2013-03-22

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

Brandon S. Diamond
University of Miami, brandonsdiamond@gmail.com

Recommended Citation

Diamond, Brandon S., "Effect of a Curricular and Professional Development Intervention on Elementary Teachers' Science Content Knowledge" (2013). *Open Access Dissertations*. 973.
https://scholarlyrepository.miami.edu/oa_dissertations/973

UNIVERSITY OF MIAMI

EFFECT OF A CURRICULAR AND PROFESSIONAL DEVELOPMENT INTERVENTION ON ELEMENTARY TEACHERS' SCIENCE CONTENT KNOWLEDGE

By

Brandon S. Diamond

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

May 2013

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

EFFECT OF A CURRICULAR AND PROFESSIONAL DEVELOPMENT INTERVENTION ON ELEMENTARY TEACHERS' SCIENCE CONTENT KNOWLEDGE

Brandon S. Diamond

Approved:

Okhee Lee, Ph.D. Professor of Teaching and Learning	M. Brian Blake, Ph.D. Dean of the Graduate School
Jennifer M. Langer-Osuna, Ph.D. Professor of Teaching and Learning	Ji Shen, Ph.D. Professor of Teaching and Learning
David L. Wilson, Ph.D. Professor of Biology	Jaime Maerten-Rivera, Ph.D. Senior Research Associate

DIAMOND, BRANDON S. (Ph.D., Teaching and Learning)

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

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Okhee Lee.
No. of pages in text. (75)

Teacher science content knowledge (SCK) is an important but rarely studied construct. The study examined two research questions using two measures of SCK (i.e., knowledge test and questionnaire) in order to explore changes in SCK over time and to evaluate the effectiveness of the intervention at improving SCK as compared to the control group of teachers who did not receive the intervention: (1) Did participation in two years of the study have additional benefits over one year of participation? and (2) Did teacher background variables and treatment group (treatment vs. control) predict teacher SCK at the start of the intervention (i.e., initial status) or over the course of the intervention (i.e., change)? The teacher knowledge test was comprised of 33 multiple choice questions and 5 short response questions, all written at approximately the fifth grade level. The Science Knowledge Scale portion of the questionnaire determined how teachers felt about their knowledge of nature of science, physical science, life science, and earth/space science. Longitudinal multilevel modeling was used to examine the change in teacher SCK over two years. Additionally, an intervention including a fifth grade science curriculum and professional development was studied to determine its effect on teacher SCK as measured by a science knowledge test and a questionnaire during the same two-year period. Each year of participation in the study significantly increased test scores, while only the first year significantly increased self-reported science knowledge. The number of science courses taken in college was a significant predictor of self-reported science knowledge. The intervention had a significant effect on the treatment group teachers' questionnaire responses and the change in those responses over time compared to the control group. The implications of these findings for research and practice are discussed.

DEDICATION

This dissertation is dedicated to my wife and children, whose patience and support allowed me to complete this work.

TABLE OF CONTENTS

Page

Chapter

Appendix

LIST OF FIGURES

LIST OF TABLES

Table Page

LIST OF ABBREVIATIONS

CK: Content Knowledge

ELL: English Language Learner

ESE: Exceptional Student Education

ESOL: English to Speakers of Other Languages

GPA: Grade Point Average

MAR: Missing at Random

MCAR: Missing Completely at Random

MCK: Mathematical Content Knowledge

MLM: Multilevel modeling

NMAR: Not Missing at Random

PCK: Pedagogical Content Knowledge

PD: Professional Development

PK: Pedagogical Knowledge

PVAF: Proportion of Variance Accounted For

PVUAF: Proportion of Variance Uniquely Accounted For

SCK: Science Content Knowledge

SKS: Science Knowledge Scale

T1: Time 1

T2: Time 2

CHAPTER 1

Introduction

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 of the research looks at how to improve teacher knowledge of science content (Fleer, 2009; Heller, Daeler, Wong, Shinohara, & Miratrix, 2012; Shallcross, Spink, Stephenson, & Warwick, 2002), how to measure it, or how to assess its impact on teaching practices or student achievement (Chinnappan & Lawson, 2005). In fact, on the rare occasions that teachers' science content knowledge (SCK) is addressed, it is usually studied in pre-service teachers rather than those currently teaching (Arzi & White, 2008; Ball, Lubienski, & Mewborn, 2001). There have also been few studies that focus on the content of professional development (PD), and whether and how much content knowledge (CK) increases during PD opportunities (Garet, Porter, Desimone, Birman, & Yoon, 2001).

The purpose of this study is to examine SCK of elementary school teachers using two measures over the first two years of a three-year curricular and PD intervention. The intervention was created as a collaboration between a university and a large urban school district to implement a new year-long fifth grade science curriculum intended to maximize inquiry-based learning and understanding of science concepts by all students, especially English language learners (ELLs). The study examined two research questions using two measures of SCK (i.e., knowledge test and questionnaire) in order to explore changes in SCK over time and to evaluate the effectiveness of the intervention at improving SCK as compared to the control group of teachers who did not receive the intervention:

1. Did participation in two years of the study have additional benefits over one year of participation?
2. Did teacher background variables and treatment group (treatment vs. control) predict teacher knowledge at the start of the intervention (i.e., initial status) or over the course of the intervention (i.e., change)?

The study could offer insights about effective interventions to improve teacher SCK. Considering that science testing at fifth grade counts toward school accountability in the state in which this study takes place, the results could also offer insights about SCK in the context of accountability policies in elementary science education.

CHAPTER 2

Literature Review

There is very little research available on practicing teachers' SCK (Arzi & White, 2008). 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 of science concepts have been identified as a main factor responsible for the differences in the quality of elementary science teaching (Shallcross et al., 2002).

Mathematics CK has been studied more extensively than SCK. Teachers' 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). Inquiry is important because it is an effective method for allowing students to construct scientific concepts (Jarvis, Pell, & McKeon, 2003). 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.

PD is important to improving the quality of U.S. schools, the effectiveness of policy for teachers and teaching practice, and 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.

This section addresses three issues: (1) effect of PD on teacher SCK; (2) the advantages and challenges associated with longitudinal research design; and (3) measures of SCK. See Appendix A for a more detailed review of the literature regarding teacher CK.

Figure 1: Model for Impact of PD on Teacher SCK and Student Achievement

Effect of Professional Development on Teacher Content Knowledge

The effects of PD on various teacher constructs, such as pedagogical knowledge, pedagogical content knowledge, 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 looked at three structural features: the form of the activity, 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 which integrate SCK and pedagogy have been found to increase teachers' confidence in teaching science (Shallcross et al., 2002). For example, 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 are 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.

PD that includes science content has also 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 SCK (Heller et al., 2012). Heller et al. 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 an estimated test score gain for teachers of 22% in the treatment group, compared to 2% in the control group, with an effect size of about 1.8 standard deviations. In addition, the treatment group students showed an improvement of about 7% over the control group students. The results suggest that PD improving teacher content knowledge also improves student content knowledge.

Longitudinal Research Design and Missing Data

A longitudinal research design is one in which inferences are based on data on the same subjects collected at different points in time, usually for the purpose of studying changes or continuity in the sample's characteristics (Arzi, 2004; Gall, Gall, & Borg, 2007). Longitudinal studies are scarce in education broadly and science education specifically due to various difficulties (Arzi, 2004). The most obvious challenges are sample attrition and continued funding (Arzi, 1988; Arzi & White, 2008). Researchers also have to find the balance between methodological continuity and the validity of repeated measurements.

Arzi and White (2008) studied teacher SCK, measured by interviews, in a 17-year longitudinal study. The teachers began the study as they entered a pre-service education program and were interviewed at five time points: at the beginning of the program, at the end of the program, at the end of the first year of teaching, at the end of the second year of teaching, and 16 or 17 years after they began teaching. The study began with 33 teachers, but the sample was down to 24 by the first year of data collection and 22 at the final data collection, for an attrition rate of 33%. The researchers were careful to make the teachers feel as if the teachers were *not* being tested as the researchers tried to measure SCK. The major findings of this study inform the importance of the current study:

- Change in content knowledge is multifaceted, including forgetting of unused knowledge and limited accretion of new material, along with improved understanding, structure reorganization and integration [of concepts].
- Development is facilitated by "critical mass" of content knowledge and interest and hence more likely to occur when teachers teach within their chosen areas of study.
- The required school curriculum is the single most powerful factor affecting teacher content knowledge, serving as both knowledge organizer and source.

The preceding findings demonstrate the value of studying teacher SCK change in a longitudinal manner, as well as an advantage of pairing PD with a mandatory curriculum change.

Sample attrition is a major concern when conducting a longitudinal study. Along with research subjects failing to answer all questions on an instrument, attrition leads to missing data. “Missing data are one of the most common analytic problems in the behavioral sciences” (Enders, 2011, p.267). There are three general classes of missing data: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). MCAR occurs when the likelihood of data being missing for a particular variable is unrelated to other measured variables and to the would-be values of that variable. MAR occurs “when the probability of missing data on a variable is related to other variables, but not to the would-be values of the incomplete variable…. [NMAR] occurs when the probability of missing data on a variable is related to the would-be value of that variable (i.e., outcome-dependent missingness)” (Enders, 2011, p. 269).

Ad hoc missing data methods, based on MCAR assumptions, are very common in the social sciences literature, especially deletion methods, which are used by default in statistical packages such as SPSS. However, computer models suggest that MCAR ignores bias that appears if missingness is MAR or NMAR, because sometimes the missing data is related to the outcome. The MAR-based analysis method maximum likelihood estimation uses all available data, including that from incomplete cases, to identify parameter estimates based on a normal distribution. NMAR methods vary widely, and there is no one model that is accepted as the best. One disadvantage of

NMAR methods is that they often rely on untestable assumptions. These considerations were important to keep in mind in the study, since missing data was an issue because the data is collected over two years.

Measures of Teachers' Content Knowledge

Teachers' CK has typically been assessed using self-reports and classroom observations. However, tests of the CK of practicing teachers are rare. This is in part due to the fact that teachers often consider questions of their CK threatening or disrespectful (Arzi & White, 2008). Each of the measures presents strengths and limitations.

Surveys or questionnaires 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.

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 certification process for about a century (D'Agostino & Powers, 2009). These tests vary by state, and over time have varied in which states require them. A meta-analysis showed that while the median correlation between test scores and college grade point average (GPA) was .55, the test scores showed a correlation with supervisor and principal ratings of only .05 for pre-service teachers and of .11 for in-service teachers. There was also a low correlation between college GPA and these supervisor ratings (.07), although the correlation was higher in the first year of teaching (.25) than with any other indicator. According to this analysis, it would seem that college GPA is a better indicator of teacher performance, but *only in the first year of teaching*.

In summary, teacher SCK is an understudied construct, in spite of the importance that has been attached to it in regards to quality of instruction. PD has been shown to improve teachers’ pedagogy and confidence, but its effect on SCK has rarely been addressed. When teacher CK is studied, the indirect method, such as surveys, is used much more often than more direct measures, such as tests. This study uses a test and a survey to measure teacher SCK and the effect of an intervention including PD on teacher SCK.

CHAPTER 3

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 is 65.3% Hispanic, 24.4% Black Non-Hispanic, 8.6% White Non-Hispanic, and 1.7% Other. For the same time period, the school district reports 72.3% of elementary students are eligible for free or reduced price lunch programs, 19.4% are designated as ELLs enrolled in English to Speakers of Other Languages (ESOL) programs, and 10.9% are 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 9 that participated in the previous project). Second, the 64 schools were randomly assigned to 32 experimental and 32 control schools. At the beginning of Year 1, all of the teachers in two of the 64 schools chose not to participate in the study, both of which had been assigned to the control group. The two schools were maintained in the control group, because we continued to collect student data and invite new teachers from those schools to participate in the study. In addition, one treatment school that participated in Year 1 of the study chose not to continue into Year 2.

