Effects of a Multiyear Curricular and Professional Development Intervention on Elementary Teachers' Science Content Knowledge

Brandon S. Diamond
University of Miami

Jaime Maerten-Rivera
University of Miami

Okhee Lee
New York University

Abstract

This study examined the effectiveness of a fifth grade science curriculum and professional development intervention at increasing teachers' science content knowledge (SCK) over the period of 3 years. The intervention included an inquiry-based science curriculum and ongoing professional development opportunities concentrating on both science content and pedagogy. SCK was measured using a science knowledge test written at the fifth grade content level and a questionnaire scale asking teachers how knowledgeable they felt teaching science at the fifth grade level. Longitudinal multilevel modeling was used to examine change in the treatment group and the control group. The treatment group demonstrated an increase in both measures of SCK after the first year, with continued improvement in self-report throughout the 3-year period. The control group demonstrated smaller consistent growth in test scores, reaching a level comparable to the treatment group at the end of 3 years, but there was no significant change in self-report over the 3-year period. These results led to the conclusion that the intervention increased the rate at which teachers learned science content, as well as increasing their confidence in their SCK.

Keywords: science content knowledge, professional development, longitudinal model, hierarchical linear modeling, science tests, questionnaires

Introduction

While it has long been recognized that teacher quality is extremely important to student success (Van Driel, Berry, & Meirink, 2014), surprisingly little research has addressed how to improve teachers' science content knowledge (SCK) until recently (Fleer, 2009). The research examining SCK usually involves preservice teachers rather than inservice teachers (Arzi & White, 2008; Ball, Lubienski, & Mewborn, 2001). In addition, few rigorous studies have examined if, and how much, teacher SCK increases during professional development (PD) opportunities (Garet, Porter, Desimone, Birman, & Yoon, 2001), and even fewer studies have examined change in teacher SCK over time. It is important to better understand the development of teacher SCK, because teacher SCK has a direct effect on student science outcomes (Diamond, Maerten-Rivera, Rohrer, & Lee, 2014).

The purpose of this study was to examine change over time in SCK of elementary school teachers according to two different measures of SCK (i.e., a science knowledge test and a science knowledge questionnaire scale) over a 3-year curricular and PD intervention using a randomized controlled trial (RCT). The intervention was created as a collaboration between a university and a large urban school district to maximize inquiry-based learning and understanding of science concepts by all students, especially English learners (ELs), at the fifth grade level. To examine the intervention's effect on teacher SCK, we examined the following two research questions (R.Q.):

> **R.Q. 1.** What was the status of teacher SCK at the start of the intervention (i.e., baseline status)?
>
> a. How knowledgeable were teachers at the start of the intervention in the treatment group and the control group?
>
> b. Did teacher background variables (i.e., number of years teaching, highest degree, number of college science courses taken) predict teacher SCK at the start of the intervention in the treatment group and the control group?
>
> **R.Q. 2.** How did teacher SCK change over the intervention?
>
> a. Was there change in teacher SCK over the course of the intervention in the treatment group and the control group?
>
> b. Did teacher background variables predict change in teacher SCK over the course of the intervention in the treatment group and the control group?

This study makes several contributions to the literature. First, the study offers insight regarding the SCK of inservice elementary teachers, along with teacher characteristics predicting SCK. Second, the study uses random selection of schools from a district and random assignment into the treatment and control groups. Third, the study examines whether a PD implemented with the recommended core features (Garet et al., 2001), including a focus on content, leads to increased SCK over a period of 3 years while using a control group for comparison. Finally, the study is unique because it utilized two different measures of SCK. One measure is a self-report questionnaire with items measuring the teachers' perceived SCK, which is similar to the ones frequently used in other studies (Banilower, Heck, & Weiss, 2007; Lee & Maerten-Rivera, 2012; Supovitz & Turner, 2000). The other measure is a paper-based test, which is a more objective measure of SCK than the self-report questionnaire. These two measures allow the evaluation of the intervention using a common style of measurement (i.e., self-report) and a style that is predictive of student success (i.e., paper-based test) (Diamond et al., 2014). The test includes multiple-choice and constructed response items to measure the science content taught at the fifth grade level according to the state science standards.

Literature Review

Surprisingly, very little of the research on teacher knowledge looks at how to improve teachers' science content knowledge (SCK; Fleer, 2009; Heller, Daeler, Wong, Shinohara, & Miratrix, 2012; Shallcross, Spink, Stephenson, & Warwick, 2002), how to measure it, or how to evaluate its impact on teacher practice or student achievement (Chinnappan & Lawson, 2005; Porter, 2012). In fact, on the rare occasions that teacher SCK is addressed, it is usually studied in preservice teachers rather than those currently teaching (Ball, Lubienski, & Mewborn, 2001).

The lack of research on CK is largely due to the introduction of the construct of "pedagogical content knowledge" (PCK) by Lee Shulman in 1986. Since then, forms of teacher knowledge other than CK have taken priority in studies (Van Driel et al., 2014), whereas CK is rarely studied except in how it relates to PCK. While CK describes how well a teacher knows a specific topic, PCK describes how well a teacher knows how to effectively teach a specific topic. For example, while CK would allow a teacher to know that a linoleum floor has friction, PCK would allow the teacher to expect students to believe it does not have friction because it can be slippery. Without a strong CK, a strong PCK is impossible to achieve (Kaya, 2009; Van Driel et al., 2014; Van Driel et al., 2002) because teachers cannot know how to effectively teach content that they do not know. Thus, while PCK is an important construct, its study does not replace the need for an understanding of CK itself. In science education, there is insufficient understanding of the SCK needed to teach science effectively. By focusing on SCK, this study is aimed at filling this gap in the literature that predominantly addresses PCK. Below, we discuss two areas of the literature: the importance of SCK, including issues of measurement, and the effect of PD on teacher SCK with consideration of longitudinal designs, which are the areas of focus in this study.

Importance of SCK

SCK and classroom instruction

Teachers' lack of SCK has been cited as a primary cause of their inability to teach science effectively (Fleer, 2009). The use of inquiry in science instruction is important because it is an effective method for enabling students to construct scientific understanding (Jarvis, Pell, & McKeon, 2003), and SCK is a major predictor of teachers' use of inquiry-based science teaching (Supovitz & Turner, 2000). Adequate SCK 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). In fact, variations of teacher SCK have been identified as the main factor responsible for the differences in the quality of elementary science teaching (Shallcross et al., 2002). While mathematics CK has been studied in teachers much more than SCK, there is still not sufficient understanding of the mathematics CK it takes to teach effectively (Ball et al., 2001). Without knowing what kind of CK is necessary for effective teaching, it is difficult to help teachers develop the knowledge they need.

Of the limited literature, more studies examined the SCK of preservice teachers than the SCK of inservice teachers (Arzi & White, 2008; Ball et al., 2001). Recently, some studies focused on the SCK of inservice teachers in various instructional contexts (Diamond, Maerten-Rivera, Rohrer, & Lee, 2013; Diamond et al., 2014; Heller et al., 2012; Jüttner, Boone, Park, & Neuhaus, 2013; Nowicki, Sullivan-Watts, Shim, Young, & Pockalny, 2013). Nowicki et al. (2013) found that elementary teachers had gaps in their SCK, and these gaps were a major obstacle to effective teaching. The gaps may be attributed to inadequate science content in preservice teacher preparation programs. For teachers entering the profession with little SCK, PD can increase their SCK (Nowicki et al., 2013).

