Elementary teachers' science content knowledge: Relationships among multiple measures

1 author:

Brandon S. Diamond
David Posnack Jewish Day School, Davie, FL

Elementary Teachers' Science Content Knowledge: Relationships Among Multiple Measures

Brandon S. Diamond, Jaime Maerten-Rivera, Rose Rohrer
University of Miami

Okhee Lee
New York University

Abstract

This study examines relationships between measures of teacher science content knowledge using multiple instruments. A teacher questionnaire and a science test were administered to 203 elementary school teachers. A random subsample of 62 teachers was observed during science instruction. All teachers were asked the number of science courses they took in college. Significant positive correlations were found between science test scores and both self-reported science knowledge and classroom observation scores and between science courses taken and self-reported science knowledge. Test scores and observations were not correlated with courses taken, nor were observations correlated with self-reported science knowledge. These results suggest the evaluation of measures of teacher content knowledge, which are not interchangeable, as is often assumed in the literature.

Keywords: teacher science content knowledge, science test, self-report, observation, science courses taken.

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), researchers and policymakers are trying to determine how to best improve the preparation of teachers. Surprisingly, very little of the research looks at how to improve teacher knowledge of science content (Fleer, 2009; Shallcross, Spink, Stephenson, & Warwick, 2002). In fact, on the rare occasions that teachers' science content knowledge (SCK) is addressed, it is usually studied in preservice teachers rather than those currently teaching (Ball, Lubienski, & Mewborn, 2001). There is also a lack of information about effective methods for measuring teacher content knowledge, in science or any other subject area. Researchers often treat various measures of SCK as interchangeable (i.e., they measure the same constructs), and choose their method for measuring SCK based on financial and logistical reasons.

The purpose of this study is to examine SCK of elementary school teachers using multiple measures. Specifically, the study examined the following two research questions:

1. What was teachers' initial SCK on each of the measures, including (a) scores on an SCK test, (b) teachers' self-reported SCK using a questionnaire, (c) observation ratings of SCK during classroom instruction for a subsample, and (d) teachers' self-reported number of science courses taken in college (both undergraduate and graduate)?
2. What were the relationships among these four measures of SCK?

Literature Review

There is very little research available on practicing teachers' overall SCK. This is in spite of the fact that lack of SCK is often cited as the cause of the inability of teachers to teach science effectively (Fleer, 2009). In fact, variations in teachers' scientific knowledge and understanding have been identified as the main factor responsible for the differences in the quality of elementary science teaching (Shallcross et al., 2002). A recent study showed that professional development that included science content had a significantly positive effect on elementary students' science content test scores regardless of the form of professional development, partly due to the effect on teacher SCK (Heller, Daeler, Wong, Shinohara, & Miratrix, 2012). While the importance of teacher SCK is becoming clearer, there is still no method that has been established as the best way to measure teacher SCK. This section first discusses the construct of teacher content knowledge (CK) in general, and then discusses the four types of measures used to assess teacher SCK in this study.

Teacher Content Knowledge

Surprisingly, very little of the research addresses how to improve teacher knowledge of science content (Fleer, 2009; Heller et al., 2012; Shallcross et al., 2002), measures, or impact on classroom practice or student achievement (Chinnappan & Lawson, 2005; Porter, 2012). The lack of research on CK is due, in part, 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). In 1986, Shulman introduced PCK as a way of understanding effective instructional practices for teaching specific subject matter in ways that students can understand (Ben-Peretz, 2011; Kaya, 2009; Shulman, 1986). Since then, forms of teacher knowledge other than CK have taken priority in studies. Perhaps this is because PCK is seen as more specific to teaching, since CK is considered a more general form of content knowledge. Content knowledge, except in how it relates to PCK, was rarely studied until recently, while PCK itself is studied and reported on widely. Jüttner, Boone, Park, and Neuhaus (2013) demonstrated in biology teachers that CK and PCK could be measured using a test. Their results demonstrated a low correlation between the two, which provides evidence that the two constructs are separate. PCK is a very important construct, but its study does not replace the need for an understanding of CK, and there is still a gap in the recent literature on CK specifically.

Without strong CK, strong PCK is impossible to achieve (Kaya, 2009; Van Driel, De Jong, & Verloop, 2002). Therefore, a better understanding of CK would potentially lead to a better understanding of PCK. Van Driel et al. (2002) found that "growth of PCK was influenced mostly by the preservice teachers' teaching experiences, university-based workshops, and meetings with mentors" (as cited in 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 (Davis, 2004; Jarvis, Pell, & McKeon, 2003; Shallcross et al., 2002). Kaya (2009) found a large effect of CK on PCK in preservice 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). The effect of teacher SCK can be even more direct. In a study of Turkish 10th grade students' views of nature of science, eight 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). These results suggest that teacher CK can have a direct effect on student learning, in addition to the indirect effect achieved as part of PCK.