Teacher and Student Participants

Table 1

Teacher Demographics during Years 1 and 2 (N = 293)

Variables	Demographic Groups	*N*	%
Gender	Male	46	15.2
	Female	247	81.8
Ethnicity	Hispanic	158	52.3
	Black Non-Hispanic	63	20.9
	White Non-Hispanic	58	19.2
	Haitian	6	2.0
	Asian	1	.3
	Other	7	2.4
Native Language(s)*	English	237	78.5
	Spanish	128	42.4
	Haitian Creole	10	3.3
	French	6	2.0
	Other	5	1.7
Other Fluent Language(s)*	English	56	18.5
	Spanish	35	11.6
	Haitian Creole	4	1.4
	French	2	.7
	Other	5	1.7
ESOL Training*	Bachelor’s or master’s degree in ESOL	27	9.0
	ESOL endorsement through college coursework	104	34.4
	ESOL endorsement through school district	142	47.0
	Grandfathered in through teaching	13	4.3
	No preparation for ESOL	29	9.6
Degrees	Bachelor’s	138	47.1
	Master’s	120	41.0
	Multiple Masters’	11	3.8
	Specialist	15	5.1
	Doctorate	2	.7
	Other	2	.7
	Missing	5	1.7

* Multiple categories could be selected.

Teacher participants. As a school-wide initiative, Year 1 of the study involved a total of 223 fifth grade teachers from 62 schools. Year 2 of the study involved a total of 197 fifth grade teachers, 70 of whom were new to the study, from 61 schools. The demographic information for all 293 participating teachers is shown in Table 1.

Student participants. The ethnic makeup of the student sample from both the treatment and control groups for Years 1 and 2 combined (2010-2012) was 66.4% Hispanic, 23.7% Black Non-Hispanic, 8.2% White Non-Hispanic, and 1.8% Other. In addition, 76.8% of elementary students in the sample were eligible for free or reduced price lunch programs, 15.4% were designated as ELLs enrolled in ESOL programs, and 10.7% were enrolled in ESE programs.

Curricular and Professional Development Intervention

While the intervention continues for three years, this study involves the first two years only. The intervention is comprised of (a) curriculum materials, including student books, teachers' guides, science supplies, and supplementary materials; (b) teacher PD workshops throughout the school year; and (c) school site support for curriculum implementation, including on-site PD. The intervention components were designed to complement and reinforce one another for the improvement of teachers' knowledge and practices in science instruction, based on state science content standards, along with English language development of ELLs in urban elementary schools.

Curriculum Materials

A comprehensive stand-alone science curriculum for grade 5 was developed from the 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. Due to a change in the state science content standards and based on feedback from treatment group teachers, curriculum materials were changed from Year 1 to Year 2 as outlined below.

Year 1. The curriculum was developed to address all of the benchmarks in the state science standards in place for the 2010-2011 school year. The curriculum was divided into strands A-H as dictated by the state science benchmarks and provided opportunities for students to explore key concepts in each (see Appendix B). It is partly through the in-depth exploration of these concepts that teachers in the treatment group are 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, develop their own understandings, and share their understanding with others. Suggestions to promote science inquiry and the 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 would 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 English language literacy development for students from diverse languages and cultures. Although the curriculum was designed to meet the needs of students who are learning English as a new language, it also worked well with native English speakers.

At the beginning of their participation in Year 1, the teachers in the treatment schools were provided with complete class sets of curriculum materials. The control schools implemented the curriculum adopted by the school district.

Year 2. The curriculum was revised to accommodate the new state science standards. The new standards organized topics into Big Ideas, with each standard in each Big Idea explaining how the standard would be assessed and what standards from previous grades would be assessed with it (Appendix C). The new standards were also more specific about what details would and would not be assessed at each grade level. In terms of the organization of the curriculum, the biggest change was the dispersal of the nature of science lessons throughout the book at the treatment group teachers' request. In the revised version, students spend much less time learning about inquiry, measurement, and graphing as a discrete unit, but spent more time developing each of these skills throughout the curriculum. While maintaining the inquiry-based nature of the curriculum, individual labs were modified in response to teacher feedback in the revision.

At the beginning of Year 2, each teacher received a new set of the revised student books, revised teachers' guide, replenishments of consumable materials, and a CD-ROM containing all the curriculum materials including supplementary resources.

Teacher Workshops

As a school-wide initiative, all fifth grade teachers in the treatment schools were invited to attend five days of PD workshops throughout each of Years 1 and 2, for a total of 10 days over the two-year period.

Year 1. The first three 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. Each year, teachers received stipends for attending the summer workshop, and schools received payments for substitute teachers during the school year.

The focus of the Year 1 workshops was on familiarizing the teachers with the science content and hands-on activities, as well as the state science content standards and assessment. Teacher workshops concentrated on CK and how students learn that content, along with opportunities for teachers to engage in active learning, both of which have been found to have positive effects on teachers' self-reported enhanced knowledge (Garet, Porter, Desimone, Birman, & Yoon, 2001). The majority of the summer and winter teacher workshops consisted of providing teachers with the opportunity to experience the inquiry-based labs of the curriculum as students, a PD method that has been found to improve teachers' comfort with science content more than learning it didactically (Cox & Carpenter, 1989). While most labs were carried out during the workshops, 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.

Year 2. During Year 2 implementation, a majority of the treatment teachers attended another three-day summer workshop. Teachers new to the intervention were given separate workshops from returning teachers, in order to maximize the learning opportunities for each group. The Year 2 January and May workshops combined all teachers into one group.

The returning teachers' summer workshop was primarily concerned with discussing the changes made to the curriculum and to the state science standards and assessment overall, as described in the curriculum materials section above. The teachers new to the intervention participated in a workshop similar to that of Year 1. The Year 2 January workshop introduced teachers to supplementary resources available to them, along with teaching them how to effectively assess student learning and understanding formatively and plan for the state assessment. The Year 2 year-end workshop in May 2012 was similar to that of Year 1, but also focused on unpacking science inquiry, highlighting opportunities for more open-ended questioning in the classroom, providing activities to help teachers become more comfortable 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 inquiry/open-ended questioning.

School Site Support

Each year, three members of the research team were each assigned to a group of schools and on average visited each school every 4 to 6 weeks, for a total of 4-6 times. The school site support team paid particular attention to high-need schools and teachers on the basis of classroom visits or in response to requests from teachers or school administrators. Assistance was offered for a range of activities, including planning lessons, co-teaching, suggestions for additional materials (e.g., home learning, lab management, supplementary assessment materials), delivery of supplies, and on-site PD.

While conducting the visits with all of the teachers, the research team members also met with school administrators to address concerns.

Data Collection, Instruments, and Data Analysis

Data Collection

Teachers in both groups had the confidentiality of their responses explained to them prior to beginning any data collection, and any concerns on the teachers' parts were addressed. Data collection activities were performed during the teacher workshops for the treatment condition and in the individual schools for the control condition during Years 1 and 2.

Year 1. In the treatment group, at the beginning of the first day of the workshop, teachers were provided with the informed consent and background information form. They were then asked to complete the questionnaire and science knowledge test. Data collection was conducted by the trained staff facilitating the workshop in a controlled environment without a time limit. At the year-end workshop, the questionnaire and science knowledge test were administered again to all teachers.

Teachers from control schools were given the same informed consent, background information form, questionnaire, and science knowledge test during school-site visits in the first quarter of each school year. Data collection was conducted by the trained staff at the school sites. 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.

Year 2. Teachers new to the schools participating in the study were administered the same informed consent, background information form, questionnaire, and science knowledge test in the same manner the instruments were administered 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.

Time point assignment. The baseline measure for each teacher is from the first instrument administration that the teacher participated in before beginning the study (see Tables 2 and 3). For teachers who participated in Year 1, that was usually the test and questionnaire from the beginning of Year 1. For teachers who did not begin participation until Year 2, the baseline was usually measured by the instruments administered at the beginning of Year 2. Time 1 (T1) was the first data collection after the baseline, which is the end of Year 1 for most teachers and the end of Year 2 for teachers new to the study in Year 2. Time 2 (T2) was the end of Year 2 for teachers who participated for two years, and was treated as missing data for teachers who participated in the study for only one year. Baseline was treated as missing data for teachers who did not take a pretest or prequestionnaire (e.g., teacher entered mid-way in the school year).

Table 2

Patterns of Test Data Collected (N =279)

	Treatment N	*Control N*	Baseline	Time 1	Time 2
Group 1	66	52	X	X	X
Group 2	46	57	X	X	
Group 3	11	9	X		
Group 4	1	1	X		X
Group 5	25	8		X	
Group 6	3	4		X	X
Totals	152	131	243	261	127

Table 3

Patterns of Questionnaire Data Collected (N =280)

	Treatment N	*Control N*	Baseline	Time 1	Time 2
Group 1	68	52	X	X	X
Group 2	46	56	X	X	
Group 3	10	8	X		
Group 4	1	1	X		X
Group 5	24	10		X	
Group 6	4	5		X	X
Totals	153	132	242	260	126

Instruments

Teacher science knowledge test. The teacher science knowledge test (TEST) consisted of 24 multiple choice and 6 short response items, including 24 items taken from Trends in International Mathematics and Science Study (TIMSS) and National Assessment of Educational Programs (NAEP) and 6 items that were project developed. The NAEP and TIMSS items selected were primarily administered to 4th grade and were items that were of difficult content and high cognitive complexity, though some items were administered to 8th grade. The items were selected to reflect state science content standards at the grade level the teachers would be teaching science, including life sciences, physical sciences, earth sciences, and nature of science. Due to the need for item security, the test will not be released until the completion of the study.

One point was awarded for each correct answer for the multiple choice items. For the short response questions, teachers were awarded the number of points, 0-3, as recommended by the original source of each question, with partial credit usually available. There were 38 total points possible. The TEST mean for the treatment and control group were calculated for each time point. Table 4 presents the descriptive statistics and reliability for each time point. 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). TEST baseline was shown to have a Cronbach's alpha of .75 for treatment and .79 for control, which indicate acceptable reliability. TEST T1 had an alpha of .61 for treatment and .79 for control, which indicate acceptable reliability for T1 overall. TEST T2 was shown to have a Cronbach's alpha of .57 for treatment and .80 for control. The T2 alpha is somewhat low, but it is expected that TEST was less reliable for the treatment group than for the control group at both T1 and T2, because it shows that the treatment group teachers were becoming more similar to each other as they learned the science (R.D. Penfield, personal communication, 2013). The items with low discrimination for the treatment group were all of low difficulty (at least 90% of teachers answered correctly), and there was a similar number of non-discriminating items in the treatment TEST compared to the control TEST.

Table 4
Descriptive Statistics of Teacher-level Variables

Measure	Treatment				Control			
	N	*M*	*SD*	*α*	*N*	*M*	*SD*	*α*
TEST Baseline	124	30.59	4.95	.75	119	30.36	5.46	.79
TEST T1	140	32.66	3.37	.61	121	31.11	4.86	.79
TEST T2	70	33.57	2.98	.57	57	32.26	5.09	.80
SKS Baseline	125	2.73	0.69	.94	117	2.89	0.64	.92
SKS T1	142	3.19	0.53	.88	123	3.15	0.57	.92
SKS T2	73	3.19	0.63	.94	58	3.19	0.47	.87
YEARS	149	13.60	7.73	--	131	14.86	9.37	--
DEGREE	155	0.52	0.50	--	133	0.53	0.50	--
COURSES	149	4.62	4.28	--	126	4.83	4.73	--

Teachers' self-reported science knowledge. The science knowledge scale (SKS) was a section of the questionnaire that consisted of four Likert-type questions asking the teachers to indicate how knowledgeable they feel about teaching the nature of science, physical science, earth/space science, and life science at their grade level (Appendix D). The questionnaire items used a 4-point rating system, with 1 = not knowledgeable and 4 = very knowledgeable. 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. Teachers who answered at least 75% of the items from the scale were included. The SKS means for the treatment and control group were calculated for each time point. Table 4 presents the descriptive statistics and reliability for each time point. SKS baseline was shown to have a Cronbach's alpha of .94 for treatment and .92 for control which indicate strong reliability. SKS T1 had an alpha of .88 for treatment and .92 for control, which indicate strong reliability. SKS T2 was shown to have a Cronbach's alpha of .94 for treatment and .87, which indicate strong reliability.