Measurement of SCK

Teacher SCK has typically been assessed using self-reports or classroom observations (Diamond et al., 2013). Few studies administered tests to inservice teachers as a measure of SCK partly because such tests are perceived to be disrespectful or threatening (Arzi & White, 2008). Recently, tests of SCK were used with inservice teachers to assess their SCK (Jüttner et al., 2013; Nowicki et al., 2013) and to examine the effect of PD on SCK (Diamond et al., 2014; Heller et al., 2012). For example, Jüttner et al. (2013) developed a test designed to distinguish between teacher SCK in biology and PCK. They used Rasch analysis to determine that these two types of knowledge were correlated but significantly different constructs that could be measured independently of each other.

Effect of Professional Development on Teacher Content Knowledge

Recognizing the importance of PD for improving teacher knowledge and practices and student outcomes, researchers identified elements of effective PD (Garet et al., 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Wayne & Youngs, 2003) and a causal model for evaluating PD programs (Desimone, 2009). To build a knowledge base about links PD, teacher knowledge and practices, and student outcomes, researchers called for more rigorous study designs (Borko, 2004; Desimone, 2009; Wayne & Youngs, 2003), including the use of randomized experiments (Wayne & Youngs, 2003). One obstacle to conducting rigorous studies is the lack of adequate measures, particularly in regard to teacher learning (Desimone, 2009).

The effect of PD on teacher knowledge and practices in science was studied extensively (Banilower et al., 2007; Diamond et al., 2014; Garet et al., 2001; Heller et al., 2012; Lee & Maerten-Rivera, 2012; Penuel et al., 2007; Supovitz, Mayer, & Kahle, 2000; Supovitz & Turner, 2000). In this section, we discuss only those studies whose researchers examined the effects of PD on teacher SCK. Additionally, some of these PD researchers examined teacher change in PCK by asking questions pertaining to increased knowledge of curriculum, instruction, and assessment.

Garet el al. (2001) studied the effectiveness of PD on teacher outcomes in their evaluation of the Eisenhower Professional Development Program. They used teacher self-report to examine elements of effective PD in terms of both core and structural features of PD. Except for the form of the activity, the core and structural features had positive effects on teachers' self-reported knowledge and practices. Two other studies examining the effects of PD on teacher SCK used a paper-based test (Diamond et al., 2014; Heller et al., 2012). Both studies found that the treatment group teachers improved their test scores after PD more than the control group teachers, but neither study examined whether there was additional improvement over time.

Banilower et al. (2007) used a measure of teachers' perceptions of pedagogical preparedness along with a measure of their perceptions of science content preparedness, which is similar to self-reported SCK, to teach topics commonly covered in K-8 science classrooms. The results indicated that participation in PD had a significantly positive effect on science content preparedness. Most of the change occurred within the first 80 hours of PD, although teachers receiving the most PD still had substantial room for further growth on the measure. The number of years teaching was a significant predictor of change in the model, with more experienced teachers feeling more prepared. While this study employed longitudinal analysis, it did not use a control group to determine causal relationships and relied solely on self-report data to measure teacher SCK.

Lee and Maerten-Rivera (2012) examined the effect of a multiyear PD intervention on elementary science teachers' perceptions of how knowledgeable they felt teaching eight different science topics at baseline and at the end of each year of participation over a 3-year period. While third and fourth grade teachers did not report increased SCK, fifth grade teachers did, which was the grade level at which a high-stakes science test counted toward school accountability. Most of the fifth grade teachers' growth occurred during their first year of participation in the intervention and was sustained in subsequent years. This pattern of growth was found in another study of PD's effects on teacher practices (Supovitz et al., 2000), where the researchers termed it "short-term growth and long-term stability" (p. 342). The researchers examined additional predictor variables in the model, including highest degree, number of science college courses, years teaching, and grade taught. Only number of science college courses was found to be a significant predictor of self-reported SCK. This study used a longitudinal design, but it did not use a control group to determine causal relationships and relied solely on self-report data to measure teacher SCK.

Summary of the Literature

While PCK is unquestionably an important construct for teacher effectiveness, SCK has not been as well studied in recent decades. This is despite the fact that it is well documented that PCK is impossible to achieve without SCK (Kaya, 2009; Van Driel et al., 2014; Van Driel et al., 2002). There is even less research on how to measure change in SCK over time or how to evaluate the effect of PD on change in SCK over time. If science teacher educators are going to improve student achievement by improving teacher practices, addressing teacher SCK is a logical place to begin.

Method

Research Setting

The research was conducted in a large urban school district in the Southeast United States with linguistically and culturally diverse student and teacher populations. The ethnic makeup of the student population in the school district in 2010 was 65% Hispanic, 24% Black non-Hispanic, 9% White non-Hispanic, and 2% Other. For the 2010-2011 school year, the school district reported that 72% of elementary students were eligible for free or reduced price lunch programs, 19% were designated as ELs enrolled in English to Speakers of Other Languages programs, and 11% were enrolled in exceptional student education programs.

Research Design

The study used a cluster randomized controlled design. First, 64 schools were randomly selected from a pool of 206 available schools (excluding 23 due to district monitoring and nine due to their participation in our previous project). Then the 64 schools were randomly assigned to 32 treatment and 32 control schools. One treatment school chose not to continue into Year 2. All schools that participated in Year 2 also participated in Year 3. Thus, 31 treatment and 32 control schools were included in the 3-year study.

Table 1. *Teacher Demographics*

Variable	Demographic group	Total *N* = 391 %	Treatment *n* = 212 %	Control *n* = 179 %
Gender	Male	16.9	15.6	18.4
	Female	83.1	84.4	81.6
Ethnicity	Hispanic	53.2	52.8	53.6
	Black non-Hispanic	22.8	22.2	23.5
	White non-Hispanic	19.2	19.8	18.4
	Haitian	1.5	0.9	2.2
	Asian	0.5	0.9	0.0
	Other (including multiracial)	2.3	2.8	1.7
	Missing	0.5	0.5	0.6
Native language(s)*	English	82.9	83.0	82.7
	Spanish	43.7	42.9	44.7
	Other	8.2	12.3	3.4

* Multiple categories could be selected.

As a school-wide initiative, Year 1 of the study involved a total of 223 fifth grade teachers from 64 schools. Year 2 of the study involved a total of 197 fifth grade teachers, 70 of whom were new to the study, from 63 schools. Year 3 of the study involved a total of 194 fifth grade teachers, 42 of whom were new to the study, from 63 schools. The demographic information for all 391 participating teachers, including those with missing data, is shown in Table 1. To determine if differences existed at baseline between teachers with missing data (i.e., teachers who left the study) and those who remained for all 3 years, analysis of variance (ANOVA) models were employed for both baseline SCK measures (i.e., a science knowledge test and a science knowledge questionnaire scale) and for number of science courses taken with attrition and group (i.e., treatment or control) included in the model. No significant differences were found, and the effect sizes were small. Furthermore, the difference in number of years teaching demonstrated a significant but small effect size.

The intervention was confined to the teachers in the treatment group over the 3-year period. As the intervention involved all fifth grade science teachers from the participating treatment schools, there was little contamination between the treatment and control group teachers. The project website was password-protected, changed each year, and was accessible only to the treatment group teachers. When the treatment teachers left schools, the intervention materials stayed at the treatment schools. Over the 3-year period, only one treatment teacher moved to a control school during her first year of participation in the intervention. While this teacher provided the baseline data, she did not provide any post-intervention data. Overall, threats to internal validity due to teacher crossover between the treatment and control groups were minimal.