Some recent studies examine teacher SCK in various contexts. The general trend seems to be that elementary school teachers tend to have major gaps in their SCK, and that these gaps are a major obstacle to effective teaching (Nowicki, Sullivan-Watts, Shim, Young, & Pockalny, 2013). This is largely attributable to poor science preparation in preservice teacher programs. Clearly the construct of teacher CK broadly and SCK specifically is worth examining, but to do so we have to have effective methods of measuring CK.

Measures of Teacher CK

Teacher CK has typically been assessed using self-reports of both courses taken and perceptions of CK, and less frequently, using classroom observations. Tests of the CK of practicing teachers are rare. Each of these measures presents strengths and limitations. Although the research does not clearly demonstrate how these measures relate to each other, they are often used interchangeably as proxies for the same construct of teacher CK, usually for financial and logistical reasons. In order to better understand these constructs and their relationships, first we will examine the use of written tests, followed by self-reported knowledge, college science courses taken, and classroom observations.

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 their certification process for about a century (D'Agostino & Powers, 2009). These tests vary by state in terms of what is included, and which states require them has varied over time. A metaanalysis 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 preservice teachers and of .11 for in-service teachers (D'Agostino & Powers, 2009). 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. Certification tests also only measure the knowledge of teachers as they enter the profession, although teacher experience significantly affects teacher effectiveness (Harris & Sass, 2011) and they can learn from postgraduate experiences such as professional development. Clearly, much more research is needed on the use of written tests before conclusions about their ability to predict teacher effectiveness can be drawn.

Some recent studies administered science content tests to teachers at various stages of development. Nowicki et al. (2013) administered a science content test developed by Horizon Research, Inc. to preservice elementary education majors and inservice teachers serving as supervising teachers and compared the test scores to in-classroom knowledge as measured by observations. The test was designed for fourth, fifth, and sixth grade students, so the difficulty was similar to that of the material that would be taught. Scores were around 80% correct for both groups, but no relationship was found between the test scores and the accuracy of science information used in class with either the preservice or inservice teachers. Jüttner et al. (2013) developed and administered a CK test and a PCK test to German inservice biology teachers, but their results were primarily concerned with the validity of the tests.

Self-report questionnaires

Self-report questionnaires are viewed by researchers as the easiest method of data collection to implement, especially in large data sets (Desimone, 2009; Supovitz & Turner, 2000). However, they are most likely to contain bias favoring socially desirable responses (Dawson, Ritzhaupt, Liu, Rodriguez, & Frey, 2013; Desimone, 2009; Supovitz & Turner, 2000). Oddly, literature searches turn up very little recent research on the validity of teachers' self-reporting.

Supovitz and Turner (2000) used a content preparation scale to measure how well-prepared teachers believed they were to teach elementary-level science topics. They found that "each standard deviation of increased content preparation a teacher reported was associated with a 20% increase in the use of both investigative teaching practices and investigative classroom culture … regardless of the intensity of teachers' professional development experiences" (Supovitz & Turner, 2000, p.974).

Banilower, Heck, and Weiss (2007) studied the National Science Foundation's Local Systemic Change Through Teacher Enhancement program, including 18,657 teachers, in part examining teacher self-perceptions of their science content preparedness. In contrast to Supovitz and Turner's finding that the intensity of professional development is not important, Banilower found that the first 80 hours of professional development had a significant positive effect on teacher science content preparedness, and additional effect was seen after 160 hours of PD.

Classroom observations

Most researchers view observation as the most unbiased, if most difficult, form of data collection (Desimone, 2009). It removes self-report bias (although observer bias comes into play) and allows researchers to directly witness what happens during PD activities and in classrooms. Observation, like interviews, allows researchers to make finer distinctions in teacher practice than questionnaires. For observations to be considered valid and reliable measures of teachers' overall instruction, "three observations are required for one stable observation, and at least three stable observations over an extended period of time are required" (Desimone, 2009, p.189). The drawback to observation is that it is both time-consuming and expensive compared to other methods of measuring professional development and teaching.