Teacher background. The background information form provided data on three of the predictor variables used in the analyses. A previous version of this intervention found that college science courses taken was a significant predictor of increase in SKS (Lee & Maerten-Rivera, 2012). Supovitz, Mayer, and Kahle (2000) found that teachers with more teaching experience tended to feel less prepared to use inquiry as a method of teaching science. In our intervention, inquiry is the primary vehicle for both teaching practices and teacher learning. Therefore, number of years teaching experience was included in the analyses (YEARS) as a predictor of baseline and change parameters. The teacher's highest degree obtained was also included (DEGREE); 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. The teacher's number of science courses taken in college, including science methods courses, was entered as the total number (COURSES).

Descriptive statistics for YEARS, DEGREE, and COURSES are displayed in Table 4. In addition, group was included as a variable (GROUP) with teachers in the treatment group coded as 1, and teachers in the control group coded as 0.

Data Analysis

Data trends. Before longitudinal analysis of the data began, the trends demonstrated by the means of each outcome variable were examined. The means for each time on both outcome variables are displayed in Table 4. The mean baseline TEST score was 30.59 out of 38 (80.5%) for treatment teachers and 30.36 (79.9%) for control teachers. This shows that the teachers began reasonably knowledgeable about fifth grade science content, but not extremely so, and that the two groups began with similar SCK. The mean T1 TEST score was 32.66 (85.9%) for treatment teachers and 31.11 (81.9%) for control teachers, and the mean T2 TEST score was 33.57 (88.3%) for treatment teachers and 32.26 (84.9%) for control teachers. Overall, there was more growth by treatment teachers than control teachers over the two-year period. In both groups there was more growth between baseline and T1 than between T1 and T2.

The mean baseline SKS score was 2.73 out of 4 for treatment teachers and 2.89 for control teachers. This shows that teachers in both groups began the study feeling generally knowledgeable about fifth grade science content, but not extremely so. The mean T1 SKS score was 3.19 for treatment teachers and 3.15 for control teachers, and the mean T2 SKS score was 3.19 for treatment teachers and 3.19 for control teachers. The overall means demonstrate most of the growth in both groups occurred from baseline to T1.

The intercorrelations among teacher-level variables are displayed in Table 5 and demonstrate that while the predictor variables were significantly correlated with each other, none were highly correlated.

Table 5

Intercorrelations Among Teacher-Level Variables

	DEGREE	COURSES
YEARS	.20**	.24**
DEGREE	--	.16**

**p < .01

Multilevel modeling. Longitudinal multilevel modeling (MLM) using HLM 7 was conducted with time nested within teachers to examine change in teacher SCK, including the examination of change that occurred in the second year of the intervention, and if treatment teachers changed more than control teachers (Raudenbush & Bryk, 2002). A level-1 or within-individual model examines how participants change over time. A level-2 or between-individual model examines what predicts differences in both baseline and change among participants.

Using multilevel modeling to examine change has advantages over more traditional methods (e.g., ANOVA, regression). First, in multilevel modeling, both initial status on the variable of interest and rate of change on the variable of interest are estimated free from measurement error variance. In contrast, more traditional methods analyze only mean change while ignoring differences in initial status and treating differences in change as error variance. Second, multilevel modeling can handle data that are unbalanced; that is, each person need not have the same number of data collection points. Rather, the model uses full information maximum likelihood estimation to compute parameter estimates based on the data collected for each person; these estimates are unbiased, provided that attrition is random (Singer & Willett, 2003). Therefore, the use of multilevel modeling allows us to use all participants in a longitudinal model, including those who only participated in one data collection time point.

The same model building process was used for each outcome variable (TEST and SKS). Before any models were created, the scores of individual teachers were graphed to determine if both outcome variables demonstrated linear, quadratic, or piecewise growth. Linear graphs would have suggested that teachers improve their SCK at a constant rate. Instead, piecewise growth was shown by discontinuous graphs, which suggested that the rate of SCK change in a teacher's first year participating in the study was different from the rate of SCK change during the teacher's second year participating in the study. This type of change is consistent with results reported by Supovitz and Turner (2000) and the previous version of this intervention (Lee & Maerten-Rivera, 2012). Graphs for treatment and control group teachers were examined separately to determine if they followed similar growth patterns. The growth patterns were similar for the two groups, so one model was built for each SCK measure with GROUP as a predictor of the rate of SCK change. While it may be surprising that both the treatment and control group teachers demonstrated growth in SCK over time, studies have shown that teaching a topic increases the confidence in that topic (Shallcross, Spink, Stephenson, & Warwick, 2002), and it is reasonable that a similar relationship may exist between teaching science and SCK. Additionally, participants taking the same test or measure repeatedly often remember their answers, which can cause participants to show an improvement as an effect of their experience with previous administrations (Gall, Gall, & Borg, 2007).

Unconditional model. The first step in the model building process used to examine both outcomes 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[1]), which is an estimate of the variance in the outcome variable attributed to individual differences. In addition, the proportion of variance accounted for (PVAF) was calculated for models throughout. In multilevel modeling, PVAF can be 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 provides a measure of effect size for the overall model. The proportion of variance uniquely accounted for (PVUAF) by each of the background predictors was computed by comparing a baseline model without a predictor of interest to a 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. Because the proportion of variance estimates provide measures of effect size, they are 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 .09 as 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 and PVUAF estimates can be interpreted using the same criteria as the R^2.

---

[1] Congruent with Raudenbush and Bryk (2002), the interclass correlation coefficient was determined from the following formula:

$$\frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

where τ_{00} = the estimated level-2 variance for the model and σ^2 = the estimated level-1 variance for the model within the parentheses.

Level-1 model. The next step for both models was to build the level-1 model. Initial status for the TEST and SKS models estimated the TEST and SKS scores, respectively, at the baseline time point. Because teacher SCK showed piecewise growth, the models for TEST and SKS were specified such that two separate linear growth factors were modeled. The first growth factor examined the primary change after participating in the first year of the study (i.e., change from baseline to T1), while the second examined the secondary change after participating in the second year of the study (i.e., change from T1 to T2). These growth factors were created by entering two time variables. The first time variable, X1, had the values of 0, 1, 1, representing baseline, T1, and T2, respectively (i.e., change from baseline to T1). The second time variable, X2, had the values of 0, 0, 1, representing baseline, T1, and T2, respectively (i.e., change from T1 to T2). For each outcome variable, the two time variables, X1 and X2, were entered into the model. Due to the piecewise nature of the change over time and the presence of three time points rather than four, we allowed the primary slope estimate to vary and fixed the variance of the secondary slope estimate in order to allow for sufficient degrees of freedom in the model. The slopes of X1 and X2 were then compared to determine the relative abilities of the predictors to change the rates of TEST and SCK change, respectively.

Level-2 model. First, YEARS, DEGREE, COURSES, and GROUP were added into the models as predictors of the intercept. Next, the four predictors were added into the models as predictors of the slope. Additionally, the correlation estimates of the intercept parameter and the slope parameter were used to determine if the baseline SCK of the teacher affected the rate of SCK change. SCK in the model below refers to either TEST or SKS, depending on the outcome being examined.

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

Level-1 model: $\mathrm{SCK}_{ti} = \pi_{0i} + \pi_{1i}(\mathrm{X1}_{ti}) + \pi_{2i}(\mathrm{X2}_{ti}) + e_{ti}$

Level-2 models: $\pi_{0i} = \beta_{00} + \beta_{01}(\mathrm{DEG}_i) + \beta_{02}(\mathrm{COURSES}_i) + \beta_{03}(\mathrm{YEARS}_i) + \beta_{04}(\mathrm{GROUP}_i) + r_{0i}$

$\pi_{1i} = \beta_{10} + \beta_{11}(\mathrm{DEG}_i) + \beta_{12}(\mathrm{COURSES}_i) + \beta_{13}(\mathrm{YEARS}_i) + \beta_{04}(\mathrm{GROUP}_i) + r_{1i}$

$\pi_{2i} = \beta_{20}$

Note that these are the full models used. For both models, π_{2i} was fixed to zero as previously mentioned, thus the error term r_{2i} is not included in the equation above. YEARS and COURSES were grand mean centered so that the intercept would reflect the average across all teachers. DEGREE and GROUP were entered into the model uncentered since they are dichotomous variables. All models were estimated using full maximum likelihood.

CHAPTER 4

Results

Impact of Time and Intervention on TEST

Unconditional model. Results from the TEST unconditional model are displayed in Table 6. The level-1 variance (σ^2) for the unconditional model was 7.02, and the level-2 variance (τ_{00}) was 12.54. The ICC was computed as .64, suggesting that 64% of the variance in TEST scores was due to between-teacher differences, whereas 36% of the variance in TEST scores was due to time. The variance of the intercept (r_{0j}) was statistically significant ($\chi^2[262] = 1278.31, p < .001$), suggesting a statistically significant amount of variance in TEST scores between teachers. (From this point forward, the word "significant" will be used to denote "statistically significant.")

Table 6
Results of Unconditional Model for TEST

Random Effect	*SD*	Variance	*df*	χ^2	*p*
Intercept (β_{00})	3.54	12.54	262	1278.31	< .001
Level 1 effect (e_{ij})	2.65	7.02	--	--	--

Level-1 (Time) Model. Next, an unconditional piecewise change model in which only the time variables were entered as uncentered predictors was examined. The results of the level-1 model are presented in Table 7. The coefficient for initial TEST status was 30.67 ($t[262] = 95.12, p < .001$), indicating that across all teachers, the average starting science test score was 30.67. The coefficient for primary change in TEST was 1.53 ($t[262] = 6.32, p < .001$), indicating that across all teachers, TEST increased by 1.53 points, on average, from baseline to T1. The coefficient for secondary change in TEST was 0.96 ($t[262] = 3.94, p < .001$), indicating that across all teachers, TEST increased by 0.96 points, on average, from T1 to T2.

Table 7

Results of Level 1 Model for TEST

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*
Intercept (β_{00})	30.67	0.32	95.12	262	< .001
For X1 slope (π_{1j})					
Intercept (β_{10})	1.53	0.24	6.32	262	< .001
For X2 slope (π_{2j})					
Intercept (β_{20})	0.96	0.24	3.94	277	< .001

Estimates of Variance Components

Random Effect	*SD*	Variance	*df*	χ^2	*p*
Intercept (r_{0j})	4.62	21.33	206	1246.42	< .001
X1 Slope (r_{1j})	2.18	4.77	206	353.10	< .001
Level 1 effect (e_{ij})	1.98	3.93	--	--	--

The estimate of the variance component for the intercept was 21.33, which was significant, suggesting that there was a significant amount of between-teacher variation in TEST at baseline. The estimate of the variance component for the primary change slope was 4.77, which was significant, suggesting that there was a significant amount of between-teacher variation around the initial true change from baseline to T1 in TEST.

Level-2 (Teacher) Model. YEARS, DEGREE, COURSES, and GROUP were then added to the model as predictors of the intercept (see Table 8 for results). The PVAF in the intercept by the teacher-level predictors was .01, suggesting that the teacher predictors accounted for 1% of the variation in the average true TEST score across teachers at baseline. The coefficient for initial TEST status was 30.25 (*t*[258] = 64.24, *p* < .001), indicating that across all teachers, the average starting SCK science test score was 30.25. Of the four predictors entered, only GROUP had a coefficient that was a significant predictor of the initial status of TEST, 1.26 (*t*[258] = 2.76, *p* = .006), but it did not account for any of the variance (PVUAF = 0), which means its effect is negligible.

