation instrument was a more reliable measure for the control group than for the treatment group. OBS is different from the other measures of SCK used in this study because it included a subsample of teachers and did not have PRE and POST scores as the observations were conducted throughout the school year.

College Science Courses Taken (COURSES). One section of the teacher background information form asked teachers to indicate how many college (both undergraduate and graduate) science courses they had taken in six categories, including other science courses and science methods courses, on a scale from 0 to 6 or more (see Appendix D). COURSES was calculated as the total of these sciences courses taken. The mean COURSES for the treatment and control groups were then calculated. Teachers from both the treatment and control groups took a mean of approximately five science courses and science methods courses in college.

Data Analysis

Data were missing from some teacher records. In order to maximize the power and accuracy of this study, we chose to use all participants with data available in each analysis, resulting in a different N for each analysis (see Table 1).

Variables. As the student achievement outcome measure, the study used the high-stakes science test that covered the topics of physical and chemical science, earth and space science, life and environmental science, and scientific thinking. In the 2010–2011 school year, the science test for fifth grade included 51 multiple-choice questions. Half of the questions measured knowledge, comprehension, or application, while the other half measured analysis, synthesis, or evaluation. The overall test had a Cronbach α of 0.89, indicating high reliability.

The descriptive statistics and variable name abbreviations for the student-level predictor variables are presented in Table 3, and the intercorrelations of the student-level variables are presented in Table 5a. For the analyses, dummy variables were created as shown in Table 3. More detailed descriptions of the descriptive statistics and the ESOL and ESE variables can be found in the methods supplement.

The high-stakes standardized reading test scores were used as a predictor. High-stakes standardized mathematics test scores were not used as a predictor, because they were highly correlated with the reading test scores. We chose reading scores over math scores because they are a more widely used predictor of overall school achievement. The reading test covered the topics of words and phrases in context; main idea, plot, and purpose; comparisons and cause–effect; and reference and research. Each year, the reading test includes 50–55 multiple-choice questions.

The intercorrelations for the teacher-level variables are displayed in Table 5b. While there are significant correlations between some of the teacher measures, none of them are large enough to prevent them from being included in the same multilevel model, with the largest correlation being between POSTTEST and OBS ($r = 0.42$).

Effect of the Intervention on Teacher SCK (Research Question 1). Two separate one-way between groups analysis of variance (ANOVA) models were examined to determine if change in TEST and SKS, as measured by the difference between PRE and POST, differed by treatment. A third one-way between groups ANOVA model was examined to determine if OBS differed by treatment. Change in OBS was not measured because OBS was measured throughout the year, rather than at PRE and POST. A previous version of this intervention found that COURSES was a significant predictor of increase in SKS, and no significant predictors were found for increases in OBS (Lee & Maerten-Rivera, 2012). Supovitz et al. (2000) found that teachers with more teaching experience tended to feel less prepared to use inquiry as a method of teaching science, and inquiry was the primary vehicle for both teacher teaching and learning in our intervention. Therefore, the variables of highest degree earned, years teaching, and number of college science courses taken were all included in the model as covariates. The descriptive statistics of the teacher variables used

Table 5a
Intercorrelations among student-level variables

	HISP	BLACK	LUNCH	ESOL	ESOL5	ESE	READ	SCI
FEMALE	−0.02	0.02	0.01	−0.03**	0	−0.10**	0.07**	−0.02*
HISP	—	−0.79**	0.08**	0.14**	0.41**	0.02	0.01	0.01
BLACK	—	—	0.12**	−0.11**	−0.36**	−0.01	−0.12**	−0.13**
LUNCH	—	—	—	0.13**	0.13**	0.05**	−0.26**	−0.26**
ESOL	—	—	—	—	−0.31**	0.01	−0.39**	−0.34**
ESOL5	—	—	—	—	—	0.02*	0.11**	0.11**
ESE	—	—	—	—	—	—	−0.28**	−0.24**
READ	—	—	—	—	—	—	—	0.81**

Table 5b
Intercorrelations among teacher-level variables

	YEARS	COURSES	POSTSKS	POSTTEST	OBS
DEGREE	0.25**	0.1	−0.03	−0.07	0.12
YEARS	—	0.23**	0.07	0.05	0.01
COURSES	—	—	0.18*	−0.02	−0.03
POSTSKS	—	—	—	0.27**	−0.01
POSTTEST	—	—	—	—	0.42**

*$p < 0.05$.
**$p < 0.01$.

in the ANOVA are shown in Table 4, and the intercorrelations between those variables are shown in Table 5b.