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 professional development predicts lesson quality (Bowes & Banilower, 2004). As mentioned above, this is not enough observing to qualify as valid and reliable (Desimone, 2009). Hierarchical generalized linear models, with lessons nested within projects, found that 20-39 hours of professional development led to teachers earning significantly higher observation scores on lesson quality than those not receiving professional development, but more than 39 hours did not lead to a higher probability of earning a high observation score (Bowes & Banilower, 2004).

This type of quantitative observation data is very rare. It is much more common to find qualitative observation data involving 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.

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) as the observations tended to be short and infrequent. Later studies that used more rigorous research methods showed that classroom observations and teacher self-reports tend to have moderate to high correlations. These studies successfully used classroom observations to validate self-report questionnaires using concurrent validity (Mayer, 1999; Ross, McDougall, Hogaboam-Gray, & LeSage, 2003); however, their instruments examined teacher behavior, not CK.

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 (Jüttner et al., 2013). 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). According to Ball et al. (2001), 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%. 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; Harris & Sass, 2011). Harris and Sass studied the effects of teachers' preservice and in-service mathematics training on their students' achievement on standardized tests. The only course type that had a significant effect for elementary teachers was mathematics content courses, which was a negative effect. Professional development was also found to lead to either no change or a decrease in student achievement. Experience in the classroom was the only teacher factor found to improve student achievement in elementary school. Recently, Nowicki et al. (2013) found that there was no relationship between the number of college science courses taken or science GPA, respectively, and science accuracy during observed lessons.

That is not to say that teacher courses taken 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). Monk 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 leveled off to a .2% increase. The largest effects were found in courses on mathematics pedagogy, rather than mathematics content.

Monk (1994) also used the LSAY to analyze high school teachers' SCK, separating classes taken into biology, chemistry, physics, earth science, elementary science education, and secondary science education. For sophomores, teacher preparation in life sciences had no effect on student performance. For juniors, life sciences preparation showed a significant negative effect. For both sophomores and juniors, teacher preparation in physical sciences showed a significant positive relationship with student performance. Monk 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 courses. Therefore, more life science experience of the teachers meant less preparation for the junior level course material.

Rice (2005) found that while 87.3% of her 414 undergraduate elementary program students participating in her study had completed at least one biology course, only 21.3% were able to identify human, dog, worm, and spider all as animals. Of those students, 70% failed to identify the worm and 23% failed to identify the human as animals. Therefore, while college coursework can lead to higher SCK, it is common for students to finish science courses without understanding important concepts.

Method

This study utilized descriptive statistics and Pearson's correlations to examine the initial states of fifth grade teachers' SCK as measured by a science test (TEST), a self-report science knowledge scale (SKS), courses taken in college (COURSES), and observations (OBS).

Research Setting

The research was conducted in a large urban school district in the Southeast United States with a linguistically and culturally diverse student population. The ethnic makeup of the student population in the school district was 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 were eligible for free or reduced price lunch programs, 19.4% were designated as English language learner (ELL) students enrolled in English to Speakers of Other Languages (ESOL) programs, and 10.9% of students were enrolled in exceptional student education (ESE) courses. The project was created as a collaboration between a university and the school district to implement a new fifth grade science curriculum intended to maximize inquiry-based learning and understanding of science concepts by all students, especially ELL students.

Participants

Participants were recruited from schools selected using a cluster randomized control trial design where 64 schools were randomly selected from a pool of 206 available schools (not including 23 due to district monitoring and nine that participated in our previous project). As a school-wide initiative, the study involved a total of 220 fifth grade teachers from the 64 schools. There were 13 (6%) teachers who refused or were unable to participate in the data collection. There were 207 teachers who completed the questionnaire, TEST or OBS. Of these 207 teachers, 185 (89%) completed the questionnaire, TEST, and reported COURSES. These 185 teachers were used in all analyses that did not include OBS.

There were 64 teachers randomly selected for observation. However, two teachers refused and we were unable to select another teacher at the school because there was only one fifth grade science teacher due to departmentalization. Four of the 62 teachers observed did not complete the questionnaire and three did not complete TEST. In addition, four teachers were missing information on other measures. Thus, there were 54 teachers included in analyses using the OBS scales. The subset of 54 teachers was compared to the other teachers using analysis of variance (ANOVA) on the total years teaching, highest degree earned, TEST, SKS, and COURSES. The observed group of teachers was statistically significantly different from the total sample on only one measure, which was the SKS. The observed group of teachers was statistically significantly different from the total sample on only one measure, which was the SKS. The observation group scored about 0.35 points higher on this measure and the difference between the groups was small based on the effect size .06. The demographic information for all participating teachers included in the analysis is shown in Table 1.