Table 8

Results of Intercept Predictor Model for TEST

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*	PVUAF
Intercept (β_{00})	30.25	0.47	64.24	258	< .001	--
YEARS (β_{01})	0.04	0.03	1.36	258	.177	.02
DEGREE (β_{02})	-0.48	0.47	-1.04	258	.301	.00
COURSES (β_{03})	-0.02	0.05	-0.29	258	.775	.00
GROUP (β_{04})	1.26	0.45	2.76	258	.006	.00
All Variables (PVAF)						.01
For X1 slope (π_{1j})						
Intercept (β_{10})	1.52	0.24	6.28	262	< .001	
For X2 slope (π_{2j})						
Intercept (β_{20})	0.94	0.24	3.89	277	< .001	
Random Effect	*SD*	Variance	*df*	χ^2	*p*	
Intercept (r_{0j})	4.60	21.14	202	1267.81	< .001	
X1 Slope (r_{1j})	2.20	4.84	206	356.18	< .001	
Level 1 effect (e_{ij})	1.97	3.89	--	--	--	

YEARS, DEGREE, COURSES, and GROUP were then added to the model as predictors of the slope (Table 9). This is the final model with all predictors included. The primary change (i.e., mean change in score from baseline to T1) is quantified by the coefficient β_{10}. The coefficient for primary change in TEST was 1.10 ($t[258] = 2.65$, $p = .009$), indicating that the mean increase in TEST for all teachers during their first year in the study was 1.10 out of 38 points. TEST had a significant mean primary change score. The PVAF by these teacher-level predictors was .07, suggesting that the predictors accounted for 7% of the variation in the average true change in TEST score across teachers from baseline to T1. TEST was then examined for predictors that had an effect on the primary change. There were no significant predictors of primary change found in the model. However, the coefficient of 0.78 for GROUP did approach statistical significance ($t[258] = 1.71$, $p = .088$). The coefficient of 0.78 indicates that the change from baseline to T1 for treatment teachers was about 0.78 points higher than that of control teachers.

Table 9

Results of Full Model for TEST

Estimates of Fixed Effects

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*	PVUAF
Intercept (β_{00})	30.67	0.58	52.77	258	< .001	--
YEARS (β_{01})	0.09	0.04	2.17	258	.031	.01
DEGREE (β_{02})	-0.57	0.66	-0.86	258	.393	.00
COURSES (β_{03})	0.00	0.07	0.06	258	.950	.00
GROUP (β_{04})	0.50	0.64	0.77	258	.441	.00
All Variables (PVAF)						.01
For X1 slope (π_{1j})						
Intercept (β_{10})	1.10	0.41	2.65	258	.009	--
YEARS (β_{11})	-0.05	0.03	-1.74	258	.084	.04
DEGREE (β_{12})	0.08	0.47	0.17	258	.863	.00
COURSES (β_{13})	-0.02	0.05	-0.34	258	.736	.00
GROUP (β_{14})	0.78	0.45	1.71	258	.088	.03
All Variables (PVAF)						.07
For X2 slope (π_{2j})						
Intercept (β_{20})	0.95	.24	3.91	277	< .001	

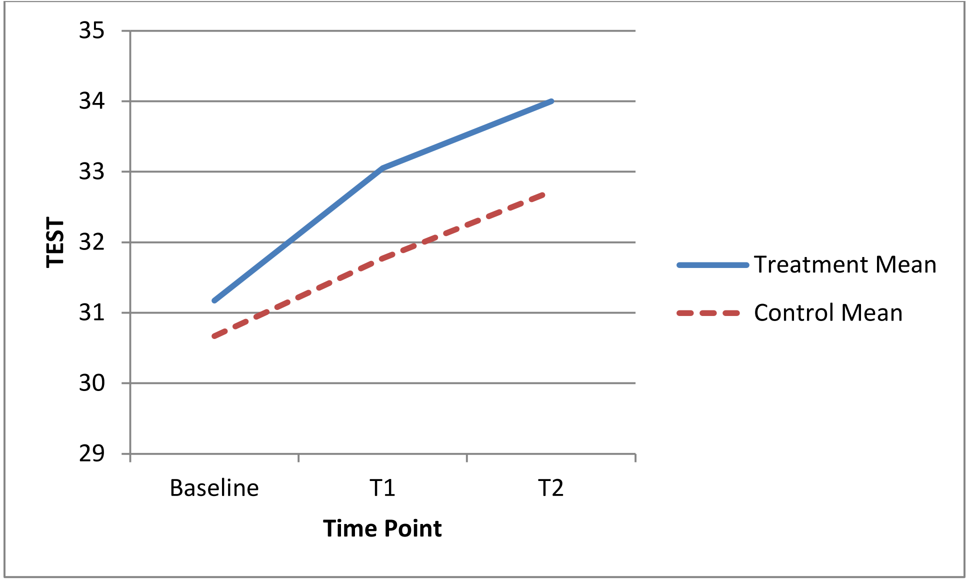
Estimates of Variance Components

Random Effect	*SD*	Variance	*df*	χ^2	*p*
Intercept (r_{0j})	4.59	21.10	202	1263.21	< .001
X1 Slope (r_{1j})	2.11	4.46	202	344.87	< .001
Level 1 effect (e_{ij})	1.97	3.90	--	--	--

The secondary change (i.e., mean change in score from T1 to T2) is quantified by the coefficient β_{20}. TEST had a significant mean secondary change score. The coefficient for secondary change in TEST was 0.95 ($t[277] = 3.91, p < .001$), which was statistically significant, indicating that across all teachers from T1 to T2, TEST increased, on average, by 0.95 points. The correlation between the intercept parameter and the slope parameter was -.79, which is a large correlation suggesting that the higher a teacher's initial TEST score was, the smaller the change was expected to be. Figure 2 graphically depicts the change in TEST based on the model results.

Figure 2

Change in TEST Over Time



Impact of Time and Intervention on SKS

Unconditional model. Results from the SKS unconditional model are displayed in Table 10. The level-1 variance (σ^2) for the unconditional model was 0.22, and the level-2 variance (τ_{00}) was 0.16. The ICC was computed as .43, suggesting that 43% of the variance in SKS was due to between-teacher differences, whereas 57% of the variance in SKS was due to time. The variance of the intercept (r_{0j}) was statistically significant (χ^2[264] = 704.27, $p < .001$), suggesting a significant amount of variance in self-reported science knowledge between teachers.

Table 10

Results of Unconditional Model for SKS

Random Effect	*SD*	Variance	*df*	χ^2	*p*
Intercept (β_{00})	0.40	0.16	264	704.27	< .001
Level 1 effect (e_{ij})	0.47	0.22	--	--	--

Level-1 (Time) Model. Next, an unconditional piecewise change model in which only the time variables were entered as uncentered predictors was examined. The results of the resulting level-1 model are presented in Table 11. The coefficient for initial SKS status was 2.79 ($t[264] = 64.79$, $p < .001$), indicating that across all teachers, the average starting SKS score was 2.79. The coefficient for primary change in SKS was 0.39 ($t[264] = 9.80$, $p < .001$), indicating that across all teachers, SKS increased by 0.39 points, on average, from baseline to T1. The coefficient for secondary change in SKS, -0.04 ($t[264] = -0.88$, $p = .380$), was not significant, indicating that across all teachers, SKS did not change, on average, from T1 to T2. The estimate of the variance component for the intercept was 0.29, which was significant, suggesting that there was a significant amount of between-teacher variation in SKS at baseline. The estimate of the variance component for the primary change slope was 0.04, which was significant, suggesting that there was a significant amount of between-teacher variation around the initial true change in SKS.

Table 11

Results of Level 1 Model for SKS

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*
Intercept (β_{00})	2.79	0.04	64.79	264	< .001
For X1 slope (π_{1j})					
Intercept (β_{10})	0.39	0.04	9.80	264	< .001
For X2 slope (π_{2j})					
Intercept (β_{20})	-0.04	0.05	-0.88	273	.380
Estimates of Variance Components					
Random Effect	*SD*	Variance	*df*	χ^2	*p*
Intercept (r_{0j})	0.54	0.29	208	614.66	< .001
X1 Slope (r_{1j})	0.20	0.04	208	248.75	.028
Level 1 effect (e_{ij})	0.39	0.15	--	--	--

Level-2 (Teacher) Model. YEARS, DEGREE, COURSES, and GROUP were then added to the model as predictors of the intercept (see Table 12 for results). The PVAF in the intercept by the teacher-level predictors was .12, suggesting that the teacher predictors accounted for 12% of the variation in the average true SKS score across teachers at baseline. The coefficient for initial SKS status was 2.80 ($t[260] = 45.44$, $p <$ .001), indicating that across all teachers, the average starting SKS was 2.80. Of the four predictors entered, only COURSES was a significant predictor of the initial status of SKS with a coefficient of 0.03 (*t*[260] = 4.14, $p < .001$), indicating that for each additional college science course taken, teachers' average SKS scores increase by 0.03 on average. The PVUAF of COURSES was .12, meaning that COURSES accounted for 12% of the between-teacher variance found in initial SKS status. GROUP was not a significant predictor of the initial status of SKS (*t*[260] = 0.04, $p = .970$), indicating that the treatment and control groups began the study with similar SKS scores.

Table 12

Results of Intercept Predictor Model for SKS

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*	PVUAF
Intercept (β_{00})	2.80	0.06	45.44	260	< .001	--
YEARS (β_{01})	0.00	0.00	0.18	260	.862	.00
DEGREE (β_{02})	-0.03	0.06	-0.43	260	.669	.00
COURSES (β_{03})	0.03	0.01	4.14	260	< .001	.12
GROUP (β_{04})	0.00	0.06	0.04	260	.970	.00
All Variables (PVAF)						.12
For X1 slope (π_{1j})						
Intercept (β_{10})	0.39	0.04	9.81	264	< .001	
For X2 slope (π_{2j})						
Intercept (β_{20})	-0.04	0.05	-0.93	273	.352	
Random Effect	*SD*	Variance	*df*	χ^2	*p*	
Intercept (r_{0j})	0.50	0.25	204	568.52	< .001	
X1 Slope (r_{1j})	0.20	0.04	208	248.71	.028	
Level 1 effect (e_{ij})	0.39	0.15	--	--	--	

YEARS, DEGREE, COURSES, and GROUP were then added to the model as predictors of primary change (see Table 13 for results). The coefficient for primary change in SKS was 0.25 (*t*[260] = 3.83, $p < .001$), indicating that the mean increase in SKS for all teachers from baseline to T1 was 0.25. SKS was then examined for predictors that had an effect on the primary change. COURSES was a significant predictor of primary change in SKS. The coefficient for COURSES was -0.02 (*t*[255] = -2.31, $p = .022$), indicating that across teachers, primary change in SKS is lower for each additional college science course taken. The slope decreases by about 0.02 points for each additional course taken; COURSES explained 26% of the observed variance in the primary change slope.