The Q–Q plot was examined and a Levene's test of homogeneity was conducted for TEST, SKS, and OBS. For each outcome the null hypothesis of equal variances was accepted, and thus the assumptions of the ANOVA were not violated. For the purpose of this research p values less than 0.05 were considered statistically significant. The partial eta squared measure of effect size was reported for each variable, which is a measure of the proportion of variance accounted for.

Effect of SCK on Student Science Achievement (Research Question 2). Cross-sectional multilevel modeling (MLM) using HLM 6 software (Raudenbush, Bryk, & Congdon, 2005) was conducted with students nested within teachers to determine the effect of POSTTEST, POSTSKS, OBS, and COURSES on student achievement outcomes as measured by the state high-stakes science test after controlling for student and teacher-level variables. The post-intervention measures of TEST and SKS were used as predictors in the model assuming that teachers who learned additional science content did so before teaching it, which makes the post-intervention scores more relevant to student learning. Multilevel modeling allowed us to model two levels (i.e., student level and teacher level) simultaneously; the intercepts and slopes of the student-level model became outcome variables at the teacher level. We considered nesting the teachers within schools as a third level, but some teachers were the only fifth grade science teachers in their schools, making the three-level models unstable.

An unconditional multilevel model in which no student-level or teacher-level predictors were included, except teacher itself, was imposed first. This provided the intraclass correlation coefficient (ICC), which estimated the proportion of the variance in science achievement that was attributed to teacher-level differences. A model-building approach was generally employed as suggested by Raudenbush and Bryk (2002). First, the student background predictors (e.g., GEN, BLK, HSP, LUNCH, ESOL, and ESE) were entered into the model; then, the reading achievement predictor was added. We selected a sequential approach in an attempt to build the complex model in a careful and deliberate way. The model was assessed in each step, and the effect of each group of predictors was examined. Only the final model is presented here for parsimony, but details of the model building can be found in the online Supporting Information.

Once the final model for student background predictors was established, the proportion of within-teacher variance accounted for (PVAF) in science achievement by the set of background predictors was computed to determine the amount of within-teacher variance they combined to explain. In addition, the proportion of within-teacher variance uniquely accounted for (PVUAF) by each of the background predictors was computed by comparing a baseline model without a predictor of interest to the fitted model with the predictor of interest and examining the change in σ^2. The PVUAF provides a measure of effect size for each of the predictors. The proportion of variance estimates is often referred to as a pseudo-R^2 (Raudenbush & Bryk, 2002; Singer & Willett, 2003). The R^2 can be interpreted as values of less than 0.09 as having a small effect, between 0.09 and 0.25 as having a medium effect, and greater than 0.25 as having a large effect (Cohen, 1988); therefore, the proportion of variance estimates can be interpreted in the same way as R^2.

The teacher-level model was built by first entering the professional predictors DEGREE and YEARS.[1] Our next step was to add the measures of teacher SCK one at a time. We chose not to include all of the SCK variables at the same time because they confounded each other when they were included in the model together. We ran separate series of models from the beginning for OBS, because only a subsample of teachers was observed.

The fixed effect for each predictor was examined, and the proportion of between-teacher variance in science achievement accounted for (PVAF) was computed for the set of Level 2 predictors. In addition, the proportion of between-teacher variance uniquely accounted for (PVUAF) by each of the teacher predictors was computed by comparing a baseline model without the predictor of interest to the fitted model with the predictor of interest and examining the change in τ_{00}.

In the analysis, the Level 1 or student-level model expressed each student's science achievement score as a function of the teacher mean science achievement score (β_{0j}), a series of predictor variables related to the student (Table 3), and error (r_{ij}). The Level 2 or teacher-level model specified the teacher mean science achievement score (β_{0j}) as a function of the teacher-level mean science achievement score (γ_{00}), a series of predictor variables related to the teacher, and error (u_{0j}). No teacher-level predictors were entered into the slope parameters (e.g., $\beta_{1j} - \beta_{8j}$). The equations below provide uncombined equations representing the a priori models. Separate models were examined where COURSES, SKS, POSTTEST, and OBS were each entered into the specified model as the SCK measure so that each could be tested separately.