Table 1. *Teacher Demographics (n = 185).*

Variables	Demographic Groups	*N*	%
Gender	Male	33	17.8
	Female	152	82.2
Ethnicity	Hispanic	102	55.1
	Black Nonhispanic	40	21.6
	White Nonhispanic	36	19.5
	Haitian	2	1.1
	Asian	1	.5
	Other	2	1.1
Native Lan-	English	146	78.9
	Spanish	85	45.9
	Haitian Creole	6	3.2
	French	3	1.6
	Other	3	1.6
Other Fluent*Lan-	English	37	20
	Spanish	22	11.9
	Haitian Creole	2	1.1
	French	1	0.5
	Other	5	2.7
ESOL Training*	Bachelor’s or master’s degree in ESOL	14	7.6
	ESOL endorsement through college	68	36.8
	ESOL endorsement through school dis-	98	53.0
	Grandfathered in through teaching	10	5.4
	No preparation for ESOL	10	5.4
Degrees	Bachelor’s	88	47.6
	Master’s	76	41.1
	Multiple Master’s	5	2.7
	Specialist	9	4.9
	Doctorate	2	1.1
	Other	2	1.1

* *Multiple categories could be selected.*

Data Collection

All teachers had the confidentiality of their responses explained to them prior to beginning any data collection, and any misgivings on the teachers’ parts were addressed. All teachers were provided with informed consent, and then asked to complete a background information form, questionnaire, and TEST. Data collection was conducted by trained staff in a controlled environment without a time limit.

Instruments

Teacher science knowledge test

The teacher TEST consisted of 30 items, including 24 multiple choice and six short or extended response items, with a total of 38 possible points. The test included 13 items from the student public release items of Trends in International Mathematics and Science Study (TIMSS), 11 items from the student public release items of National Assessment of Educational Programs (NAEP), and six items that were project developed. The test items addressed the topics covered in the state science content standards. The test had a combination of items of low (four items), moderate (16 items), and high (10 items) difficulty for fifth grade. The test was designed to assess how knowledgeable a teacher was on the fifth grade science content assessed in the state. Project personnel were trained to use the scoring rubric provided by the public release items to grade the short and extended response items. Each written response was scored by one research member and, when necessary, answers were discussed with the team before scores were assigned. When checking 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). Cronbach α for the scores of this scale was .80, indicating fairly strong reliability.

Teachers' self-reported science knowledge

SKS is a section of the questionnaire that consists of four Likert-type questions asking the teachers to indicate how knowledgeable they believe they are about teaching the nature of science, physical science, earth/space science, and life science at their grade level (see Appendix A). This measure was developed to align with the knowledge test since it measured the teachers' perceived knowledge of the general topics taught at their grade level, while the test used items covering these topics at the grade level. Teachers could choose any number from 1-4, with 1=not knowledgeable and 4=very knowledgeable. Cronbach α for the scores of this scale was .94, indicating strong reliability.

Classroom observations

One randomly selected teacher from each of the 64 schools was observed three times from September to January and scored on teacher knowledge of science content during a science lesson using a previously validated research-generated classroom observation guideline (see Appendix B). Three members of the research team observed the selected teachers. The observers were trained together during a one week, four hours per day, training program led by experienced former observers. The training objective was to ensure consistent scoring by learning how the OBS scales are meant to be interpreted, practicing the coding of observation notes from our previous research, and practicing observations using videotaped lessons. For this study, the OBS score was calculated as the mean of the three OBS scores for CK. When there were fewer than three observations done, the mean of the performed observations was used. Two cases had two observations, and one case had one observation completed. The mean was used to compensate for the differences that were seen from one lesson to another. Ten percent of the observations were conducted with pairs of observers and scores were compared in order to ensure observer agreements. Scores that were within one point were considered agreement. Based on this criterion, there was 95% agreement of scores between observers, and any discrepancies were discussed to make the scoring more similar in the future. Cronbach α for the scores of this scale was .76, indicating fairly strong reliability.

College science courses taken

One section of the teacher background information form asked teachers to choose how many college (both undergraduate and graduate) science courses they had taken in six categories, on a scale from zero to six or more (see Appendix C).

Data Analysis

Of the 207 teachers who provided information for the study, 185 provided complete data for TEST, SKS, and COURSES. These 185 teachers were included in the analysis. For observation, the 54 teachers out of 64 observed with data on all four measures were included.