Table 13
Results of Full Model for SKS

Estimates of Fixed Effects

Fixed Effect	Coefficient	*SE*	*t*	*df*	*p*	PVUAF
Intercept (β_{00})	2.91	0.07	38.94	260	< .001	--
YEARS (β_{01})	0.00	0.01	-0.02	260	.988	.00
DEGREE (β_{02})	-0.08	0.08	-0.97	260	.333	.01
COURSES (β_{03})	0.04	0.01	4.65	260	< .001	.14
GROUP (β_{04})	-0.14	0.08	-1.77	260	.079	.02
All Variables (PVAF)						.15
For X1 slope (π_{1j})						
Intercept (β_{10})	0.25	0.07	3.83	260	< .001	--
YEARS (β_{11})	0.00	0.00	0.23	260	.820	.00
DEGREE (β_{12})	0.07	0.07	0.95	260	.345	.11
COURSES (β_{13})	-0.02	0.01	-2.31	260	.022	.26
GROUP (β_{14})	0.19	0.07	2.64	260	.009	.35
All Variables (PVAF)						.51
For X2 slope (π_{2j})						
Intercept (β_{20})	-0.05	0.05	-1.00	273	.319	

Estimates of Variance Components

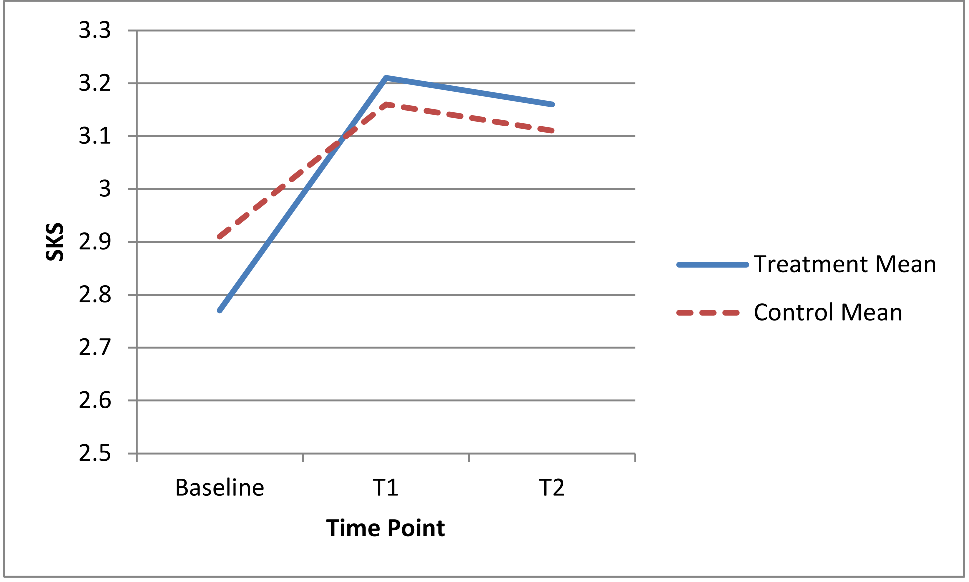
Random Effect	*SD*	Variance	*df*	χ^2	*p*
Intercept (r_{0j})	0.49	0.24	204	551.13	< .001
X1 Slope (r_{1j})	0.14	0.02	204	232.99	.080
Level 1 effect (e_{ij})	0.39	0.15	--	--	--

GROUP had a significant effect on primary change in SKS. The coefficient for GROUP was 0.19 ($t[260] = 2.64$, $p = .009$), indicating that teachers in the treatment group demonstrated additional change. The primary change slope of treatment group teachers increased by 0.19 points on average compared to control group teachers. GROUP explained 35% of the observed between-teacher variance in SKS change over the first year. The coefficient for secondary change in SKS was not significant ($t[273] = -1.00$, $p = .319$), indicating that teachers' SKS did not increase from T1 to T2.

Figure 3 graphically depicts the change in SKS based on the model results. The correlation between the intercept parameter and the slope parameter was -.93, which is a large correlation suggesting that the higher a teacher's initial SKS score was, the lower the change was expected to be.

Figure 3

Change in SKS Over Time



CHAPTER 5

Discussion and Implications

This analysis examined the effect of 2 years of participation in the study on teacher SCK. The analysis further examined the effect of 2 years of curricular and PD intervention on treatment teacher SCK compared to the control group. The data for both questions were longitudinal; therefore, multilevel modeling was used to analyze teacher predictors of SCK.

Conclusions and Discussion

Key findings for the two research questions are highlighted. Then, the discussion focuses on how these findings add new insights to the existing literature while also confirming the results of previous research.

Effect of time on teacher SCK. The first research question asked whether an additional year of participation in the study had an effect on teacher SCK beyond that of the first year. Each year participating in the study significantly increased TEST, demonstrating that according to that measure, a second year did have an additional effect over the first year. The fact that the reliability of TEST scores decreased in the treatment group but not in the control group is evidence that the intervention was partly responsible for this change, because it was probably caused by compression of the range of treatment teacher scores as they approached the maximum possible points. However, because we cannot attribute the changes in SCK entirely to the intervention, it seems that the changes may have been due, at least in part, to the experience of teaching fifth grade science for an additional year.

The first year of participation significantly increased SKS, but the second year did not. This lack of change in the second year may be partly explained by the strongly negative correlation between the SKS intercept and slope. The higher a teacher's SKS was at baseline, the smaller the observed change from baseline to T1 was expected to be. Since the first year led to higher average SKS, it is reasonable that the higher starting SKS at T1 would lead to a smaller change from T1 to T2.

We could not find any literature directly describing a relationship between CK and number of years teaching, using a test or another measurement tool, which is not surprising considering the little amount of current research on practicing teachers' CK in general. Shallcross et al. (2002) did find that teaching a topic increases the teacher's confidence with that topic, which is related to SKS, but they were not specifically measuring SCK. Arzi and White (2008) also found that the required curriculum is a major source of teacher CK. Our results support this idea that teachers learn science while teaching it, and that they continue to become more knowledgeable of science content each year, at least for the first two years. It is also interesting to note that years teaching was not a significant predictor of initial status of SKS, which means that while the teachers were more likely to feel that they know more science if they taught science for each year of the study, they were not more likely to feel that they had more science knowledge at the beginning of the study based on their total teaching experience. This finding raises questions about how much science the teachers taught in previous years. The finding also raises questions about how many of the teachers were veterans of teaching fifth grade, as opposed to having experience in other grades, because the experience of teaching fifth grade science may have increased grade-specific pedagogical CK.

Effect of predictors on teacher SCK. The second research question asked whether any of the predictors YEARS, DEGREE, COURSES, or GROUP had effects on teacher SCK. None of the predictors had a significant effect on TEST. While the result that number of science courses taken in college predicted higher initial SKS is consistent with the literature (Shallcross, Spink, Stephenson, & Warwick, 2002), COURSES also predicted lower primary change in SKS. This suggests that taking courses in college makes the teachers feel better about their SCK while decreasing their ability to improve their self-perceived SCK. Alternatively, this result could be due to the fact that a higher initial SKS caused by COURSES left little room for improvement in SKS.

The effect of GROUP on SKS indicates that the intervention had a significant effect on teachers' self-reported science knowledge. Since the intervention combined the teaching of SCK and pedagogy through hands-on inquiry-based approaches, this result is consistent with previous research indicating that this format of PD is highly effective in improving teachers' perceptions of their CK (Appleton, 2008; Shallcross, Spink, Stephenson, & Warwick, 2002). However, while the effect of the intervention on teachers' science knowledge test scores approached significance, the fact that it was not significant is inconsistent with this same literature. An examination of Year 3 results will be interesting, because the addition of a third year of data may make the analysis more sensitive to changes in teacher SCK. The addition of a fourth time point will also make the longitudinal analysis more stable, and therefore more accurate.

Contributions to the literature. This study makes important contributions to the literature in several ways. First, very few studies have measured the SCK of practicing teachers, and even fewer have had large sample sizes. This study measured the SCK of a large sample of practicing fifth grade teachers using two different measures of SCK. In doing so, we created an available data set with which future researchers can compare teacher SCK of other samples. Second, random assignment of schools into treatment and control groups allowed the test of causality of the intervention on teacher SCK. Third, 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. Finally, this study adds to the literature examining the benefits of PD for teachers' SCK, and demonstrates that there are advantages to long-term PD.

Limitations

This study had several limitations that may be addressed by future research. One limitation is a ceiling effect on the SCK test. On the pretest, five teachers (1.7% of the sample) earned perfect scores and 57 (19.5% of the sample) answered almost 90% of the questions correctly, leaving very little room for improvement. However, the test performed as expected, because it measured the ability of teachers to answer fifth grade level questions, so high scores are appropriate. Another potential limitation is another ceiling effect on the SKS on the questionnaire. Before beginning the intervention, 23 teachers (7.9% of the sample) gave themselves a 4 out of possible 4 on all items making up the scale. The addition of items for which teachers are less likely to choose 4 to the questionnaire would allow for a more sensitive analysis, and would likely demonstrate more change in teacher SCK. These ceiling effects may also partially explain the negative correlation between initial scores and change in scores for both measures.

As mentioned in the intervention section of this study, the state science standards changed from Year 1 to Year 2. This may have brought the science knowledge test slightly out of alignment with the standards in Year 2, potentially decreasing the TEST scores in that year. Teachers may also have felt less confident in their science knowledge due to the change in standards, which would have negatively affected their Year 2 SKS scores. Conversely, the posttest scores may have been lower if we had administered different forms of the test at each time point.

Finally, the addition of a question such as "how many years have you taught fifth grade science?" would have been enlightening, because it may have allowed us to determine why years teaching affected change in TEST but not initial TEST.

Implications for Practices

This study also has important implications for improving teacher SCK and practices. The results show that there is an additional benefit on fifth grade teachers' SCK for each year they teach. This finding suggests that elementary school principals should consider keeping teachers in the same grade and subject, because experience contributes to teacher SCK. An examination of the teacher attrition from Year 1 to Year 2 revealed that more than half of the teachers that left the study were reassigned to different subjects within the same school, rather than leaving the school for any reason. Finally, the results suggest that the intervention including research-based curriculum and professional development led to teachers becoming more comfortable with science content.

REFERENCES

Akerson, V. (2005). How do elementary teachers compensate for incomplete science content knowledge? *Research in Science Education, 35*(2), 245-268. doi: 10.1007/s11165-005-3176-8

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. (1988). From short- to long-term: Studying science education longitudinally. *Studies in Science Education, 15*, 17-53.

Arzi, H. J. (2004). On the time dimension in educational processes and educational research. *Canadian Journal of Science, Mathematics & Technology Education, 4*(1), 15-21. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=12526484&site=ehost-live

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). Handbook of research on teaching. In V. Richardson (Ed.), (4th ed 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

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

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

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

Brten, I., & Samuelstuen, M. S. (2007). Measuring strategic processing: Comparing task-specific self-reports to traces. *Metacognition and Learning, 2*(1), 1-20. doi: 10.1007/s11409-007-9004-y

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

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

Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning, 1*(3), 229-247. doi: 10.1007/s11409-006-9002-5

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

Davis, E. A. (2004). Knowledge integration in science teaching: Analysing teachers' knowledge development. *Research in Science Education (Australasian Science Education Research Association), 34*(1), 21-53. doi: 10.1023/B:RISE.0000021034.01508.b8

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

Dogan, N., & Abd-El-Khalick, F. (2008). Turkish grade 10 students' and science teachers' conceptions of nature of science: A national study. *Journal of Research in Science Teaching, 45*(10), 1083-1112. doi: 10.1002/tea.20243

Drechsler, M., & Van Driel, J. (2008). Experienced teachers' pedagogical content knowledge of teaching Acid–base chemistry. *Research in Science Education, 38*(5), 611-631. doi: 10.1007/s11165-007-9066-5

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabilitation Psychology, 56*(4), 267-288. doi: 10.1037/a0025579

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

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed ed.). New York: Pearson.

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

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

Goubeaud, K. (2010). How is science learning assessed at the postsecondary level? assessment and grading practices in college biology, chemistry and physics. *Journal of Science Education and Technology, 19*(3), 237-245. doi: 10.1007/s10956-009-9196-9

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

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), pp. 371-406. Retrieved from http://www.jstor.org/stable/3699380

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

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

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, 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.

Lombaerts, K., De Backer, F., Engels, N., van Braak, J., & Athanasou, J. (2009). Development of the self-regulated learning teacher belief scale. *European Journal of Psychology of Education - EJPE (Instituto Superior De Psicologia Aplicada), 24*(1), 79-96. Retrieved from http://iiiprxy.library.miami.edu:4420/login.aspx?direct=true&db=aph&AN=40514538&site=ehost-live

Marek, E. A., & Methven, S. B. (1991). Effects of the learning cycle upon student and classroom teacher performance. *Journal of Research in Science Teaching, 28*(1), 41-53. doi: 10.1002/tea.3660280105

Martin, W., Strother, S., Beglau, M., Bates, L., Reitzes, T., & Culp, K. M. (2010). Connecting instructional technology professional development to teacher and student outcomes. *Journal of Research on Technology in Education, 43*(1), 53-74. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=53500537&site=ehost-live

McCray, J. S., & Chen, J. (2012). Pedagogical content knowledge for preschool mathematics: Construct validity of a new teacher interview. *Journal of Research in Childhood Education, 26*(3), 291-307. doi: 10.1080/02568543.2012.685123

Mihladiz, G., & Timur, B. (2010). Assesing in-service science teachers’ pedagogical content knowledge: Views of pre-service science teachers. *Eurasian Journal of Physics and Chemistry Education, 1*(1), 89-100.