Level 1 model:

$$\begin{aligned} Science_{ij} = {} & \beta_{0j} + \beta_{1j}(\mathrm{FEMALE}) + \beta_{2j}(\mathrm{HISP}) + \beta_{3j}(\mathrm{BLACK}) + \beta_{4j}(\mathrm{LUNCH}) + \beta_{5j}(\mathrm{ESOL}) \\ & + \beta_{6j}(\mathrm{ESOL5}) + \beta_{7j}(\mathrm{ESE}) + \beta_{8j}(\overline{\mathrm{READING}}) + r_{ij} \end{aligned}$$

Level 2 models:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{DEGREE}) + \gamma_{02}(\text{YEARS}) + \gamma_{03}(\text{SCK Measure}) + u_{0j}$$
$$\beta_{1j} = \lambda_{10} + u_{1j}$$
$$\beta_{8j} = \gamma_{80} + u_{8j}$$

Results

Descriptive statistics for each of the four SCK measures are reported in Table 4. The TEST and SKS information is provided for PRE and POST. COURSES are given for just one time. Observations were conducted over the course of the school year and therefore are presented as means rather than as PRE and POST.

Effect of the Intervention on Teacher SCK

Tables 6a and 6b present the results from the one-way between groups ANOVA analyses. Teachers participating in the intervention demonstrated a statistically significant larger score increase on the TEST ($F_{(1,\ 166)} = 67.18$, $p = 0.029$, $\eta_p^2 = 0.03$) and on the SKS ($F_{(1,\ 169)} = 0.33$, $p = 0.004$, $\eta_p^2 = 0.05$). Both were small, positive effects, accounting for 3% and 5% of the variance, respectively. Table 6c shows that there was no significant difference between the treatment and control groups in the mean OBS scores ($F_{(1,\ 53)} = 0.79$, $p = 0.878$). Years teaching, highest degree earned, and number of college science courses taken were not statistically significant predictors in any of the ANOVA models.

Effect of Teacher SCK on Student Science Achievement

Unconditional Model. The Level 1 variance (σ^2) for the unconditional model was 2,585.25, and the Level 2 variance (τ_{00}) was 1,480.34. The ICC was computed as 0.36, suggesting that 36% of the variance in science achievement was due to between-teacher differences (i.e., teacher-level variation), whereas 64% of the variance in science achievement was due to within-teacher differences (i.e., student-level variation). The variance of the intercept (μ_{0j}) was statistically significant ($\chi^2[177] = 2879.08, p < 0.001$), suggesting a significant amount of variance in science achievement between teachers.

Student-Level Results. The model for student-level results including teacher professional predictors at Level 2 is displayed in Table 7. Details of the student-level results at each step of the analysis and an explanation of the final student-level model can be found in the Supporting Information. Tables with additional information about the final models and about intermediate models in the model-building process are also included in the Supporting Information.

Teacher-Level Professional Predictors. Teacher-level professional predictors were entered into the model as predictors of the intercept (Table 7). The Level 1 variance (σ^2) for this model was 1,053.46, and the Level 2 variance (τ_{00}) was 1,402.92. The variance of the intercept (μ_{0j}) was statistically significant ($\chi^2[82] = 1{,}864.68$, $p < 0.001$), suggesting a significant amount of variance in science achievement between teachers. The coefficient for YEARS was 0.76 ($t[174] = 2.29$, $p = 0.023$), indicating that we would expect an increase of 0.76 points in the

Table 6a
Analysis of variance in changes in TEST between treatment and control groups

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
DEGREE	4.08	1	4.08	0.3	0.588	0.002
YEARS	7.87	1	7.87	0.57	0.452	0.003
COURSES	1.64	1	1.64	0.12	0.731	0.001
GROUP	67.18	1	67.18	4.86	0.029	0.028
Error	2295.96	166	13.83			
Total	2,741	171				

Table 6b
Analysis of variance in changes in SKS between treatment and control groups

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
DEGREE	0.03	1	0.03	0.10	0.749	0.001
YEARS	0.03	1	0.03	0.10	0.750	0.001
COURSES	0.96	1	0.96	2.93	0.089	0.017
GROUP	2.74	1	2.74	8.36	0.004	0.047
Error	55.47	169	0.33			
Total	85.12	174				

Table 6c
Analysis of variance in OBS between treatment and control groups

Source	*SS*	*df*	*MS*	*F*	*p*	η_p^2
DEGREE	0.42	1	0.42	0.53	0.471	0.01
YEARS	<0.001	1	<0.001	<0.001	0.996	<0.001
COURSES	0.02	1	0.02	0.03	0.859	0.001
GROUP	0.02	1	0.02	0.02	0.878	<0.001
Error	42.00	53	0.79			
Total	614.58	58				

expected mean classroom science achievement given one additional year of teaching experience. YEARS uniquely accounted for 3% of the teacher-level variation in mean science achievement. The coefficient for DEGREE was insignificant, and DEGREE did not account for any unique variance. The professional variables combined to account for a total of 3% of the variation in classroom science achievement.