Teacher science knowledge test

The test score for each teacher was determined by adding the total number of points earned, with 38 as the maximum points. For the multiple-choice items, one point was awarded for each correct answer. For the written response items, teachers were awarded the number of points recommended by the scoring rubric for the original source of each item, with partial credit usually available. The mean test score for all teachers was then calculated.

Self-reported science knowledge

SKS for each teacher was calculated as the average score (1-4) across the four science topics on which teachers reported how knowledgeable they felt. The mean SKS across all teachers was then calculated.

Classroom observations

OBS for each teacher was calculated as the average of the three ratings (1-5) recorded for teacher SCK during the three observations.

College science courses taken

The percent of teachers reporting taking courses in the six subject areas is presented in the results section. In addition, the number of courses taken by each teacher was calculated as the total number of courses taken. Finally, the mean number of courses taken by all teachers in the study was calculated.

Relationships between measures

Pearson correlations were computed to assess relationships between fifth grade teachers' scores on TEST, SKS, OBS, and COURSES. Note that the observations were conducted with a subset of the sample and thus the number of teachers included in analyses using OBS data is lower than that of the other analyses. In this study, correlation coefficients less than .30 are considered weak, between .30 and .50 moderate, and greater than .50 strong (Cohen, 1988).

Results

Results are reported with regard to each of the two research questions.

Teacher SCK Scores

Table 2 presents an analysis of teachers' self-reported science knowledge on the questionnaire (see Appendix A). The mean scores were similar across science content areas, ranging from 2.76

in earth and space science to 2.95 in life science, with similar standard deviations. This result indicates that teachers tended to be fairly confident in their knowledge in all science content areas.

Table 3 presents the percentage of teachers reporting different science courses taken at the undergraduate and graduate level (see Appendix C). This question was asked to get an idea of the science background of the teachers. Most teachers had taken at least one course in methods of teaching science in elementary school, though 11% reported that they had not taken a course in this area. Nearly half of the teachers reported that they had not taken any courses in physical (49%) or earth/space (50%) science, while 38% reported not having taken a course in life science.

Table 2. *Teachers' Science Knowledge across Science Content Areas (n = 185).*

Measures	Cronbach α	*M*	*SD*	*n*
All science topics (SKS)	.94	2.84	.67	185
Nature of science	--	2.87	.72	185
Physical science	--	2.79	.74	185
Earth and space science	--	2.76	.71	185
Life science	--	2.95	.73	184

Note. The responses are based on a four-point rating system (1 = not knowledgeable; 2 = somewhat knowledgeable; 3 = knowledgeable; 4 = very knowledgeable).

Table 3. *Percentage and Number (n) of Teachers Taking Each Type of Science Course (n = 185).*

Course	*0*	*1*	*2*	*3*	*4*	*5*	*6 or more*
Methods of teaching science in elementary school	11 (21)	57 (105)	16 (30)	6 (11)	4 (7)	2 (3)	4 (8)
Methods of teaching science in secondary school	80 (149)	14 (25)	3 (6)	2 (3)	1 (2)	0 (0)	0 (0)
Physical science (physics, chemistry)	49 (90)	30 (55)	10 (19)	7 (12)	2 (3)	2 (3)	2 (3)
Earth/space science (astronomy, geology)	50 (92)	31 (57)	14 (25)	4 (7)	1 (1)	0 (0)	2 (3)
Life science (biology, ecology, environmental science, marine science)	38 (71)	34 (63)	16 (30)	5 (10)	2 (4)	1 (1)	3 (6)
Other science	97 (179)	3 (5)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)

The mean scores reported for the measures of SCK are presented in Table 4. First, the test scores ranged from 9 to 38 with a mean of 30.81 out of 38 possible points, or 81.1% correct. This mean score seems reasonably low given that the test items were written at the fifth grade level. Test scores were not examined by science content areas (i.e., earth and space, life science, etc.) because there were not enough questions in each content area to provide reliable subscores, though the overall test scores were reliable. Second, the mean questionnaire score of 2.84 suggests that teachers generally believed that they were knowledgeable about the topics they taught in fifth

Table 4. *Descriptive Statistics of Teacher Science Content Knowledge Measures.*

Measure	*N*	*M*	*Min*	*Max*	*SD*	*α*
Test Score (38 max)	185	30.81	9	38	5.09	.79
SKS (1-4)	185	2.84	1	4	.67	.94
Observation (1-5)	54	3.19	1	5	.87	.75
Courses Taken	185	4.77	0	24	4.67	--

grade (see Appendix A). Third, the mean OBS score of 3.19 demonstrates that the teachers' knowledge of the lesson content was generally accurate, they did not make many scientific errors during the observed lessons, nor did they add much information that was not included in the project curriculum (see Appendix B). Finally, COURSES ranged from 0 to 24 with the mean between 4 and 5 (see Appendix C).