Curricular and Teacher Professional Development Intervention

In our 3-year intervention, curriculum, PD, and school site support were designed to complement one another with the goal of improving teacher SCK and practices according to state science standards for diverse student groups, including ELs, in a high-stakes assessment and accountability policy context. The intervention was grounded in three areas: (a) reform-oriented practices to promote students' science inquiry and understanding, (b) state science standards and high-stakes science assessment, and (c) science instruction for diverse student groups with a focus on ELs. Throughout the 3 years of the intervention, the control schools implemented the district-adopted textbook-based science curriculum that included kit-based hands-on science activities. The teaching methods employed by the control teachers varied widely by teacher, but the topics taught by both treatment and control group teachers were specified by the state science standards. Below, key aspects of our intervention are described.

Science inquiry

Our intervention was designed to promote science inquiry with students from diverse languages and cultures as they ask questions, develop their understanding, and share their understanding with others (see Figure 1). The way students engage in inquiry varies depending on their prior experience with science, science disciplines or topics, and the kinds of questions explored. In particular, the last two steps of the inquiry process, inquiry extension and application, provide students with opportunities to use their initiative and creativity. The intervention included multiple opportunities for students to develop and refine questions and apply their understanding to explain phenomena. For example, the curriculum provided teachers with probing questions or prompts in the teachers' guide. Also, at PD workshops, teachers practiced how to encourage students to generate their own questions and seek their own answers.

While all students are able to "do" science, they need support to learn to engage in inquiry. To develop students' inquiry abilities, teachers provide a differentiated level of assistance. Some students are more ready to do inquiry on their own and need little assistance, whereas others require extensive assistance over an extended period of time. As students develop inquiry abilities, they need increasingly less help until they are able to do inquiry on their own.

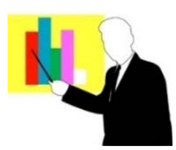
Science Inquiry	
1. Questioning 	**Ask a question** ▪ What do I want to find out? **Make a prediction** ▪ What do I think will happen? ▪ What is my reasoning for the prediction?
2. Planning 	**Read the materials and procedures** ▪ Do I have all of the necessary materials? ▪ Have I read the procedures? ▪ Summarize the procedures in my own words.
3. Implementing  	**Gather the materials** ▪ What materials do I need to implement the plan? **Implement the plan** ▪ What steps do I need to take to implement the plan? **Observe and record the data** ▪ What happens when I implement the plan? ▪ What do I observe? ▪ How do I display the data? (a graph, chart, table)
4. Concluding 	**Draw a conclusion** ▪ What did I find out? ▪ What is the evidence? ▪ Was my prediction supported or not? (Remember: It is okay if the prediction is not supported.)
5. Reporting 	**Share my results (informal)** ▪ What results do I want to share with others orally? **Produce a report (formal)** ▪ What results do I want to share with others in writing?
6. Inquiry Extension 	**Reflect on your results** ▪ If I would do this investigation again, how would I improve it? ▪ What would be a good follow-up investigation based on what I have learned?
7. Application 	**Make connections** ▪ How does this investigation relate to what happens in the real world? ▪ How could I apply the results in new situations?

Figure 1. *Intervention inquiry framework.*

Curriculum

A comprehensive, stand-alone, year-long science curriculum for fifth grade was developed for this intervention. The intervention provided each treatment teacher with a teachers' guide, student books, and supplies. The class set of student books was provided each school year, and consumable supplies were replenished at the start of each subsequent school year. Supplementary materials were made available via CD and online via the project's website. The science standards with which the curriculum was aligned are summarized in Appendix A of the supplementary materials.

Teachers in the control group implemented the district-adopted science curriculum. The fifth grade textbook consisted of 18 chapters in four units: life science, physical science, earth science, and human body. The textbook included hands-on activities with science supplies provided with adoption, but no replenishments.

The intervention curriculum differed from the district-adopted textbook with regard to three key features. First, the intervention curriculum was closely aligned to the state science standards and stayed within the content limits of these standards. In contrast, the district-adopted textbook went beyond the content limits, leaving teachers to decide curriculum coverage. Second, the intervention curriculum incorporated inquiry both as a learning outcome and as a means of fostering science understanding. Also, the intervention curriculum provided enough science supplies for students to conduct the hands-on activities that were integral to developing scientific understanding. In contrast, the district-adopted textbooks had fewer and less fully developed hands-on activities that were optional and not closely connected to science concepts addressed in lessons/chapters. Finally, the intervention curriculum focused on language development for all students and for ELs in particular. Intervention students shared ideas orally in small group discussions, wrote their findings extensively in the student book, and reported their conclusions in whole group discussions. The intervention curriculum offered supplementary materials, including language development activities for ELs of lower proficiency levels.

Professional Development Workshops

Teacher workshops focused on curriculum implementation by enabling teachers to realize the intentions of the curriculum and to utilize the curriculum as a scaffold to promote teacher learning in addition to student learning (i.e., the notion of "educative curriculum material" by Davis & Krajcik, 2005). The workshops incorporated critical features of effective PD (Desimone, 2009; Garet et al., 2001; Penuel et al., 2007). First, the workshops focused on reform-oriented teaching practices and teacher SCK to promote students' science inquiry and understanding (i.e., content focus). Second, teachers were actively engaged in hands-on science inquiry and planning for classroom implementation (i.e., active learning). Third, the curriculum was aligned with state science standards and high-stakes science assessment, while also supporting literacy and English language development (i.e., coherence). Fourth, the workshops were offered during the summer and throughout the school year over the 3-year period (i.e., duration). Finally, all fifth grade science teachers from the treatment schools participated in the workshops (i.e., collective participation).

During Year 1, treatment teachers received 5 days of workshops, approximately 6 hours each day, including 3 days in the summer shortly before school started, 1 day in January, and 1 day in May. Given the large number of teachers, the teachers were split into two groups. Teachers were given a stipend for attending the summer workshops, and schools received substitute payments during school days. The workshops focused on familiarizing teachers with the curriculum by demonstrating strategies for hands-on inquiry investigations, as well as highlighting the curriculum's alignment to the state science standards and high-stakes science assessment. The workshops focused on allowing the teachers to carry out the science investigations their students would be doing throughout the school year and to discuss how to handle potential difficulties that might arise.

The summer workshop covered the science investigations that were carried out in the first 4 months of school, and the January workshop covered the science investigations that were carried out between Winter Break and the state science assessment. The intervention especially helped teachers recognize how students' science inquiry abilities were related to the state science standards and, thus, could enhance performance on the high-stakes science assessment. Additionally, strategies for English language development were introduced as they were embedded in science inquiry and understanding.

During Year 2, teachers in the treatment group were offered 5 days of workshops that followed the same schedule as in Year 1. The 3-day summer workshops differed between teachers new to the intervention and returning teachers. Teachers new to the intervention were introduced to the curriculum similarly to the teachers during Year 1, in a different room but at the same time as returning teachers. With returning teachers, the focus was on highlighting changes to the new state science standards and the new high-stakes science assessment aligned with the new standards. Building on PD during the first year, the summer workshop for returning teachers highlighted further professional growth in science instruction (e.g., student initiative in science inquiry, common misconceptions), integration of science with English language development, and incorporation of students' home language and culture in science instruction. At the 1-day January workshop, new and returning teachers were combined into one group to become familiar with the curriculum that was revised to align with the new state science standards and high-stakes science assessment, and returning teachers shared their experiences from the previous year with teachers new to the intervention. The 1-day May workshop with both the new and returning teachers focused on unpacking science inquiry, highlighting opportunities for more open-ended questioning in science instruction, helping teachers become more comfortable with and willing to implement inquiry, making teachers aware of what is a good student response, providing constructive feedback to students, and developing and using a rubric for assessing science inquiry.