COURSES. Each measure of SCK was added into the model one at a time (see Table 8). The coefficient for COURSES (see Model 1) was 0.05 ($t[174] = 0.07$, $p = 0.941$), indicating that the number of science courses taken in college by a teacher was not a statistically significant predictor of mean classroom science achievement. COURSES did not account for any unique variance.

POSTSKS. The coefficient for POSTSKS (see Model 2) was 7.63 ($t[174] = 1.53$, $p = 0.127$), indicating that the science knowledge self-reported by a teacher was not a statistically significant

Table 7
Results of teacher professional variable model for full sample of teachers

Fixed Effect	Estimates of Fixed Effects					
	Coefficient	*SE*	*t*	*df*	*p*	PVUAF
Student-level variables						
Intercept (γ_{00})	327.64	5.68	57.64	175	<0.001	—
READ (γ_{01})	0.74	0.01	56.77	177	<0.001	0.51
FEMALE (γ_{02})	−10.47	0.88	−11.84	5773	<0.001	0.02
HISP (γ_{03})	−5.85	1.65	−3.55	5773	0.001	0.00
BLACK (γ_{04})	−11.12	2.00	−5.56	5773	<0.001	0.00
LUNCH (γ_{05})	−5.15	1.21	−4.24	5773	<0.001	0.00
ESOL (γ_{06})	−11.79	2.12	−5.56	177	<0.001	0.01
ESOL5 (γ_{07})	−0.12	1.14	−0.11	5773	0.915	0.00
ESE (γ_{08})	−6.06	2.36	−2.56	177	0.012	0.02
Total						0.59
Teacher-level variables						
DEGREE (γ_{01})	3.47	5.69	0.61	175	0.542	0.00
YEARS (γ_{02})	0.76	0.33	2.29	175	0.023	0.03
Total						0.03

Random Effect	Estimates of Variance Components				
	SD	Variance	*df*	χ^2	*p*
Intercept (u_{0j})	37.46	1402.92	79	1864.68	<0.001
READ (u_{1j})	0.10	0.01	82	126.70	0.001
ESOL (u_{6j})	10.31	106.33	82	94.89	0.156
ESE (u_{8j})	16.85	283.91	82	169.12	<0.001
Level 1 effect (r_{ij})	32.46	1053.46	—	—	—

predictor of mean classroom science achievement. The unique variance accounted for by POSTSKS was 1%.

POSTTEST. The coefficient for POSTTEST (see Model 3) was 2.16 ($t[174] = 3.36$, $p = 0.001$), indicating that the teacher science knowledge test score was a statistically significant predictor of mean classroom science achievement, and that for each additional point earned by the teacher on POSTTEST, mean classroom science achievement was expected to increase by 2.16 points. POSTTEST uniquely accounted for 6% of the total teacher-level variation in mean classroom science achievement.

OBS. Model 4 was built using the same process as above, beginning with the student level model, because it used a subsample of the teachers and students included in the models previously described. The Model 4 coefficient for OBS was 5.33 ($t[50] = 1.06$, $p = 0.296$), indicating that the OBS score of a teacher was not a statistically significant predictor of mean classroom science achievement. The unique variance accounted for by OBS was 1%.

Discussion and Implications

This study examined the effect of a curricular and PD intervention on teacher SCK. It further examined the influence of teacher SCK on student science achievement as measured by one state's