Relationships Between Four Measures of SCK

The results of the Pearson correlations between the four measures are presented in Table 5. The correlations between TEST and SKS (r = .30) and between TEST and OBS (r = .31) were both statistically significant. Both correlations were positive and fairly moderate in size. Additionally, the correlation between COURSES and SKS (r = .30) was statistically significant. This correlation was positive and fairly moderate in size. The correlations between COURSES and TEST (r = .01) and between COURSES and OBS (r = -.03) were not statistically significant. In addition, the correlation between OBS and SKS (r = -.14) was not statistically significant. These results indicate that increases in teachers' self-reports of science knowledge on the questionnaire were related to increases in the score they received on the science knowledge test. Additionally, increases in teachers' self-reports of knowledge from the questionnaire were related to increases in the science content presented in observed lessons. Finally, increases in the number of college science courses taken were related to increases in teachers' self-reports of SCK from the questionnaire. However, additional college science courses taken were not related to increases in the SCK test score, nor increases in the science content presented in observed lessons.

Table 5. *Intercorrelations Among Teacher Science Content Knowledge Measures.*

	TEST	SKS	OBS	COURSES
Test (TEST)	--	--	--	--
Science Knowledge Scale (SKS)	.30** (185)	--	--	--
Observation (OBS)	.31* (54)	-.14 (54)	--	--
College Science Courses (COURSES)	.01 (185)	.30** (185)	-.03 (54)	--

* $p < .05$ ** $p < .01$

Values in parentheses represent the number of teachers for each correlation.

Discussion and Implications

Despite the importance of teacher SCK to students, this construct and the proxies used to measure it are poorly understood. In addition to providing descriptive statistics for the SCK of a large sample of practicing fifth grade science teachers, this study examined relationships among measures that have not been extensively described in the literature.

Discussion

Teacher SCK scores

The results of the four different measures of teacher SCK provide information on where elementary teachers' SCK stands. The science knowledge test that was administered to the teachers covered the topics at the fifth grade level and, on average, teachers answered 80% of the questions correctly. This demonstrates that the teachers were generally knowledgeable of fifth grade science content, but did not have complete mastery of the science content even at this elementary level. The lack of previous research using a written test to assess practicing teacher SCK makes this information a valuable addition to the literature and serves as a foundation to begin to build upon.

The teachers reported that they felt generally knowledgeable of the science content, contrary to the 2000 National Survey of Science and Mathematics Education, which showed that fewer than 30% of elementary teachers surveyed felt confident in their SCK (Fulp, 2002; Rice, 2005). This may mean that teachers in our sample differed from the national population, or it may reflect a difference in the wording of the survey questions. The topic that the teachers in our study believed they were most knowledgeable in was life science, which was also an area in which most teachers reported taking more college science courses. This finding is consistent with the literature, which shows that more courses in college leads to more teacher confidence in a subject area (Shallcross et al., 2002).

The observation results demonstrated that the teachers tended to cover the information in the curriculum accurately, and tended not to deviate from the curriculum by adding correct or incorrect information. This type of quantitative observation data about teacher knowledge is very rare. Observations are often used to evaluate the fidelity of implementation of curricula and pedagogical practices of teachers (Stearns et al., 2012). However, observations may provide valuable information about teacher SCK that is distinct from the information found using more common measures, because it captures teacher SCK (and PCK) as it is revealed during teaching, as opposed to in more controlled environments.

Most of the teachers reported taking at least one methods of teaching science in elementary school course, although 11% reported never having taken this type of course. About 50% of the teachers reported not taking any course in physical science or earth/space science, and 38% reported not taking a life science course. The National Science Teachers Association (1998) recommends that elementary teachers should have taken a course in science teaching methods and one course each in the life sciences, earth/space sciences, physical sciences, and environmental sciences in order to be adequately prepared to teach science. In this study, many of the teachers took the methods course, but did not take the recommended number of courses in the different science areas. Further, in the school year prior to the study beginning, 29% of the teachers never participated in professional development activities in science or science education. This finding is consistent with a national survey involving elementary school teachers (Choy, 2006; Hill, 2009; Smith, Banilower, McMahon, & Weiss, 2002) as well as the literature indicating inadequate professional development of elementary school teachers in SCK and science teaching (Garet, Porter, Desimone, Birman, & Yoon, 2001; Kennedy, 1998; Loucks-Horsley, Hewson, Love, & Stiles, 1998). Ongoing education in science content and pedagogy is as important for teachers as university courses, which were often taken many years in the past and which rarely address the science content or the pedagogy relevant to elementary science teaching.