During Year 3, summer workshops differed between teachers new to the intervention and returning teachers. Teachers new to the intervention participated in a 3-day summer workshop to familiarize themselves with the curriculum. Returning teachers participated in a 1-day summer workshop to help them implement student-centered inquiry and language development strategies for ELs. They were provided with students' work samples to discuss how to promote student-centered inquiry and how to pay explicit attention to language development strategies for ELs. Then, they used a rubric to assess students' work samples with regard to science inquiry and language development. At the 1-day meeting in May, both the new and returning teachers convened to offer feedback on the intervention and reorient to the district adopted curriculum to which they would be transitioning back the following school year. A more detailed description of the PD can be found in previously published articles to be named once blind review is complete.

School site support

Members of the research team were assigned to school site support to provide teachers with mentoring and to ensure fidelity of implementation during all 3 years of the intervention. Three members of the research team were assigned in Years 1 and 2 and two members in Year 3. Visits to schools were scheduled generally every 4 to 6 weeks. Some schools were visited more often if a school or teacher requested assistance or if the project staff or the school district administrators identified a school or teacher to be high-needs compared to the other schools or teachers in the treatment group. Support included co-planning, co-teaching, modeling, delivering supplies, providing additional resources (e.g., home-learning suggestions, lab management strategies, and supplemental assessment materials), and disseminating and assisting with interpretation of district interim assessment data. This support also provided teachers with the opportunity to ask questions about the curriculum that were not covered in the workshops (e.g., questions about topics not covered by science investigations). The project staff also met regularly with school administrators during school visits to address areas of concern.

Data Collection, Instruments, and Data Analysis

Data Collection

Data collection activities were performed during the teacher workshops for the treatment group and in the individual schools for the control group during each of the 3 years. For each data collection event, teachers were given the informed consent and background information form if they were new to the study and were given the science knowledge test and the questionnaire every time. In Year 1, the treatment group's data were collected at the beginning of the first day of the summer workshop and at the year-end workshop. The control group's data were collected at school sites in the first quarter of each school year and again at the end of the school year. In Years 2 and 3, teachers new to the schools participating in the study were administered the instruments in the same manner as in Year 1, both at the beginning and end of the school year. Returning teachers were administered the instruments at the end of the year only, again using the same procedures as described for Year 1. Control group teachers were given a small stipend for their participation.

In both the treatment and control groups, only teachers who taught science at fifth grade were included. Some schools were departmentalized where one science teacher taught multiple classes of fifth grade science. In Year 1, six treatment and five control schools each had only one science teacher for all fifth grade students. In Year 2, five treatment schools and four control schools each had one science teacher. In Year 3, seven treatment schools and five control schools each had one science teacher.

Time point coding

Table 2 displays the patterns of data collection for the study. Each pattern represents the teachers who participated in data collection at each time point. This table demonstrates how many teachers were missing the data from each time point by totaling the number of teachers who provided data for each time point. For example, most of the teachers (111 in total) provided data for a pretest and a single posttest (Group 6), while no teachers provided data for a second or third year without providing data for a first year (Groups 14 and 15).

When a teacher completed the test and questionnaire prior to beginning the intervention, the time for these measures was coded as baseline or T0. For teachers who participated in Year 1, T0 was the test and questionnaire administered at the beginning of Year 1. For teachers who did not begin participation until Year 2 or 3, T0 was measured by the test and questionnaire administered at the beginning of Year 2 or 3, as appropriate. Time 1 (T1) was the first data collection after participation in the study, which was the end of Year 1 for most teachers, the end of Year 2 for teachers new to the study in Year 2, or the end of Year 3 for teachers new to the study in Year 3. Time 2 (T2) was the end of each teacher's respective second year in the study, and Time 3 (T3) was the end of each teacher's respective third year in the study. T0 was treated as missing data for teachers who did not take a pretest or prequestionnaire (e.g., teacher entered midway in the school year). Therefore, teachers who participated in 1 year of the study were expected to have data for T0 and T1 only, while only teachers who participated in all 3 years of the study were expected to have data at all four time points from T0 to T3.

Table 2. *Patterns of Data Collected.*

Table 2a. *Patterns of Test Data Collected (N =355*).

	Treatment	Control	Time 0	Time 1	Time 2	Time 3
Group 1	49	38	X	X	X	X
Group 2	28	31	X	X	X	
Group 3	1	1	X	X		X
Group 4	0	0	X		X	X
Group 5	1	3		X	X	X
Group 6	49	62	X	X		
Group 7	0	0	X		X	
Group 8	0	0	X			X
Group 9	16	6		X	X	
Group 10	0	0		X		X
Group 11	0	0			X	X
Group 12	11	11	X			
Group 13	26	22		X		
Group 14	0	0			X	
Group 15	0	0				X
Total	181	174	281	333	172	93

Table 2b. *Patterns of Questionnaire Data Collected (N =359*).

	Treatment	Control	Time 0	Time 1	Time 2	Time 3
Group 1	51	38	X	X	X	X
Group 2	32	33	X	X	X	
Group 3	0	0	X	X		X
Group 4	0	0	X		X	X
Group 5	0	3		X	X	X
Group 6	49	66	X	X		
Group 7	0	0	X		X	
Group 8	0	0	X			X
Group 9	14	5		X	X	
Group 10	0	0		X		X
Group 11	0	0			X	X
Group 12	10	12	X			
Group 13	26	19		X		
Group 14	1	0			X	
Group 15	0	0				X
Total	183	176	291	336	177	92

Instruments

Teacher SCK test. The teacher SCK test was aligned with the fifth grade science content standards of the state in which the research took place at the time of developing the measure. The topics included nature of matter, energy, force and motion, processes that shape the Earth, Earth and space, processes of life, living things interacting with the environment, and nature of science (see Appendix A). Two researchers took the lead in searching for test items that mapped onto these topics from two main sources: National Assessment of Educational Progress (NAEP) and Trends in International Mathematics and Science Study (TIMSS). A pilot test was developed with 34 items, including five short and extended response items. The test was piloted with a sample of 311 respondents. The psychometric properties of the test and the items were examined for the pilot test, and a comparison of the difficulty of the items with the ability of the sample was conducted. A panel of researchers, district personnel, and classroom teachers reviewed the pilot test information and chose 30 items that mapped onto the science topics assessed at fifth grade for the final version of the test, which included 24 multiple-choice and six short or extended response items. Each multiple-choice item was worth 1 point, one constructed response item was worth 1 point, two were worth 2 points each, and three were worth 3 points each for a total of 38 points. The TEST score was calculated by summing the points earned. Table 3 displays the test specifications, along with the item difficulty level and source information.

Table 3. *Teacher SCK Test Specifications*

Strand	Low difficulty	Medium difficulty	High difficulty	Total
Nature of matter	8^{t*}, 12^{D*}	7^{t}, 15^{n}	18^{t}	5
Energy	21^{N}	26^{t}, 29^{T}	6^{n*}, 10^{N}	5
Force and motion		13^{D}, 25^{D}, 24^{D}, 9^{T}	2^{T}, 28^{N}	6
Processes that shape the Earth		17^{N}	16^{T*}	2
Earth and space		11^{t}, 23^{T}	30^{T}	3
Processes of life		4^{t*}		1
Living things interacting with environment	(21), 20^{n}	14^{N}	22^{N}, 1^{D}	4
Nature of science	(8*)	3^{n}, 5^{t}, (17), (25) 27^{D*}	(6*), 19^{n}	4
Total	4	16	10	30

() Nature of science is embedded in these items
* Short response
^{D} Project developed item
^{n} NAEP item fourth grade
^{N} NAEP item eighth grade
^{t} TIMSS item third/fourth grade

The test was assessed for unidimensionality by conducting a confirmatory factor analysis (CFA) for all four time points. The model fit the data well at T0, T1, and T2, indicating that the items were all measuring a single construct. However, the model did not fit the data well at T3, which might be attributed to the considerably smaller sample size compared to those at the two previous time points due to teacher attrition over time. Table 4 presents the descriptive statistics and the Cronbach's alpha reliability estimate for each time point for the sample used in the study. Additional detail on TEST development procedures, TEST, and the scoring rubric can be found in previously published articles to be named once blind review is complete.