Table 8
Comparison of fixed effects and standard error estimates for four models

	Model 1	Model 2	Model 3	Model 4
	COURSES	POSTSKS	POSTTEST	OBS
	MLM	MLM	MLM	MLM
Student-level variables				
γ_{10} (Effect of READ on β_{0i})	0.74 (0.01)**	0.74 (0.01)**	0.74 (0.01)**	0.72 (0.02)**
γ_{20} (Effect of FEMALE on β_{0i})	−10.47 (0.88)**	−10.47 (0.88)**	−10.48 (0.88)**	−8.53 (1.44)**
γ_{30} (Effect of HISP on β_{0i})	−5.84 (1.65)**	−5.86 (1.65)**	−5.90 (1.65)**	−6.44 (2.75)*
γ_{40} (Effect of BLACK on β_{0i})	−11.11 (2.00)**	−11.13 (2.00)**	−11.05 (2.00)**	−12.93 (3.27)**
γ_{50} (Effect of LUNCH on β_{0i})	−5.15 (1.21)**	−5.15 (1.21)**	−5.15 (1.21)**	−6.63 (2.10)**
γ_{60} (Effect of ESOL on β_{0i})	−11.76 (2.12)**	−11.93 (2.11)**	−11.57 (2.13)**	−9.13 (2.69)**
γ_{70} (Effect of ESOL5 on β_{0i})	−0.12 (1.14)	−0.13 (1.14)	−0.10 (1.14)	0.22 (1.95)
γ_{80} (Effect of ESE on β_{0i})	−6.03 (2.36)*	−5.85 (2.36)*	−6.15 (2.36)*	−4.65 (3.91)
Teacher-level variables (β_{0i})				
γ_{00} (Mean β_{0i})	327.48 (6.05)**	303.85 (16.55)**	257.72 (21.58)**	305.58 (17.45)**
γ_{01} (Effect of DEGREE on β_{0i})	3.46 (5.71)	4.00 (5.68)	5.40 (5.57)	−6.67 (2.75)
γ_{02} (Effect of YEARS on β_{0i})	0.76 (0.34)*	0.71 (0.33)*	0.73 (0.32)*	0.80 (0.48)
γ_{03} (Effect of SCK on β_{0i})	0.05 (0.61)	7.63 (4.98)	2.16 (0.64)**	5.33 (5.04)
PVAF of SCK variable	0.00	0.01	0.06	0.01

Note: Model 4 is based on a subsample of the total sample used in Models 1, 2, and 3. Therefore, while the first three models are mostly the same, Model 4 is a completely separate model that was built from scratch.
$^{*}p < 0.05$.
$^{**}p < 0.01$ Standard errors are in ( ).

high-stakes test. The study was conducted with a culturally and linguistically diverse student sample in elementary schools from a large urban school district.

Discussion

Effect of the Intervention on Teacher SCK. The results indicate that the intervention had a significant effect on teacher SCK as measured by test scores and self-reported science knowledge. The effect sizes for both measures suggest that the intervention had a small, but meaningful, impact on teacher SCK. Since the intervention combined the teaching of SCK and pedagogy through hands-on inquiry-based methods, this result is consistent with previous research indicating that this format of PD is effective in improving teacher knowledge (Appleton, 2008; Shallcross et al., 2002). While the difference in classroom OBS ratings between the treatment and control groups was not statistically significant, it is interesting to note that there was a difference in the *SD*s ($SD = 0.71$ in the treatment group, $SD = 1.04$ in the control group). This difference suggests that control group teachers were more likely to add both correct and incorrect science information to their lessons than treatment group teachers.

It is not surprising that the control group also improved in TEST and SKS. Arzi and White (2008) found that the required curriculum was a major source of teacher CK, and Shallcross et al. (2002) found that teaching a topic increased the teacher's confidence with that topic, which is related to SKS. However, the greater increase in each of these measures found in the treatment group suggests that the PD was successful at improving teacher SCK. It is also interesting to note that number of years teaching had a small but positive impact on student science achievement (i.e., more years teaching led to higher student achievement), and it would be interesting to explore this relationship further.

Effect of Teacher SCK on Student Science Achievement. While lack of teacher SCK is considered a cause of low student science achievement (Fleer, 2009), the relationship between the two has not been well described in the literature. Teacher SCK as measured by POSTTEST had a significant effect on student science achievement outcomes. The teacher science test scores predicted 6% of the variance found between teachers' mean student science test scores, regardless of participation in the intervention. The finding of a variable that predicts 6% of the variance found between teachers is important, because it suggests that teachers participating in PD that directly targets their SCK at the grade level they are teaching may have a direct positive effect on student achievement. However, further research must be done to determine whether it is the teacher SCK, the teacher test-taking skills, or a combination of the two that causes this effect.

Teacher self-reported science knowledge did not predict student science achievement. Courses taken by a teacher also did not predict improved student performance in any of the models, which is consistent with the literature (Ball et al., 2001; Rice, 2005), but it is still a very commonly used proxy of CK in the literature. OBS also did not predict student science achievement outcomes. Although observations are considered one of the most unbiased methods of data collection, they are also one of the most expensive (Desimone, 2009), and this study suggests they are less effective for measuring SCK than more cost-effective, easier methods such as science knowledge tests.

Implications

Contributions and Limitations. Although "teachers are now seen as the single most important factor in terms of school variables for producing student learning" (Porter, 2012, 36 minutes, 55 seconds), very few studies have measured the SCK of practicing teachers, and even fewer have used large samples. This study makes an important contribution to the literature. First, the study examined the effect of an intervention on teacher SCK using a large sample of practicing fifth grade teachers and four different measures of SCK. Second, this study examined the effect of SCK on student science achievement, a relationship that has not been well described in the literature. Third, random assignment of schools into the treatment and control groups allowed us to test the causality of the intervention improving teacher SCK. Finally, the randomly selected sample of teachers in a large school district allows generalizability of the results to the district in this study and other similar districts.