Relationships between four measures of SCK

Giving teachers a science knowledge test seems to be a somewhat reliable method of determining how they would perceive their own SCK and how they would portray science knowledge in their teaching. This suggests that a teacher who is better able to answer questions on a science test is also more likely to perceive their science knowledge with more confidence and demonstrate more accurate science knowledge during instruction.

Giving teachers a questionnaire to describe their SCK would serve as a fairly reliable predictor of how they would perform on a test, but not necessarily how they would teach science content in the classroom. This seems like an odd finding, due to the fact that both SKS and OBS were moderately related to TEST, but it seems that TEST does not link the two measures.

A significant positive correlation was found between COURSES and SKS. However, COURSES was not related to TEST or to OBS. This finding, consistent with the findings of Nowicki et al. (2013), suggests that taking more science courses in college leads to teachers having increased confidence in their SCK or that people who feel more knowledgeable about science enroll in more courses. Either way, science courses in college do not lead to improved performance on written tests or in transmitting information in the classroom. This supports the findings in mathematics education research, in which additional coursework by the teacher is not necessarily beneficial to the students (Ball et al., 2001), but increases teacher confidence in the subject area (Shallcross et al., 2002). The lack of correlation between test scores and college science courses taken may be partly explained by Rice's (2005) finding that preservice elementary teachers who had taken at least one biology course either did not learn or did not retain basic biology concepts past their graduation. The lack of correlation may also be explained by the fact that the test was written at a fifth grade level and teachers probably did not need extensive college science courses to do well. However, teacher confidence is in itself a positive factor for teacher attitudes toward teaching (Shallcross et al., 2002).

Overall, these findings suggest that while there is overlap in the information about teacher SCK measured by the four instruments, the overlap is far from complete as the relationships between the measures are not strong. Each measure seems to capture distinct aspects of SCK, that can potentially be put together to form a more complete picture of the construct. Interestingly, the science knowledge test seems to be the only connection between the questionnaire and the observations, yet it is the least commonly used in research. In contrast, college science courses taken seems to have weak connections to the other measures of SCK, yet it is the most commonly used in the literature. The ways that SCK has traditionally been evaluated by researchers needs to be reexamined in light of these findings.

Implications for Future Research

The results of the study make important contributions to the literature, because very few studies have measured the overall SCK of practicing teachers, and even fewer have had large sample sizes. This study measures the SCK of a large sample of practicing fifth grade teachers using four different measures of SCK. We used four measures of SCK simultaneously and compared them to show how they relate to each other. In doing so, we have created an available data set with which future researchers can compare teacher SCK of other samples. Additionally, future researchers can use our results to make methodological decisions about which measures of teacher SCK to use in their studies. Since various methods of measurement seem to measure different aspects of SCK, researchers should use more than one measure when studying SCK.

This study had several limitations worth mentioning. One limitation is a ceiling effect on the SCK test. On the test, five teachers earned perfect scores and 48 answered more than 89% of the questions correctly, leaving very little room for improvement. However, the test performs as expected, because it is measuring the ability of teachers to answer fifth grade level questions, so high scores are appropriate. Another potential limitation is another ceiling effect on SKS. Twenty-one teachers gave themselves 4 out of a possible 4 on all items making up the scale. Finally, three observations per year in the study might not be sufficient to capture variations among science topics and class activities for curriculum implementation. In addition, the number of teachers observed was a variable subsample of the larger group, decreasing the power of some analyses.

Future research would overcome these limitations. The teacher science knowledge test could have more difficult items to allow discrimination and demonstrate more change in teacher SCK. The questionnaire could add items for which teachers are less likely to choose 4, which allows for a more sensitive analysis and demonstrates more change in teacher SCK. More frequent classroom observations (than the three per year in the study) would be desirable for reliable results. In addition, the use of multiple measures could help determine how to reliably measure teacher CK. Various measures of teacher knowledge have been treated as interchangeable by researchers (Desimone, 2009), but none of the measures in this study had strong enough correlations to be treated as such. Only moderate correlations between the measures suggest that they are each capturing different aspects of teacher knowledge. Therefore, it would be beneficial for studies to use multiple measures of SCK to get a more complete picture of the construct. It would also be beneficial for studies to examine the relative strengths of the different measures as predictors of student achievement. If teachers are the most important factor in student learning, it is imperative that we do everything we can to find ways to help teachers become the best they can be.