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

Moseley, C., & Utley, J. (2006). The effect of an integrated science and mathematics content-based course on science and mathematics teaching efficacy of preservice elementary teachers. *Journal of Elementary Science Education, 18*(2), 1-12. doi: 10.1007/BF03174684

Nehm, R. H., & Schonfeld, I. S. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science Teacher Education, 18*(5), 699-723. doi: 10.1007/s10972-007-9062-7

Nuangchalerm, P. (2011). In-service science teachers' pedagogical content knowledge. *Studies in Sociology of Science, 2*(2), 33-37. Retrieved from http://search.proquest.com/docview/920383888?accountid=14585

Pelaez, N. J., Boyd, D. D., Rojas, J. B., & Hoover, M. A. (2005). Prevalence of blood circulation misconceptions among prospective elementary teachers. *Advances in Physiology Education, 29*(3), 172-181. doi: 10.1152/advan.00022.2004

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

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

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

Rumsfeld, D. (2002). *U.S. department of defense office of the assistant secretary of defense (public affairs): News transcript.* Retrieved February 26, 2011, from http://www.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636

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

Sloan, H. A. (1993). *Direct instruction in fourth and fifth grade classrooms.* (PhD, Purdue University). *ProQuest Dissertations and Theses,* , 125.

Stearns, L. M., Morgan, J., Capraro, M. M., & Capraro, R. M. (2012). A teacher observation instrument for PBL classroom instruction. *Journal of STEM Education : Innovations and Research, 13*(3), 7-16. Retrieved from http://search.proquest.com/docview/1015211574?accountid=14585

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

Tchoshanov, M. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics, 76*(2), 141-164. doi: 10.1007/s10649-010-9269-y

Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Human Development (0018716X), 44*(2), 84-97. doi: 10.1159/000057048

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

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. L. Gabel (Ed.), *Handbook of research on science teaching* (pp. 177-177-210). New York: Macmillan.

Yip, D. (1998). Identification of misconceptions in novice biology teachers and remedial strategies for improving biology learning. *International Journal of Science Education, 20*(4), 461-477. doi: 10.1080/0950069980200406

Appendix A

Teacher Science Knowledge: Conceptualization, Professional Development, and Measures

"As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know." – Donald Rumsfeld (2002)

As much as Donald Rumsfeld was ridiculed for the above statement, it actually gets right to the heart of a question this paper attempts to answer. What is knowledge, especially in regards to the types of knowledge necessary for teachers to effectively teach science? While most people believe that teachers should know the content that they are teaching, there is not yet strong research to show the importance of teacher content knowledge (CK) (Porter, 2012). However, since "[t]eachers are now seen as the single most important factor in terms of school variables for producing student learning" (Porter, 2012, 36:55), it is imperative that researchers begin to examine the importance of teacher CK.

Before this essential research can proceed, a valid method of measuring teacher CK must be developed. If teacher CK does affect student achievement outcomes, then it is also important to find ways to improve teacher CK through professional development (PD). Furthermore, it is equally important to examine whether changes in teacher CK through PD affects student achievement outcomes. This literature review will examine each of these issues in order to demonstrate the need for further research in their respective areas, especially in regards to science content knowledge (SCK). Additionally, the review will examine whether studies involved pre-service or in-service teachers and the grade levels taught by these teachers. The first question that must be answered in order to begin to think about the others is: what is teacher science knowledge?

Conceptualization of Teacher Science Knowledge

Types of Teacher Knowledge

Surprisingly, very little of the research on teacher knowledge looks at how to improve teacher knowledge of science content (Fleer, 2009; Heller, Daeler, Wong, Shinohara, & Miratrix, 2012; Shallcross, Spink, Stephenson, & Warwick, 2002), how to measure it, or its impact on classroom practice or student achievement (Chinnappan & Lawson, 2005; Porter, 2012). In fact, on the rare occasions that teachers' science content knowledge (SCK) is addressed, it is usually studied in pre-service teachers rather than those currently teaching (Ball, Lubienski, & Mewborn, 2001). The lack of research on CK is largely due to the introduction of the term "pedagogical content knowledge" (PCK) by Lee Shulman in 1986. In the 1970's and 1980's, scholars such as Bloom, Gagne, and Schwab attempted to describe and classify CK, without empirically studying how teacher CK affected student outcomes (Shulman, 1986). Since Shulman's 1986 paper describing the classification of PCK was published, forms of teacher knowledge *other* than CK have taken priority in studies. CK is rarely anymore studied except in how it relates to PCK. While PCK is an important construct, its study does not replace the need for an understanding of CK itself, and there is a major gap in the recent literature on this highly important topic. In order to understand the importance of studying CK itself, it would be helpful to first examine the types of teacher knowledge as they are discussed in the field.

Doing so will illuminate not only what we do know about teacher knowledge, but, more importantly, what we don't.

The most commonly accepted breakdown of teacher knowledge is into the domains of: subject matter CK, curricular knowledge, pedagogical knowledge (PK), and PCK (Shulman, 1986). Knowledge of the subject matter includes knowledge of the content, the organization of the discipline, and the ways a discipline evaluates new knowledge (Ben-Peretz, 2011). Curricular knowledge refers to how well a teacher understands the various curricular materials available to help students reach understanding of the material (Shulman, 1986), and is more closely related to PCK than to CK. PK is knowledge of methods and techniques for effective teaching. PCK was introduced by Shulman in 1986 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). It is important to understand that PCK refers to the skills necessary to teach particular topics to particular students, as opposed to PK, which is the skill set needed to teach in general (Van Driel, De Jong, & Verloop, 2002).

Shulman (1986) introduced the idea of PCK because he felt that, until then, researchers and teachers looked at content *or* pedagogy, and there was no paradigm for examining their interaction. Kaya (2009) describes four reinterpretations of PCK from 1992-1999, each building on the last. In the last one of the series, Magnusson, Krajcik, and Borko describe five components of science PCK as "orientation toward science teaching, knowledge of the curriculum, knowledge of science assessment, knowledge of science learners, and knowledge of instructional strategies" (Kaya, 2009, p. 962). Kaya goes on to describe the most common definition of PCK among science educators as "subject matter knowledge and pedagogical knowledge that consists of knowledge of students' learning difficulties and conceptions, and instructional strategies" (Kaya, 2009, p. 962), along with knowledge of curriculum and assessment. In other words, PCK is knowing what misconceptions students hold and how to direct them, being prepared for the types of questions and struggles students will have with the material, and in general knowing how to teach a particular subject well.

Lack of SCK is often cited as the cause of the inability of teachers to teach science (Fleer, 2009). 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). Without a strong understanding of content, a strong PCK is impossible to achieve (Kaya, 2009; Van Driel et al., 2002). Van Driel et al (2002) found that "growth of PCK was influenced mostly by the pre-service teachers' teaching experiences, university-based workshops, and meetings with mentors" (Kaya, 2009, p. 965), but that the variations in PCK observed were due to differences in CK. Others have found that teachers without sufficient SCK are unable to identify and correct students' misconceptions. Kaya (2009) found a large effect of CK on PCK in pre-service teachers. "Teachers' lack of SCK limit[s] their ability to anticipate the directions in which pupils' scientific learning might proceed" (Shallcross et al., 2002, p.1295).

CK alone is not enough to achieve strong PCK, because knowing a topic is not the same as knowing how to teach a topic (Van Driel et al., 2002). For example, Van Driel et al. (2002) found that because pre-service high school chemistry teachers had such an implicit understanding of the relationships of microscopic and macroscopic perspectives, they tended to jump between the two without explicitly telling students they were doing so. That is why, while strong CK is important for effective teaching, without strong PCK, teachers cannot effectively transmit that knowledge (Davis, 2004). Pre-service teachers are often lacking in their PCK of effective science teaching, largely because it is impossible to address all possible pedagogical situations in a teaching program (Kaya, 2009). However, this limited scaffold of PCK that teachers begin with is important as a starting point for new teachers to develop from (Davis, 2004).

In order for teachers to be able to successfully carry out a project-based or inquiry-based science curriculum, it is important for teachers to have both extensive CK and PCK (Basista & Mathews, 2002; Kanter & Konstantopoulos, 2010). Teachers lacking in either skill will not be able to diagnose and change students' misconceptions of science topics during inquiry lessons. In fact, a lack of SCK often leads to teaching science as a series of facts or not teaching it at all (Shallcross et al., 2002). Considering the importance of addressing misconceptions in order to support effective teaching, and their direct relationship to SCK, we will now examine what misconceptions are and the challenges they put forth.

Teacher Misconceptions

A scientific misconception is "a vague and imperfect or mistaken understanding of something; one that is commonly held by learners, is difficult to teach away, and is at variance with current scientific knowledge" (Pelaez, Boyd, Rojas, & Hoover, 2005, p. 172; Wandersee, Mintzes, & Novak, 1994). In other words, it is a flaw in SCK. Misconceptions can be very broad. For example, many teachers think science is a known set of facts or laws, rather than a dynamic body of knowledge (Jarvis, Pell, & McKeon, 2003). Correcting student misconceptions is often a major challenge, because teachers often hold the same misconceptions (or alternative conceptions) themselves (Davis, 2004; Jarvis et al., 2003; Shallcross et al., 2002). When teachers are asked about a topic they do not understand well, they often change the subject to a related topic they have a stronger understanding of, essentially side-stepping the weaker topic (Davis, 2004). Student teachers do this on exams, as well (Pelaez et al., 2005). Several studies have also shown that pre-service teachers tend to have very superficial and often incorrect knowledge of science content (Kaya, 2009).

While many pre-service programs now require more science coursework, teachers who have been in service for a while often did not have that requirement (Jarvis et al., 2003; Shallcross et al., 2002). Even those who did take required coursework rarely took enough to satisfy the breadth of content science teachers are responsible for teaching (Shallcross et al., 2002).

Jarvis et al. (2003) studied elementary teachers' understanding of various scientific concepts they had been teaching before and after a six-month in-service course. Before the test, half of the teachers had no idea why a bulb lights, 31% thought current is used up as it passes through a bulb, 12% did not know what direction a current flows, and 81% were unable to correctly answer two out of four questions about series and parallel circuits. They also found that some teachers did not know the difference between melting and dissolving. Some teachers named speed, energy, and momentum as examples of forces. Opposing forces and the directions of forces were often ignored. Over 16% of teachers did not understand that there is gravity on the moon. When a question was asked to determine if teachers understood how to control variables, the mean correct score was a 41%. Asking teachers to design a scientific investigation resulted in a mean correct score of about 69%. After receiving PD on science content, teachers showed gains in every topic, but it was determined that at least 10 days of training was necessary to develop scientific understanding.

The previous study was not an isolated incident. When testing science teacher knowledge of the circulatory system, Yip found that novice secondary biology teachers were unable to relate blood flow, blood pressure, and blood vessel diameter (Pelaez et al., 2005; Yip, 1998). The majority of those novice teachers also had misconceptions about the nutritional process, gas exchange, homeostasis, and reproduction and variation (Yip, 1998). Pelaez et al. (2005) found that pre-service elementary teachers had erroneous ideas about blood pathway, blood vessels, gas exchange, gas molecule transport and utilization, and lung function, and that asking students to draw the concepts revealed more errors than other methods of assessment. Many of these misconceptions persisted after re-teaching, and most of them would not have been detected in an ordinary classroom setting.