Table 4. *Descriptive Statistics of Teacher-Level Variables*

Measure	Treatment				Control			
	n	*M*	*SD*	*α*	*n*	*M*	*SD*	*α*
TEST T0	143	29.04	7.83	.70	150	28.59	8.18	.78
TEST T1	171	32.18	5.50	.63	166	30.55	5.95	.78
TEST T2	98	32.06	7.24	.51	80	31.98	5.79	.78

TEST T3	50	32.30	5.49	.51	44	32.75	4.03	.76
SKS T0	143	2.70	0.71	.94	150	2.71	0.84	.95
SKS T1	171	3.15	0.54	.88	166	3.05	0.67	.94
SKS T2	98	3.24	0.60	.93	80	3.18	0.51	.90
SKS T3	50	3.46	0.43	.84	44	3.18	0.91	.98
COURSES	203	4.67	4.45	--	171	4.94	4.92	--
YEARS	204	14.59	7.97	--	177	14.42	9.21	--
DEGREE	209	0.49	--	--	178	0.51	--	--

Teachers' self-reported SCK. The science knowledge scale (SKS) was a section of the questionnaire consisting of four Likert-type questions asking the teachers to indicate how knowledgeable they felt about teaching physical science, earth and space science, life science, and nature of science at their grade level (see Appendix B of the supplementary materials). The questionnaire items used a 4-point rating system, with 1 = not knowledgeable, 2 = somewhat knowledgeable, 3 = knowledgeable, and 4 = very knowledgeable. SKS for each teacher was calculated as the average score (1-4) for the four items. An SKS score was computed for teachers who answered at least three of the four items from the scale. Table 4 presents the descriptive statistics and Cronbach's alpha reliability estimate for each time point. SKS scores were shown to have strong reliability for both the treatment and control groups at all time points.

Correlations between measures were examined in depth in Author (2013). The Pearson correlation between TEST and SKS was .30 ($p < .01$), which is considered a moderate correlation. While these measures are correlated, their correlation is low enough to justify using both measures in the analysis.

Teacher background. The teacher background form collected information on the predictor variables used in the analyses. Predictors were chosen based on previous studies (Banilower et al., 2007; Lee & Maerten-Rivera, 2012; Nowicki et al., 2013). All predictors were self-reported by the teachers during their first data collection. The teacher's total number of science courses taken in college (COURSES), including science methods courses, was examined as a predictor. The teacher's total number of years teaching (YEARS) was also included as a predictor. In addition, the teacher's highest degree obtained was included (DEGREE) as a predictor; teachers with a master's degree or higher were coded as 1, while teachers with a bachelor's degree as their highest degree were coded as 0. Descriptive statistics for COURSES, YEARS, and DEGREE are displayed in Table 4. The intercorrelations among these variables were examined; while the predictor variables were significantly correlated, none were highly correlated.

Data Analysis

While random selection and assignment were conducted at the school level, analysis was conducted at the teacher level, because some schools only had one departmentalized science teacher. Inclusion of groups of one teacher would have made the model unstable, and the PD might have affected teachers from the same school differently.

Patterns of change

As a first step in conducting longitudinal data analyses, the general trends in the data over time were examined. This was done in two ways. First, individual teachers' scores on the outcome variables were graphed to examine the change over time. Second, the group mean and standard deviation for each outcome variable at each time point was examined; this information is displayed in Table 4. The graphs of individual teachers from the treatment group for both outcome variables suggest that change was piecewise (i.e., discontinuous), with most growth occurring from T0 to

T1 and less growth from T1 to T3 (see Figure 2). The group means also suggest a similar pattern. In contrast, the graphs of individual teachers from the control group for both outcome variables suggest that change was linear (i.e., a similar increase each year; see Figure 3). The group means also suggest a similar pattern. Since there were missing data (e.g., a teacher had outcome data for T0 and T1, but missed T2 and T3), this first step provides a general idea of the pattern of growth; however, multilevel modeling (MLM) provides more conclusive results, as described next.

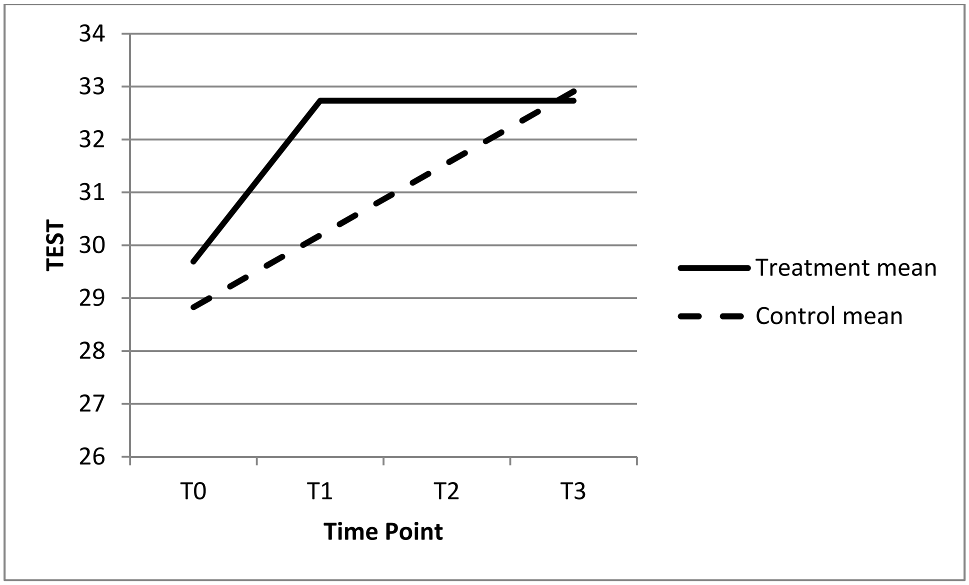


Figure 2. Predicted TEST scores over time.

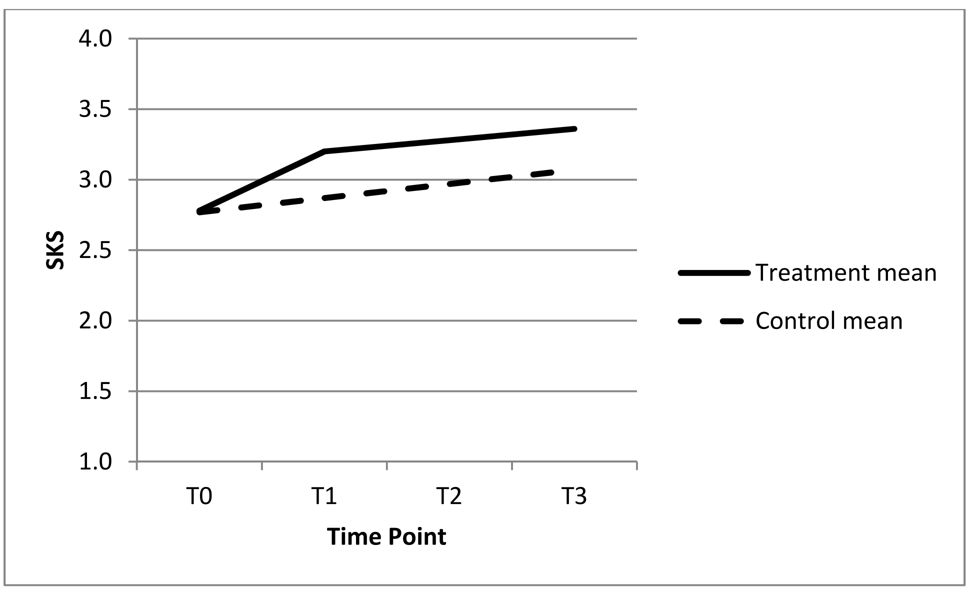


Figure 3. Predicted SKS scores over time.