This study had several limitations. First, there was a ceiling effect on the SCK test. Five teachers earned perfect scores on PRETEST and 48 answered almost 90% of the questions correctly, leaving very little room for improvement. Second, there was another ceiling effect on the SCK scale. Twenty-one teachers rated their knowledge as a four out of possible four on all items making up PRESKS. These ceiling effects likely contributed to the small effect sizes observed in the comparisons between the treatment and control groups. The addition of more difficult items to both measures would allow for a more sensitive analysis and would likely demonstrate more change in teacher SCK. Third, more frequent classroom observations (than the three in the study) would be desirable for more reliable results, considering variations among science topics and class activities for curriculum implementation (Desimone, 2009). In addition, the number of teachers observed was a subsample of the larger group, requiring separate analyses using a smaller sample than other analyses. OBS data may also be affected by the fact that teachers scheduled the time when the observer came, and thus may have chosen lessons they were more comfortable with or knowledgeable in to be observed. Finally, the pre–post design could also be improved by including more longitudinal measures in the form of additional years of data collection.

Implications for Further Research. The results of this study can have an impact on future research. First, while our sample was large compared to most educational studies, a larger sample spanning a larger geographic region would allow for more generalizable results. Second, the direct effect of teacher SCK on student achievement indicates the need for further research on teacher SCK to supplement the existing research on pedagogical CK. Third, future studies could also examine the similarities and differences between teacher and student SCK as constructs. Finally, quality tests of teacher SCK should be developed, because that seems to be the most predictive method of measuring teacher SCK when student achievement on standardized science tests is the outcome of interest. The question of how SCK is measured should be readdressed in the science education community, because improving teacher SCK is a promising avenue for improving student achievement outcomes.

Implications for Practice. The results of this study can have a potential impact on how school districts design and implement their PD programs. As elementary school teachers are increasingly expected to have strong CK, many teachers may benefit from PD that updates SCK. The fields of science are constantly changing. For example, genomics used to be an obscure branch of biology but is now considered a central field that could come up even in an elementary science classroom. Also, elementary school teachers may be unaware that birds are now considered by most biologists to be a type of dinosaur and therefore a type of reptile (Futuyma, 2009). Teachers who have more content-based PD will have more accurate and up-to-date SCK than those who depend solely on what they learned in college. The results of this study indicate that the intervention had a significant effect on teacher SCK, so the intervention can be used as an example for curriculum writers and PD providers who wish to improve teacher SCK. As teacher SCK is a changeable attribute determining children's education, it is imperative that educators understand how to measure and improve this construct efficiently and effectively.

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090281 to the University of Miami. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Notes

[1]It should be noted that gender and ethnicity were considered in the model but neither were significant predictors, and together they accounted for a small proportion of the variance. Therefore, they were not used in the model.

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. Educational Assessment, 8(3), 231–257.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. Applied Measurement in Education, 14(3), 219–234.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. Educational Measurement: Issues and Practice, 19(3), 16–26.

Appleton, K. (2008). Developing science pedagogical content knowledge through mentoring elementary teachers. Journal of Science Teacher Education, 19(6), 523–545. doi: 10.1007/s10972-008-9109-4

Arzi, H.J., & White, R.T. (2008). Change in teachers' knowledge of subject matter: A 17-year longitudinal study. Science Education, 92(2), 221–251. doi: 10.1002/sce.20239

Bacharach, V., Baumeister, A., & Furr, R.M. (2003). Racial and gender science achievement gaps in secondary education. Journal of Genetic Psychology, 164(1), 115–126.

Ball, D.L., Lubienski, S.T., & Mewborn, D.S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), Handbook of research on teaching (4th ed., pp. 433–456). Washington, DC: American Educational Research Association.

Basista, B., & Mathews, S. (2002). Integrated science and mathematics professional development programs. School Science & Mathematics, 102(7), 359. Retrieved from http://iiiprxy.library.miami.edu:4420/login.aspx?direct=true&db=aph&AN=7887984&site=ehost-live

Begle, E.G. (1979). Critical variables in mathematics education: Findings from a survey of the empirical literature. Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Ben-Peretz, M. (2011). Teacher knowledge: What is it? how do we uncover it? what are its implications for schooling? Teaching & Teacher Education, 27(1), 3–9. doi: 10.1016/j.tate.2010.07.015

Boyd, D.J., Grossman, P.L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher preparation and student achievement. Educational Evaluation and Policy Analysis, 31(4), 416–440. doi: 10.3102/0162373709353129

Chapin, J. (2006). The achievement gap in social studies and science starts early: Evidence from the early childhood longitudinal study. The Social Studies, 97(6), 231–238.