In a study involving 44 in-service biology teachers, Nehm and Schonfeld (2007) found an alarming number of misconceptions about evolution before the teachers began a graduate-level evolution course. "Commonly held misconceptions included that evolution is a weak scientific idea because it is only a theory, there are no transitional intermediates in the fossil record, mutations are always harmful, and humans and dinosaurs coexisted" (Nehm & Schonfeld, 2007, p.708). These misconceptions were often used to explain evolutionary concepts. The teachers also failed to use important concepts in natural selection, such as competition for resources, in their explanations. The surprising finding, however, is that teachers rarely changed the way they taught evolution, even after taking the course and showing a significant increase in knowledge. This finding highlights the importance of pedagogical development along with SCK development.

Clearly, the one area of teacher SCK that has been well studied is that of misconceptions, or lack of SCK, on specific science topics. Effective teaching requires strong PCK, and that PCK requires the ability to correct misconceptions in students. Teachers often hold the same misconceptions as their students, a situation that can presumably be at least partially remedied by improving teacher SCK. While the literature has several examples of individual topics that small groups of teachers tend to misunderstand, we still do not know how to measure the overall SCK of teachers in a statistically robust manner, the direct importance of teacher SCK to student achievement, or how to actually improve teacher SCK. Let us turn next to what is necessarily the next part of the process. It is easier to improve and demonstrate the importance of SCK if we can measure it first.

Measures of Teacher Content Knowledge

Teacher CK has typically been assessed using self-reports of both courses taken and perceptions of CK, interviews, and less frequently, using classroom observations. However, tests of the content knowledge of practicing teachers are rare. Each of the measures presents strengths and limitations. Since courses taken seem to be the most frequently used proxy of CK in large scale studies, we will look at that measure first, followed by surveys, interviews, classroom observations, and written tests.

Courses Taken as Proxy of CK

Mathematics CK (MCK) has been studied in teachers much more than SCK has, so the field of science education can draw on mathematics education for some foundational research. In fact, teacher MCK was one of the first variables of teaching investigated in the 1960s (Ball et al., 2001). While it is recognized that sufficient 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), there is still not sufficient understanding of the MCK it takes to teach effectively. Without knowing what kind of MCK is necessary for effective teaching, it is difficult to help teachers develop the knowledge they need.

Many of these studies use the proxy of number of mathematics courses taken in college as their measure of MCK. When researchers have counted the number of mathematics courses taken as a measure of teacher MCK, the effects have typically been small (Ball et al., 2001; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009). In 1979, Begle found that taking mathematics courses led to positive effects on student achievement in 10% of cases and negative effects in 8%, while majoring or minoring in mathematics yielded positive main effects in 9% of cases and negative main effects in 4% (Ball et al., 2001). 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).

That is not to say that college courses taken by teachers are unimportant. The Longitudinal Study of American Youth (LSAY) found that 90% of variation in student achievement in reading and mathematics seems to be attributable to differences in teacher qualifications (Ball et al., 2001). Monk (1994) analyzed the number of mathematics and science courses taken in college by high school teachers measured via the LSAY, and found that up to five courses showed positive effects of a 1.2% gain in student test score per course, but then the positive effects level off to a .2% increase. The largest effects were found in courses on mathematics pedagogy, rather than mathematics content. This finding suggests that for mathematics teachers, PK may be more important than CK. Monk (1994) also used the LSAY to analyze high school science teacher knowledge, 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 performance. Monk suggested that life sciences might have had a negative effect on juniors because teachers who took more life science courses often took fewer physical science courses, and juniors were usually enrolled in physical science classes. Therefore, more life science experience of the teacher meant less preparation for the junior level course material.

A more direct indication of why number of courses is not predictive of teacher SCK may be Rice's (2005) work, which found that while 87.3% of her 414 participating undergraduate elementary program students had completed at least one biology course, only 21.3% were able to identify human, dog, worm, and spider all as animals. 70% failed to identify the worm and 23% failed to identify the human as animals. Therefore, it seems that students either are not learning or are not retaining, even over the short term, basic concepts important to the subjects they are learning. In either case, the teachers that would have fallen into these categories do not have the SCK necessary to teach these topics.

Self-Report Surveys

Self-report 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, partially due to the expected low consistency between acting and reflecting upon acting caused by the role signs play in the regulation of conduct (Valsiner, 2001). Oddly, literature searches turn up very little recent research on the validity of teachers' self-reporting. 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 tend to have moderate to high correlations.

Most recent studies seem to concentrate on the validity of *student* self-reporting, so we will examine that next. Unfortunately, while student learning strategies are usually assessed by self-report (Brten & Samuelstuen, 2007), the results from student data are very inconclusive. Ivar and Samuelstuen (2007) found that among Norwegian tenth grade students there was a 76%-83% correlation between self-reports of strategies and separately scored traces of the same strategies (Brten & Samuelstuen, 2007). On the other hand, Cromley & Azevedo (2006) found that prospective self-report data on reading strategies was not significantly correlated with the results of concurrent multiple-choice or think-alouds.

There are a few studies that have looked at teachers specifically. In a validity test of the Teachers' Sense of Self-Efficacy Scale (TSES) spanning five countries (Canada, Cyprus, Korea, U.S., and Singapore), the researchers developed a model that was shown to be good both within and between countries (Klassen et al., 2009). Another study found a good fit for a model of teachers' self-report of their own learning behaviors (Gordon, Dembo, & Hocevar, 2007). The Self-Regulated Learning Teacher Beliefs Scale developed in Belgium to assess primary school teacher beliefs about the introduction of self-regulated learning in daily classroom practice was also internally consistent with a good fit (Lombaerts, De Backer, Engels, van Braak, & Athanasou, 2009). Teachers' self-reported enthusiasm for teaching mathematics is strongly correlated with self-reported and student-reported quality of instructional behavior, but teachers' 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). Overall, teacher self-report scales tend to be statistically valid more often than those designed for students, but they are still not considered by the research community to provide as good information as classroom observations or interviews.

Interviews

Interviews have the advantage of allowing the development of a trusting relationship between the interviewer and interviewee, which in turn elicits comprehensive and truthful information about implementation (Desimone, 2009). The primary disadvantages of interviews are that they are time consuming, expensive, and subject to interviewer bias. Recent research suggests that interviews and surveys tend to elicit very similar responses, contrary to popular opinion (Desimone, 2009). However, interviews are better than surveys for capturing in-depth and nuanced constructs, and are appropriate for providing narratives and anecdotes, and are overall better than surveys for qualitative research. As is true of other measurement methods, SCK has not often been evaluated through the use of interviews, so uses of this technique in evaluating PCK are briefly described below.

Interviews have been used to evaluate the science PCK of experienced secondary school chemistry teachers, allowing the researchers to explore what the teachers knew about students' difficulties and how to overcome them (Drechsler & Van Driel, 2008). One case study used interviews to discover that pre-service teachers who observed in-service elementary teachers found the in-service teachers lacking in their SCK and science PCK, but considered their own SCK sufficient and their PCK as something that would develop when they began teaching (Mihladiz & Timur, 2010). McCray and Chen (2012) found that McCray's interview for measuring preschool teachers' mathematics PCK is a strong predictor of the teachers' math-related language use in both structured and non-structured pedagogical settings. The interview also showed small but significant relationships with child outcomes measured by the *Test of Early Mathematics Ability*, 3rd ed. So interviews can be used validly to measure some forms of teacher knowledge, but classroom observations are generally seen as less biased.

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 in PD activities and classrooms. Observation, like interviews, allows researchers to make finer distinctions in teacher practice than surveys. For observations to be considered reliable and valid 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 drawback to observation is that it is "the most time-consuming and expensive method of measuring [PD] and teaching" (Desimone, 2009, p.188). Observations also tend to be episodic in nature, leading to an incomplete picture of the teacher's observed attributes.

The National Science Foundation's Local Systemic Change through Teacher Enhancement funded 88 individual projects, and evaluators of each project were required to observe a single lesson of a random sample of teachers targeted by their project, partly to determine whether teacher participation in PD predicts lesson quality (Bowes & Banilower, 2004). Hierarchical generalized linear models, with lessons nested within projects, found that 20-39 hours of PD led to teachers earning significantly higher observation scores on lesson quality than those not receiving PD, but more than 39 hours did not lead to a higher probability of earning a high observation score.

This type of quantitative observation data is very rare. It is much more common to find qualitative observation data that examines only a few teachers (Akerson, 2005; Appleton, 2008; Nuangchalerm, 2011). Observations are also used to evaluate the fidelity of implementation of curricula and pedagogical practices of teachers (Stearns, Morgan, Capraro, & Capraro, 2012). An easier, cheaper, and less biased form of measurement may be tests of CK.

Written Tests

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 certification process for about a century (D'Agostino & Powers, 2009). These tests vary by state, and which states require them has varied over time. A meta-analysis of 123 studies involving both elementary and secondary teachers showed that while the median correlation between certification test scores and college GPA was .55, the tests scores showed a correlation with supervisor and principal ratings of only .05 for pre-service teachers and of .11 for in-service teachers. There was also a low correlation between college GPA and these supervisor ratings (.07), although the correlation was higher in the first year of teaching (.25) than with any other indicator. According to this analysis, it would seem that college GPA is a better indicator of teacher performance than test scores, but *only in the first year of teaching*. There is not, however a test that is accepted as being well-designed for assessing the SCK of in-service teachers, so it is necessary for further research to be done before strong conclusions can be reached about the usefulness of this kind of measure.

Effects of Teacher Attributes on Student Achievement Outcomes

Very little research exists demonstrating the effect of teacher CK, of any subject, on student outcomes. A study of Turkish 10th grade students' views of nature of science found that 8 out of 14 items showed that teachers' and students' views were statistically the same, and the differences in some others were small (Dogan & Abd-El-Khalick, 2008).

A rare recent study showed that PD that includes science content has a significant positive effect on elementary student science content test scores regardless of the form of PD, partly due to the effect on teacher SCK (Heller et al., 2012). The study examined the differences among three 24-hour PD programs that varied in the PCK components while maintaining the SCK components identical, coming from the WestEd *Making Sense of SCIENCE* course for elementary teachers on electric circuits. The Teaching Cases course allowed teachers to discuss narrative cases drawn from actual classroom episodes and written by classroom teachers. The Looking at Student Work course led teachers through a structured, collaborative analysis of their own students' work while concurrently teaching an electric circuits unit. The Metacognitive Analysis course led teachers in reflective discussions about their own learning processes to connect their own experiences to how students learn. All three courses improved teacher SCK significantly more than control teachers according to test scores. Additionally, all three courses improved student test score gains significantly more than control test score gains. HLM analysis revealed that teacher SCK was a significant predictor of student test scores ($p<.001$).

As mentioned previously, there are studies of teacher MCK where studies on SCK are lacking. A study of first and third grade teachers showed that there is a relationship between teacher "mathematics content knowledge for teaching" and student achievement (Hill, Rowan, & Ball, 2005). Teachers in the bottom 20% had first grade students with significantly lower test scores than those in the upper 80%, and teachers in the bottom 30% had third grade students with lower test scores than those in the upper 70%. This finding showed that there was a threshold level of MCK where the teacher had enough

MCK to teach the material required for that grade, but MCK above that threshold did not affect student achievement further.

Tchoshanov (2011) argued that, while there is extensive research extant on the connection between teacher MCK and student achievement, since most of these studies used proxies such as number of courses taken, the research does not reveal the specific MCK needed for effective teaching. As such, he created a teacher MCK test to differentiate between three types of MCK: knowledge of facts and procedures, knowledge of concepts and connections, and knowledge of models and generalizations. When correlated with student passing rates on a state standardized test, only knowledge of concepts and connections was significant, and he stated that knowledge of facts and procedures clearly has little impact on student achievement. Knowledge of models and generalizations was also found to not have a significant effect on student achievement. The implication for SCK is that there may also be different impacts based on different types of teacher CK and different types of student achievement, but first the general relationship between SCK and student achievement should be established.