Multilevel modeling. Longitudinal MLM using HLM 7 was conducted with time nested within teachers (Raudenbush & Bryk, 2002) to examine baseline status in teacher SCK (R.Q. 1a) and change in teacher SCK over time (R.Q. 2a). The level-1 model estimated the within-teacher change over time (intraindividual), while the level-2 model estimated the between-teacher variation in change (interindividual). MLM can handle data that are unbalanced (that is, each person need not have the same number of data collection points) by using full information maximum likelihood estimation to compute parameter estimates based on the data collected for each person (Singer & Willett, 2003). Therefore, MLM allows the use of all participants in a longitudinal model, including those who participated in only one data collection time point.

The first step in the model building process was to estimate a two-level model (i.e., time nested within individual) that contained no independent variables in order to determine the intraclass correlation coefficient (ICC), which is an estimate of the variance in the outcome variable attributed to individual differences. The next step was to build the level-1 or growth model. Since treatment teachers displayed piecewise growth, the treatment group models were specified such that two separate linear growth factors were modeled with the first estimating the primary change (i.e., change from T0 to T1) and the second estimating the secondary change (i.e., average yearly change from T1 to T3). These growth factors were created by entering two time variables; TIME1 had the values of 0, 1, 1, 1, representing T0, T1, T2, and T3, respectively, and TIME2 had the values of 0, 0, 1, 2, representing T0, T1, T2, and T3, respectively. Control teachers exhibited linear growth; therefore, a single TIME1 variable, with the values 0, 1, 2, 3 to denote the four time points, was entered in the level-1 model. Separate models were necessary because two different growth patterns were modeled (i.e., piecewise growth for the treatment group and linear growth for the control group). A three-level model was not used because several of the schools had only one teacher involved in the study (i.e., departmentalized; see the details in Data Collection).

COURSES, YEARS, and DEGREE were added into the level-2 model as predictors of the intercept (R.Q. 1b). Then, the three predictors were added into the model as predictors of any slopes that had a statistically significant amount of variation (R.Q. 2b). COURSES and YEARS were grand mean centered so that the intercept and slope(s) would reflect the average across all teachers. DEGREE was entered into the model uncentered because it is a dichotomous variable.

The proportion of variance accounted for (PVAF) provides a measure of effect size for each of the predictors. The PVAF is computed by comparing a baseline model without predictors to a fitted model with all the predictors and by examining the change in variance. The PVAF was computed for all predictors in the models. The PVAF is often referred to as a pseudo-R2 (Raudenbush & Bryk, 2002; Singer & Willett, 2003). The R2 can be interpreted as values of less than .09 having a small effect size, between .09 and .25 having a medium effect size, and greater than .25 having a large effect size (Cohen, 1988). Therefore, the PVAF estimates were interpreted using the same criteria.

The resulting full level-1 and level-2 models are given by:

Level-1 model: $SCK_{ti} = \pi_{0i} + \pi_{1i}(TIME1_{ti}) + \pi_{2i}(TIME2_{ti})^* + e_{ti}$

Level-2 model: $\pi_{0i} = \beta_{00} + \beta_{01}(COURSES_i) + \beta_{02}(YEARS_i) + \beta_{03}(DEG_i) + r_{0i}$

$\pi_{1i} = \beta_{10} + \beta_{11}(COURSES_i) + \beta_{12}(YEARS_i) + \beta_{13}(DEG_i) + r_{1i}$

$\pi_{2i} = \beta_{20} + \beta_{21}(COURSES_i) + \beta_{22}(YEARS_i) + \beta_{23}(DEG_i) + r_{1i}$

* The TIME2 variable was only used for the treatment group, as described above.

Results

In answering the research questions for the study, the results are presented for TEST and SKS with regard to the treatment and control groups. The results for the unconditional models and full models are presented.

TEST Results

Treatment group

The unconditional model for the treatment group on TEST provided the level-1 variance estimate ($\sigma 2$) of 35.83 and the level-2 variance estimate ($\tau 00$) of 12.76. The ICC was computed as .26, indicating that 26% of the variance in TEST was due to between-teacher differences, and 74% of the variance was due to time. This result suggests that there was change in TEST over time.

Results from the full model are presented in Table 5. The full model indicated that across all treatment teachers, the average baseline TEST was 29.69 (R.Q. 1a). The mean increase in TEST during their first year from T0 to T1 was 3.04 points, which was statistically different from zero (R.Q. 2a). The average yearly change from T1 to T3 was not significantly different from zero. The time variables accounted for 63% of the within-teacher variation in TEST. It is noted that from here on, "significant" results indicate statistical significance.

COURSES, YEARS, and DEGREE were added as predictors of the intercept (R.Q. 1b). Only YEARS was a significant predictor of baseline TEST; for each additional year of teaching, TEST was expected to increase 0.20 points on average. For example, a teacher with 5 years teaching experience was expected to score one point higher on TEST than a teacher with 1 year teaching experience. The PVAF of YEARS was .05, indicating that YEARS accounted for 5% of the variation in baseline TEST across treatment teachers, which is a small effect. COURSES, YEARS, and DEGREE were not found to be significant predictors of primary (T0-T1) or secondary (T1-T3) change in TEST (R.Q. 2b).

Control group

The unconditional model for the control group on TEST provided the level-1 variance estimate ($\sigma 2$) of 22.37 and the level-2 variance estimate ($\tau 00$) of 25.30. The ICC was computed as .53, indicating that 53% of the variance in TEST was due to between-teacher differences, and 47% of the variance was due to time.

Results from the full model are presented in Table 6. The full model indicated that across all control teachers, the average baseline TEST was 28.83. The average yearly increase in TEST during their 3 years in the study was 1.36 points. The PVAF for the level-1 predictors was .23, indicating that the time variable accounted for 23% of the within-teacher variation in TEST.

COURSES, YEARS, and DEGREE were not significant predictors of teacher baseline status or of teacher change over time.