Chinnappan, M., & Lawson, M. (2005). A framework for analysis of teachers' geometric content knowledge and geometric knowledge for teaching. Journal of Mathematics Teacher Education, 8(3), 197–221. doi: 10.1007/s10857-005-0852-6

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. (2nd ed.). Hillsdale, NJ: Erlbaum Associates.

Corn, J.O. (2010). Investigating the quality of the school technology needs assessment (STNA) 3.0: A validity and reliability study. Educational Technology Research and Development, 58(4), 353–376. doi: 10.1007/s11423-009-9140-y

Cox, C., & Carpenter, J. (1989). Improving attitudes toward teaching science and reducing science anxiety through increasing confidence in science ability in inservice elementary school teachers. Journal of Elementary Science Education, 1(2), 14–34. doi: 10.1007/BF03173020

D'Agostino, J.V., & Powers, S.J. (2009). Predicting teacher performance with test scores and grade point average: A meta-analysis. American Educational Research Journal, 46(1), 146–182. doi: 10.3102/0002831208323280

Desimone, L.M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. Educational Researcher, 38(3), 181–199. doi: 10.3102/0013189X08331140

Diamond, B.S., Maerten-Rivera, J., Rohrer, R., & Lee, O. (2013). Elementary teachers' science content knowledge: Relationships among multiple measures. Florida Journal of Educational Research, 51. Retreived from http://feraonline.org/fjer/2013/Diamond.51%281%29.pdf

Duncan, G., Yeung, W.J., Brooks-Gunn, J., & Smith, J. (1998). How much does childhood poverty affect the life chances of children? American Sociological Review, 63(3), 406–423.

Fleer, M. (2009). Supporting scientific conceptual consciousness or learning in 'a roundabout way' in play-based contexts. International Journal of Science Education, 31(8), 1069–1089. doi: 10.1080/09500690801953161

Futuyma, D.J. (2009). Evolution. (2nd ed.). Sunderland, MA: Sinauer Associates.

Garet, M.S., Porter, A.C., Desimone, L.M., 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. doi: 10.3102/00028312038004915

Gordon, S.C., Dembo, M.H., & Hocevar, D. (2007). Do teachers' own learning behaviors influence their classroom goal orientation and control ideology? Teaching and Teacher Education, 23(1), 36–46. doi: 10.1016/j.tate.2004.08.002

Guo, G. (1998). The timing of the influences of cumulative poverty on children's cognitive ability and achievement. Social Forces, 77(1), 257–287.

Gustin, W.C., & Corazza, L. (1994). Mathematical and verbal reasoning as predictors of science achievement. Roeper Review, 16(3), 160–163.

Hanushek, E., & Rivkin, S. (2010). Constrained job matching: Does teacher job search harm disadvantaged urban schools? (Working Paper No. 15816). Cambridge, MA: National Buerau of Economic Research.

Heller, J.I., Daeler, K.R., Wong, N., Shinohara, M., & Miratrix, L.W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. Journal of Research in Science Teaching, 49(3), 333–362. doi: 10.1002/tea.21004

Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure. Theory Into Practice, 41(1), 30–41.

Jüttner, M., Boone, W., Park, S., & Neuhaus, B.J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). Educational Assessment, Evaluation and Accountability, 25(1), 45–67. doi: 10.1007/s11092-013-9157-y

Kanter, D.E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content and pedagogical content knowledge and inquiry-based practices. Science Education, 94(5), 855–887. doi: 10.1002/sce.20391

Kaya, O.N. (2009). The nature of relationships among the components of pedagogical content knowledge of preservice science teachers: 'Ozone layer depletion' as an example. International Journal of Science Education, 31(7), 961–988. doi: 10.1080/09500690801911326

Klassen, R.M., Bong, M., Usher, E.L., Chong, W.H., Huan, V.S., Wong, I.Y.F., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. Contemporary Educational Psychology, 34(1), 67–76. doi: 10.1016/j.cedpsych.2008.08.001

Koretz, D., & Hamilton, L. (1999). Assessing students with disabilities in Kentucky: The effects of accommodations, format, and subject (CSE technical report 498). University of California, Los Angeles: Center for the Study of Evaluation.

Koretz, D., & Hamilton, L. (2001). The performance of students with disabilities on the New York Regents comprehensive examination of English (CSE technical report 540). University of California, Los Angeles: Center for the Study of Evaluation.