Now would be a good time to pause and summarize what is important for teachers to know and why. A strong PCK is necessary for a teacher to be able to teach effectively, especially when it comes to such activities as correcting misconceptions and inquiry-based learning. PCK is impossible to achieve without strong CK, which studies of teacher misconceptions show is often lacking in science teachers. Improving teacher SCK through PD has been shown to improve student achievement (Heller, 2012). The final piece of this puzzle is to examine how PD can be used to improve teacher knowledge in general and CK specifically.

Impact of Professional Development on Teachers' Content Knowledge

Professional development is important to improving the quality of U.S. schools, the effectiveness of policy for teachers and teaching practice, and student achievement (Desimone, 2009). Of course, the PD must be effective in order to have an impact.

Impact of PD on teachers' knowledge and skills

The effectiveness of professional development on teacher outcomes was studied using data from a Teacher Activity Survey conducted to evaluate the Eisenhower Professional Development Program (Garet, Porter, Desimone, Birman, & Yoon, 2001). The study used teacher self-report to examine both core and structural features of PD. Core features included (a) a focus on content knowledge and how students learn that content, (b) opportunities for teachers to engage in active learning, and (c) coherence with other activities for teacher learning and development. All three of these core features had positive effects on teachers' self-reported enhanced knowledge and skills in the areas of curriculum, instructional methods, approaches to assessment, use of technology, strategies for teaching diverse student populations, and deepening knowledge of mathematics. Enhanced knowledge and skills, in turn, had a positive effect on teaching practice. They also identified structural features including (a) the form of the activity, (b) sufficient duration in terms of both the number of contact hours and the span across the calendar year, and (c) collective participation of teachers from the same school, department, or grade level. The results indicated that the reform activities were longer and had a greater effect on teacher knowledge and skills than other forms of activity. Longer time span and more contact hours each led independently to improved outcomes for all three core features.

Another study found a similar relationship between PD fidelity to the curriculum and teacher outcomes, and additionally found a positive relationship between the measured teacher outcomes and student outcomes (Martin et al., 2010). The study included 287 elementary and middle school teachers and 9 high school teachers participating in four-hour PD sessions. They "obtained classroom visit data for 272 teachers, PD fidelity data for 269 teachers, and lesson plans from 180 teachers" (Martin et al., 2010, p. 59), and used standardized test scores as the student outcome measure. The researchers identified five concepts that served as the foundation to the PD: modeling instruction, community building, technology utilization, connection to practice, and inquiry-based learning. No significant correlations were found between PD fidelity and classroom visit activities. All of the PD fidelity factors were related to higher student test scores.

Kanter and Konstantopoulos (2010) studied the effect of a graduate-level course for in-service teachers learning the *I, Bio* curriculum on teacher SCK and PCK, as well as the effects of teacher SCK and PCK on minority student achievement. Both teacher SCK and PCK improved with the PD, but there was no control group to compare to. Regression analysis showed that teacher average post-test CK and average post-test PCK were predictors of minority student achievement in the specific content of the curriculum, according to the tests designed by the researchers.

Aside from improving teacher CK and PCK, PD often attempts to improve the perceptions of the targeted teachers. In fact, there is much more research available on the effects of PD on teachers' perceptions of their CK than on PD's effect on CK itself. It can also be argued that the self-report measure used so frequently as a proxy for teacher

CK is governed more by the teacher's perceptions than by the actual CK of the teacher. Therefore, we will examine this phenomenon next.

Teachers' perceptions of their own CK have a direct effect on how they teach. Teachers who are confident in their CK feel more competent as teachers, regardless of their actual competence in the subject (Shallcross et al., 2002). The findings of Shallcross et al. (2002) suggest that teachers who think they know what they are talking about but do not are less anxious than teachers who actually know what they are talking about but are unsure of themselves. Teaching a topic also increases confidence with that topic. Professional development courses which integrate science content and pedagogical knowledge have been found to increase teachers' confidence in teaching science.

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 tests, the teachers were only given daily mini-quizzes, so there is no data for their resulting cumulative CK. However, surveys demonstrated that the teachers *felt* significantly more comfortable with the content and more prepared to teach it. It has also been found that taking a content-based earth science course has a significant effect on pre-service elementary teachers' self-efficacy, especially their belief that they have the ability to influence student outcomes (Moseley & Utley, 2006).

Lack of confidence can be as detrimental as confidence is helpful to teaching ability. Elementary school teachers are often hesitant to teach science due to low self-confidence in science or due to lack of SCK (Appleton, 2008; Heller et al., 2012) or due to negative experiences in secondary school (Shallcross et al., 2002). Early childhood teachers have been found to not recognize their own informal science knowledge as valuable in teaching science (Fleer, 2009).

Teachers tend to report that they are less confident in physical sciences than life sciences, and they tend to prefer to teach accordingly, but primary school teachers are usually expected to spend most of their time teaching physical science (Shallcross et al., 2002). This lack of confidence is in spite of the fact that pre-service programs tend to emphasize physical sciences over life sciences. It would seem that this is caused by a difference between the ways the two general disciplines are taught in college. One study found that undergraduate chemistry and physics professors were significantly less likely to use formative assessments than biology professors, and that they were twice as likely to grade on a curve (Goubeaud, 2010). Both of these practices lead to a less reliable assessment of what pre-service teachers actually know. This phenomenon may partly explain why the number of courses taken is not an accurate indicator of teacher SCK.

Conclusions and Implications

Conclusions

Most current research is on PCK and not CK. Although it is widely acknowledged that PCK is impossible without sufficient CK, few researchers seem to have studied CK since the advent of the term PCK. Research that is related to CK is that on misconceptions, which shows that many teachers have similar misconceptions to their students, making it impossible for the teachers to correct the students' misconceptions (Davis, 2004; Jarvis et al., 2003; Shallcross et al., 2002).

Most research is on pre-service teachers, not in-service teachers (Ball et al., 2001). No matter the training, there are going to be differences between prospective teachers and practicing teachers. Not only have the practicing teachers had an opportunity to develop PK, in turn allowing the development of PCK, but they also have been out of school longer and are likely to have less up-to-date SCK. Science is changing constantly, so even five years out of school can be enough for a teacher who is not vigilant to end up out-of-date on her CK.

There have also been few studies that concentrate on the content of PD, and how much content teachers actually learn during PD opportunities (Garet et al., 2001). A review of over 1,300 empirical studies that attempted to address the link between PD and teacher learning only turned up two studies (Marek & Methven, 1991; Sloan, 1993) that focused on science and were empirical randomized controlled trials or quasi-experimental designs with valid measures of student and teacher outcomes (Heller et al., 2012). Therefore, the link between PD and CK is another area that requires additional research. Much more common than studies that test the effect of PD on CK are studies that examine the effect of PD on teacher perceptions. Changes in teacher perceptions have been shown to lead to changes in teacher practices. However, perception is not a substitute for the still poorly understood field of teacher CK.

Implications

While some researchers are now beginning to examine the importance of teacher SCK, there is nowhere near enough data published on this essential topic. Future research should accomplish several goals. First, effective assessment methods of SCK must be developed, so that both current levels and improvement of teacher SCK can be evaluated and monitored in the future. Second, these new assessment methods must be applied to determine how strong teacher SCK is on average, and disaggregated by teacher subgroups. Third, it must be determined how strongly teacher SCK affects student achievement, because that is the primary goal of education.

Effectively measuring teacher SCK and connecting it to student achievement can lead to a profound shift in the PD that teachers currently receive in the United States, from concentration on teacher perceptions and pedagogy to concentration on SCK. Teachers with a weak background in science would benefit greatly from this, but teachers with strong SCK could also receive update classes discussing new advances in science. The lack of effect of college courses taken on teacher SCK should also lead to changes in the way prospective teachers are taught science, so that teachers leave their preparation programs with a solid foundation in SCK. If teachers are, in fact, the most important factor leading to student learning, it is imperative that we give teachers the tools they need in the most effective way possible.

Appendix B

Concepts included in the 2010-2011 Sunshine State Science Standards
(In the order presented in the P-SELL curriculum)

Strand H
- Scientific process and inquiry
- Use of basic laboratory tools
- Making observations, collecting data, and analyzing data
- Communicating results

Strand A
- All matter is composed of elements
- Atoms and molecules are too small to be seen with the eye
- The relationship between mass, volume, and density
- Matter can be changed from one state to another by heating and cooling
- The ways in which molecules combine in mixtures, solutions, and compounds to form new materials

Strand B
- The sources of energy that different organisms need to survive
- How energy flows through living and non-living systems
- How food chains are connected into food webs
- The different forms that energy can take
- How energy is transformed as it passes through systems
- The difference between renewable and non-renewable resources

Strand C
- Transformations of potential and kinetic energy
- How to measure an object's motion
- Forces that influence how an object moves
- How different forces influence the movement of an object
- The result of more than one force acting on an object
- The difference between balanced and unbalanced forces

Strand D
- Earth is constantly changing
- There is a complex interaction between and within Earth's major systems
- The role of weathering and erosion in shaping and reshaping Earth's surface
- How the atmosphere impacts weather and climate

Strand E

- How the Earth’s tilt causes changes in seasons and length of day

Strand F

- How and why living things are classified
- The major groups of living things
- How living things are similar and different

Strand G

- What are adaptations
- How variations in ecosystems affect populations
- What is photosynthesis
- How changes in an organism’s habitat can be beneficial or harmful
- All living things must compete for Earth’s limited resources
- The size of a population depends on resources available it its community

Appendix C

Concepts included in the 2011-2012 Next Generation Sunshine State Science Standards
(In the order presented in the P-SELL curriculum)

Big Ideas 1 and 2

- Scientific process and inquiry
- Making observations, inferences, and predictions
- Making accurate measurement when collecting data
- Analyzing data and communicating results
- Reflecting on results and making connections

Big Idea 8

- Use of basic laboratory tools and appropriate units to make accurate measurements
- The nature of matter
- The ways in which molecules combine in mixtures and how mixtures can be separated

Big Idea 9

- All matter is composed of elements
- Atoms and molecules are too small to be seen with the eye
- Matter can be changed from one state to another by heating and cooling
- Chemical changes result in the formation of new substances

Big Idea 13

- Interpreting and making graphs
- The motion of objects, including distance and speed
- Isaac Newton's Laws of Motion
- How friction and magnetism influence the movement of an object

Big Ideas 10 and 11

- The different forms that energy can take
- How energy is transformed as it passes through systems
- The reflection and refraction (bending) of light
- The use of subatomic particles to explain electrical energy
- Transformations of potential and kinetic energy

Big Idea 7

- How water changes states as part of the water cycle
- The meteorological and geographical factors that affect weather and climate

Big Idea 6

- Earth is constantly changing
- Process of rock formation
- Mineral properties used for classification
- The role of weathering and erosion in shaping and reshaping Earth’s surface
- The different forms that energy can take
- The difference between renewable and non-renewable resources

Big Idea 5

- The different systems involving the Earth that extends beyond our planet
- The different objects of the Solar System and Earth’s position in it
- How the rotation and revolution of the Earth and apparent movements of the Sun, moon, and stars are connected

Big Idea 14

- The functions of the organs of some most of the human body systems, including skeletal, muscular, circulatory, respiratory, digestive, urinary, nervous, and reproductive systems
- How and why living things are classified
- The major groups of living things
- How living things are similar and different
- The function of organs and other physical structures of plants and animals
- Structures in plants and their roles in food production, support, water and nutrient transport, and reproduction

Big Idea 17

- The sources of energy that different organisms need to survive
- The flow of energy from the sun as it is transferred along a food chain
- What is photosynthesis?
- What are adaptations?
- How variations in ecosystems affect populations
- How changes in an organism’s habitat can be beneficial or harmful
- All living things must compete for Earth’s limited resources
- The size of a population depends on resources available in its community

Appendix D

Science Knowledge Scale

Please indicate how knowledgeable you feel about teaching each of the following science topics at your grade level.

	Not knowledgeable	Somewhat knowledgeable	Knowledgeable	Very knowledgeable
a. Nature of Science	1	2	3	4
b. Physical Science	1	2	3	4
c. Earth and Space Science	1	2	3	4
d. Life Science	1	2	3	4