Table 5. *Results of Full Model for the Treatment Group.*

	TEST					SKS				
	Estimates of fixed effects									
Fixed effect	Coeff.	*SE*	*t*	*p*	PVAF	Coeff.	*SE*	*t*	*p*	PVAF
Intercept (β_{00})	29.69	1.04	28.61	< .001	--	2.78	0.08	32.74	< .001	--
COURSES (β_{01})	-0.02	0.16	-0.13	.895	.00	0.05	0.01	4.19	< .001	.15
YEARS (β_{02})	0.20	0.08	2.37	.019	.03	0.00	0.01	0.64	.520	.00
DEGREE (β_{03})	-1.49	1.44	-1.04	.303	.00	-0.16	0.12	-1.35	.181	.02
For TIME1 slope (π_{1j})										
Intercept (β_{10})	3.04	1.01	3.01	.003	--	0.42	0.08	5.52	< .001	--
COURSES (β_{11})	0.26	0.16	1.65	.101	.03	-0.03	0.01	-2.75	.007	.17
YEARS (β_{12})	-0.16	0.08	-1.95	.053	.08	0.00	0.01	-0.15	.880	.00
DEGREE (β_{13})	0.86	1.41	0.61	.539	.01	0.09	0.10	0.90	.371	.03
For TIME2 slope (π_{2j})										
Intercept (β_{20})	-1.13	0.82	-1.37	.173	--	0.08	0.03	2.53	.012	--
COURSES (β_{21})	-0.05	0.13	-0.38	.701	.00					
YEARS (β_{22})	0.06	0.07	0.89	.374	.01					
DEGREE (β_{23})	-0.07	1.15	-0.06	.955	.00					
	Estimates of variance components									
Random effect	*SD*	Variance	*df*	χ^2	*p*	*SD*	Variance	*df*	χ^2	*p*
Intercept (r_{0j})	7.13	50.88	73	380.13	< .001	0.56	0.31	118	367.57	< .001
TIME1 slope (r_{1j})	5.81	33.72	73	176.47	< .001	0.30	0.09	118	172.49	.001
TIME2 slope (r_{2j})	3.78	14.31	73	148.24	< .001	--	0[a]	--	--	--
Level 1 effect (e_{ij})	3.52	12.36	--	--	--	0.36	0.13	--	--	--

All fixed effects *df* = 159, except for the TIME2 slope, which had *df* = 412.

Note. [a]The variance was not significantly different from zero and thus was fixed to zero.

Table 6. *Results of Full Model for the Control Group.*

	TEST					SKS				
	Estimates of fixed effects									
Fixed effect	Coeff.	*SE*	*t*	*p*	PVAF	Coeff.	*SE*	*t*	*p*	PVAF
Intercept (β_{00})	28.83	0.87	33.27	< .001	--	2.77	0.09	31.62	< .001	--
COURSES (β_{01})	-0.11	0.13	-0.83	.406	.01	0.05	0.01	3.55	.001	.10
YEARS (β_{02})	0.10	0.07	1.46	.146	.01	0.00	0.01	-0.17	.868	.00
DEGREE (β_{03})	-0.20	1.23	-0.16	.873	.00	-0.01	0.12	-0.08	.936	.00
For TIME1 slope (π_{1j})										
Intercept (β_{10})	1.36	0.39	3.50	.001	--	0.10	0.05	1.93	.054	--
COURSES (β_{11})	0.04	0.06	0.64	.524	.05	-0.01	0.01	-0.87	.386	.00
YEARS (β_{12})	-0.03	0.03	-1.01	.316	.04	0.00	0.00	0.32	.749	.02
DEGREE (β_{13})	0.53	0.55	0.97	.334	.07	0.15	0.07	2.15	.032	.18
	Estimates of variance components									
Random effect	*SD*	Variance	*df*	χ^2	*p*	*SD*	Variance	*df*	χ^2	*p*
Intercept (r_{0j})	6.56	43.04	133	529.92	< .001	0.62	0.39	133	302.22	< .001
TIME1 slope (r_{1j})	1.82	3.32	133	190.39	.001	0.23	0.05	133	166.68	.025
Level 1 effect (e_{ij})	4.17	17.42	--	--	--	0.48	0.23	--	--	--

All fixed effects *df* = 166.

SKS Results

Treatment group

The unconditional model for the treatment group on SKS provided the level-1 variance estimate (σ2) of 0.24 and the level-2 variance estimate (τ00) of 0.18. The ICC was computed as .42, indicating that 42% of the variance in SKS was due to between-teacher differences, and 58% of the variance was due to time. Again, this result suggests that there was change in SKS over time.

Results from the full model are presented in Table 5. The full model indicated that across all treatment teachers, the average baseline SKS was 2.78 points. The mean increase in SKS during their first year in the study was 0.42 points. The mean increase in SKS during their second and third years in the study was 0.08 points. The PVAF for the level-1 predictors was .23, indicating that the time variables accounted for 23% of the within-teacher variation in SKS.

Of COURSES, YEARS, and DEGREE, only COURSES was a significant predictor of baseline SKS; for each additional college science course taken, baseline SKS was expected to increase 0.05 points on average. The PVAF of COURSES was .15, indicating that COURSES accounted for 15% of the variation in baseline SKS across treatment teachers, which is a medium effect.

Of COURSES, YEARS, and DEGREE, only COURSES was a significant predictor of primary (T0-T1) change in SKS; for each additional college science course taken, SKS was expected to increase 0.03 points less on average during the first year in the study. The PVAF of COURSES was .17, indicating that COURSES accounted for 17% of the variation in primary change on SKS across treatment teachers, which is a medium effect.

Control group

The unconditional model for the control group on SKS provided the level-1 variance estimate (σ2) of 0.33, and the level-2 variance estimate (τ00) of 0.33. The ICC was computed as .53, indicating that 53% of the variance in SKS was due to between-teacher differences, and 47% of the variance was due to time.

Results from the full model are presented in Table 6. The full model indicated that across all control teachers, the average baseline SKS was 2.77 points. The mean SKS did not change during the study. The PVAF for the level-1 predictors was .25, indicating that the time variable accounted for 25% of the within-teacher variation in SKS.

Of COURSES, YEARS, and DEGREE, only COURSES was a significant predictor of baseline SKS; for each additional college science course taken, SKS was expected to increase 0.05 points on average. The PVAF of COURSES was .10, indicating that COURSES accounted for 10% of the variation in baseline SKS across control teachers, which is a medium effect.

Of COURSES, YEARS, and DEGREE, only DEGREE was a significant predictor of change in SKS; teachers with an advanced degree were expected to increase SKS by 0.15 points more than teachers without an advanced degree on average over the course of the study. The PVAF of DEGREE was .18, indicating that DEGREE accounted for 18% of the variation in change on SKS across control teachers, which is a medium effect.

Discussion and Implications

This study examined the effect of a multiyear PD on teacher SCK between a treatment group and a control group. The study used two measures of SCK: a paper-based test and a questionnaire scale.

Discussion

We examined the status of teacher SCK at the start of the intervention (i.e., baseline status, R.Q. 1a). The treatment and control groups started off with similar SCK at the fifth grade level as measured by TEST: The teachers in the treatment group, on average, scored 78% correct (mean of 29.69 out of 38 maximum points), and the teachers in the control group scored 76% correct (mean of 28.83). These scores are similar to the 82% correct on the baseline test for inservice elementary school teachers in Nowicki et al. (2013). In addition, teachers in both treatment and control groups felt generally knowledgeable about science content at the fifth grade level, with a mean baseline SKS score of 2.78 on a 4-point scale in the treatment group and a mean of 2.77 in the control group.

We also examined whether teacher SCK was predicted by teacher background variables at the start of the intervention (R.Q. 1b). YEARS (i.e., total number of years teaching) was not a predictor of baseline TEST in the control group, but it was a predictor in the treatment group. While there was a difference in significance found across the two groups, the effect size was small in the treatment group. In both the treatment and control groups, COURSES (i.e., number of college science courses taken) was a significant predictor of SKS with additional courses leading to increased baseline SKS. This finding suggests that taking science courses in college increases self-perceived SCK, which is consistent with the literature (Diamond et al., 2013).