Kunter, M., Tsai, Y., Klusmann, U., Brunner, M., Krauss, S., & Baumert, J. (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. Learning and Instruction, 18(5), 468–482. doi: 10.1016/j.learninstruc.2008.06.008

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? Educational Researcher, 31(3), 3–12.

Lee, J. (2007). The testing gap: Scientific trials of test-driven school accountability systems for excellence and equity. Charlotte, NC: Information Age Publishing.

Lee, O., & Maerten-Rivera, J. (2012). Teacher change in elementary science instruction with English language learners: Multi-year professional development intervention across multiple grades. Teachers College Record, 114(8), 1–44.

Madaus, G., & Clarke, M. (2001). The impact of high-stakes testing on minority students. In M. Kornhaber & G. Orfield (Eds.), Raising standards or raising barriers: Inequality and high stakes testing in public education. (pp. 85–106). New York, NY: The Century Foundation.

Maerten-Rivera, J., Adamson, K., Ahn, S., (2012 November). *An examination of the validity and reliability of constructs used to measure teacher's science knowledge and practices in a longitudinal intervention*. Paper presented at the annual meeting of the Florida Educational Research Association (FERA), Gainesville, FL.

Maerten-Rivera, J., Huggins, C., Adamson, K., (2013 November). *The development and validation of measures of teachers' science knowledge and teaching practices collected through surveys and classroom observations*. Paper presented at the annual meeting of the Florida Educational Research Association (FERA), Gainesville, FL.

Maerten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. Science Education, 94(6), 937–963.

Mau, W., & Lynn, R. (2000). Gender differences in homework and test scores in mathematics, reading and science at tenth and twelfth grade. Psychology, Evolution and Gender, 2, 119–125.

Monk, D.H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. Economics of Education Review, 13(2), 125–145. doi: 10.1016/0272-7757(94)90003-5

Nowicki, B.L., Sullivan-Watts, B., Shim, M.K., Young, B., & Pockalny, R. (2013). Factors influencing science content accuracy in elementary inquiry science lessons. Research in Science Education, 43(3), 1135–1154. doi: 10.1007/s11165-012-9303-4

O'Reilly, T., & McNamara, D. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. American Educational Research Journal, 44(1), 161–196.

Porter, A.C. (2012). Educational policy breakfast series: Teacher quality/effectiveness: Defining, developing, and assessing policies and practices. New York, NY: New York University.

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

Raudenbush, S., Bryk, A., & Congdon, R. (2005). Hierarchical linear and nonlinear modeling (HLM). Lincolnwood, IL: Scientific Software International.

Reis, S.M., & Park, S. (2001). Gender differences in high-achieving students in math and science. Journal for the Education of the Gifted, 25(1), 52–73.

Rice, D. (2005). I didn't know oxygen could boil! what preservice and inservice elementary teachers' answers to 'simple' science questions reveals about their subject matter knowledge. International Journal of Science Education, 27(9), 1059–1082. doi: 10.1080/09500690500069426

Rice, J.K. (2003). Teacher quality: Understanding the effectiveness of teacher attributes. Washington, DC: Economic Policy Institute.

Shallcross, T., Spink, E., Stephenson, P., & Warwick, P. (2002). How primary trainee teachers perceive the development of their own scientific knowledge: Links between confidence, content and competence? International Journal of Science Education, 24(12), 1293–1312. doi: 10.1080/09500690110110106

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4–14. doi: 10.3102/0013189X015002004

Singer, J.D., & Willett, J.B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. New York, NY: Oxford University.

Supovitz, J.A., Mayer, D.P., & Kahle, J.B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. Educational Policy, 14(3), 331–356. doi: 10.1177/0895904800014003001

Supovitz, J.A., & Turner, H.M. (2000). The effects of professional development on science teaching practices and classroom culture. Journal of Research in Science Teaching, 37(9), 963–980. doi: 10.1002/1098-2736(200011)37:9<963::AID-TEA6>3.0.CO;2-0

Van Driel, J.H., De Jong, O., & Verloop, N. (2002). The development of preservice chemistry teachers' pedagogical content knowledge. Science Education, 86(4), 572. Retrieved from http://iiiprxy.library.miami.edu:4420/login.aspx?direct=true&db=aph&AN=7184184&site=ehost-live

Wayne, A., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. Review of Educational Research, 73, 89–122.

Yuill, N., & Oakhill, J. (1991). Children's problems in text comprehension: An experimental investigation. Cambridge, England: Cambridge University.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.