This study included an examination of whether teacher SCK changed over the intervention (R.Q. 2a). In the treatment group, the overall pattern of teacher change on both measures was growth during the first year of the PD, followed by relative stability during the subsequent 2 years while the intervention continued. The primary change suggests that the PD was successful at quickly improving both TEST and SKS in treatment teachers. This result is consistent with the pattern of short-term growth and long-term stability found in studies of other PDs' effects on teacher outcomes (Lee & Maerten-Rivera, 2012; Supovitz et al., 2000). While these previous PD studies did not involve a control group, the control group teachers in our study demonstrated growth in TEST over time, which may seem surprising. However, the finding that both groups improved TEST over time supports the idea that teachers learn science while teaching it and continue to gain SCK each year, which is consistent with the finding by Arzi and White (2008) that the required curriculum was a major source of teacher SCK. Since both the treatment and control teachers were teaching from a curriculum based on state science standards, both groups of teachers were expected to increase their SCK over time. One difference was that the treatment group teachers gained knowledge during the first year of the intervention, while the control group teachers gained the same level of knowledge over the 3-year period. This finding can be useful for districts trying to improve teacher SCK quickly, especially in urban settings with high rates of teacher mobility.

It should also be noted that the treatment teachers' SKS continued to rise during the second and third years of the study, whereas the control teachers' change in self-reported SKS was not significant at any point of the study, so that the treatment teachers felt considerably more knowledgeable than the control teachers at the end of the study. This finding demonstrates that the treatment teachers believed that the PD helped their SCK. This improvement in self-reported SCK might lead to enhanced self-confidence in teaching science, while low self-confidence in SCK has been cited as a deterrent to elementary science teaching (Appleton, 2008; Heller et al., 2012).

In this study, we examined whether change in teacher SCK was predicted by teacher background variables over the 3-year period of the intervention (R.Q. 2b). In the treatment group, YEARS was a significant predictor of the primary change slope for TEST with each additional year of teaching decreasing the change estimate by 0.16 points; this means that teacher SCK was expected to change less for each year of teaching. Thus, the intervention had greater effects for teachers with fewer years teaching. In the treatment group, COURSES was a significant predictor of the primary change slope for SKS with each additional college science course decreasing the change estimate by 0.03 points; this means that teacher SCK was expected to change less for each college science course taken. Thus, the intervention had greater effects for teachers with fewer college science courses. This finding could be interpreted that since the teachers with more courses began the study with higher SKS, there was less room for improvement. In the control group, neither YEARS nor COURSES was a significant predictor of change in SKS, which was expected, because SKS did not show a significant change in the control group.

Figure 2 graphically depicts the expected change in TEST for both the treatment and control groups based on the full model results. Figure 3 graphically depicts the expected change in SKS for both the treatment and control groups based on the full model results. In both the treatment and control groups, the higher the baseline scores were, the less room there was for improvement, so less improvement was seen in teachers with high baseline scores. Therefore, these correlations help explain why the intervention had a greater effect over time on treatment teachers with fewer years of teaching or fewer science courses for TEST and SKS, respectively.

Implications

Contributions to the literature

This study makes important contributions to the literature in several ways. First, while few studies have measured the SCK of inservice teachers with a large sample size, this study measured the SCK of a large sample of inservice fifth grade teachers. Second, while even fewer of the available studies on SCK examined the data longitudinally, this study examined the change in teacher SCK over time. Third, random assignment of schools into the treatment and control groups allowed the test of the causal effect of the intervention on teacher SCK. Fourth, the sample of teachers from randomly selected schools in a large school district allows generalizability of the results to the district in this study and other similar districts. Finally, the study used different measures of SCK, including the paper-based test (a more objective measure) and the questionnaire scale (a more subjective measure). Overall, the study could serve as a basis for future research on scale-up interventions. In addition, the study generated an available data set with which future research could compare teacher SCK of other samples.

Implications for PD

This study adds to the literature examining the benefits of PD for teacher SCK. While teachers develop SCK over time, PD including science content significantly increases the rate at which teachers develop SCK. Considering the high rate of teacher mobility between grades, between schools, and between careers, an efficient and effective PD opportunity is likely to prove beneficial to students. Teacher mobility is particularly high in urban settings, such as the school district in this study, where students are often underserved in science classrooms. The teacher mobility issue is especially important in light of the fact that teacher SCK is predictive of student science achievement (Diamond et al., 2014). While it seems that teachers will eventually learn science content, at least three cohorts of students could benefit from teachers with higher SCK as the result of content-based PD. Additionally, the limitations of our instruments (described below) may have caused underestimation of the full potential of teacher SCK growth from sustained PD.

Directions for future research

This study has several limitations that may be addressed in future research. One limitation is a ceiling effect on TEST. On the pretest, five teachers, including three from the treatment group and two from the control group (1.4% of the sample), earned perfect scores, and 78, including 37 from the treatment group and 41 from the control group (22.0% of the sample), answered at least 89% of the items correctly, leaving little room for improvement. At T1, while four treatment teachers and six control teachers earned perfect scores, more treatment teachers (74) than control teachers (55) answered at least 89% of the items correctly, leaving even less room for improvement in the treatment group. This ceiling effect may explain why the reliability of TEST scores decreased in the treatment group; the low reliability was probably caused by compression of the range of treatment teachers' scores as they approached the maximum possible points. This ceiling effect may also explain why the treatment teachers' TEST scores leveled off after the first year, although they might have continued to increase their SCK in the 2 subsequent years.

TEST performed as expected because it measured the ability of teachers to answer fifth grade level items, so high scores are appropriate. Future research could add higher level items to increase the discrimination of TEST and potentially decrease the ceiling effect. The ceiling effect was compounded by the fact that the same items were used for each TEST administration, which allowed both treatment and control teachers to learn the items and increase their test scores. Future research could use equated forms of TEST to reduce the memory effect of teachers recalling their answers to the previous form or discussing answers with others prior to retaking the test.

While the intervention seems to have improved the SCK of treatment teachers, it is difficult to distinguish how much of that improvement was caused by PD and how much was caused by teaching the intervention curriculum. It would therefore be useful to have three treatment groups, one receiving the PD only, one receiving the curriculum only, and one receiving both. However, this research design would require a substantially larger sample size of schools and teachers in the three treatment groups, in addition to the control group.

Finally, the information about how many years teaching fifth grade science would have been useful because it would have allowed us to examine the relationship between years teaching the science content at fifth grade, which was the grade level under examination in this study, and both baseline status and change, respectively.

References

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

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.

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can professional development make the vision of the standards a reality? The impact of the National Science Foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching, 44*(3), 375-395. doi:10.1002/tea.20145

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15. Retrieved from http://www.jstor.org/stable/3699979

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.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*(3), 3-14.

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*(1), 1-20. Retrieved from http://feraonline.org/fjer/2013/Diamond.51%281%29.pdf

Diamond, B. S., Maerten-Rivera, J., Rohrer, R.E., & Lee, O. (2014). Effectiveness of a curricular and professional development intervention at improving elementary teachers' science content knowledge and student achievement outcomes: Year 1 results. *Journal of Research in Science Teaching, 51*, 635-658. doi:10.1002/tea.21148

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

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. doi:10.3102/00028312038004915

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

Jarvis, T., Pell, A., & McKeon, F. (2003). Changes in primary teachers' science knowledge and understanding during a two year in-service programme. *Research in Science & Technological Education, 21*(1), 17-42. doi:10.1080/02635140308341

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

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

Lee, O. & Maerten-Rivera, J. (2012). Teacher change in elementary science instruction with English language learners: Multiyear professional development intervention across multiple grades. *Teachers College Record, 114*(8), 1-42.

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

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921-958. Retrieved from http://search.proquest.com/docview/61927580?accountid=14585

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.

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., Berry, A., & Meirink, J. (2014). Research on science teacher knowledge. In N.G. Lederman, & S. K. Abell (Eds.), *Handbook of research on science education* (Vol. II, pp. 848-870). New York, NY: Routledge.

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.

Corresponding Author:
Brandon Diamond
David Posnack Jewish Day School
5810 South Pine Island Road
Davie, FL 33328
E-mail: dr.bdiamond@gmail.com