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

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.

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

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

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

Choy, S. P. (2006). *Teacher professional development in 1999-2000*. Washington, DC : National Center for Education Statistics : U.S. Dept. of Education, Institute of Education Sciences.

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

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

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

Dawson, K., Ritzhaupt, A., Liu, F. Rodriguez, P., & Frey, C. (2013). Using TPCK as a Lens to Study the Practices of Math and Science Teachers Involved in a Year-long Technology Integration Initiative. *Journal of Computers in Mathematics and Science Teaching, 32*(4), 395-422. Retrieved October 5, 2013 from http://search.proquest.com/docview/1438894700?accountid=14585

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

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

Fulp, S. L. (2002). *2000 national survey of science and mathematics education: Status of elementaryschool science teaching.* Chapel Hill, NC: Horizon Research.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945. doi: 10.3102/00028312038004915

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics, 95*(7–8), 798-812. doi: 10.1016/j.jpubeco.2010.11.009

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. (2009). Fixing teacher professional development. *Phi Delta Kappan, 90*(7), 470-476.

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

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

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

Kennedy, M. M. (1998). Education reform and subject matter knowledge. *Journal of Research in Science Teaching, 35*(3), 249-263. doi: 10.1002/(SICI)1098-2736(199803)35:3<249::AID-TEA2>3.0.CO;2-R

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing professional development for teachers of science and mathematics*. National Inst. for Science Education, Madison, WI: Corwin Press, Inc.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), pp. 29-45. Retrieved from http://www.jstor.org/stable/1164545

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

Nowicki, B. L., Sullivan-Watts, B., Shim, M. K., Young, B., Pockalny, R. (2013). Factors Influencing Science Content Accuracy in Elementary Inquiry Science Lessons. *Research in Science Education, 43*(3), 1135-1154. doi: 10.1007/s11165-012-9303-4

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

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

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

Ross, J. A., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education, 34*(4), 344.

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
Smith, P. S., Banilower, E. R., McMahon, K. C., & Weiss, I. R. (2002). *The national survey of science and mathematics education: Trends from 1977 – 2000.* Chapel Hill, NC: Horizon Research, Inc.
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., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*(9), 963-980. doi: 10.1002/1098-2736(200011)37:9<963::AID-TEA6>3.0.CO;2-0
Van Driel, J. H., De Jong, O., & Verloop, N. (2002). The development of preservice chemistry teachers' pedagogical content knowledge. *Science Education, 86*(4), 572. Retrieved from http://iiiprxy.library.miami.edu:4420/login.aspx?direct=true&db=aph&AN=7184184&site=ehost-live

Corresponding Author:
Brandon S. Diamond
b.diamond@bio.miami.edu
Department of Biology, Cox Science Center
University of Miami
1301 Memorial Dr.
Coral Gables, FL 33146
(786) 408-4762

Appendix A

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

	Not knowledge-able	*Somewhat knowledge-able*	*Knowledge-able*	*Very knowledge-able*
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

Appendix B

Classroom Observation Guideline

Scale: Teacher Knowledge of Science Content

How accurate and comprehensive is the teacher's mastery of the science content of the lesson?

This scale indicates the extent to which the teacher has an accurate and comprehensive grasp of the science content of the lesson. While elementary grade teachers are not expected to match the degree of mastery that a scientist or other specialist would have in the field, they should possess accurate information about the topic they are teaching. Their mastery of the content should be at least slightly above that expected of students upon successful completion of the lesson. Teachers should be able to answer students' questions or at least indicate to students how one might go about finding out the answer and/or what factors limit the possibilities for doing so. Responding "I don't know" is preferable to offering incorrect information, but such a response should be accompanied by suggestions (or asking students for suggestions) of how students and the teacher might find out more.

A high score on this scale would be characterized by the teacher responding to students' questions with relevant information, enriching the lesson by providing deeper knowledge of the phenomena, or linking it to other phenomena or experiences known to students.

A low score would be characterized by shallow or tenuous knowledge of science content. Uncertainties are not pursued with students as potential paths toward deeper understanding of the topic. Furthermore, the teacher transmits inaccuracies in the information presented to students, or makes statements that indicate a fundamental misunderstanding of the facts or processes involved.

This scale focuses more on teacher behavior than on students. As with all of the scales, however, the interaction between teacher and students is the focus of observation; in this case, how the teacher's mastery of the content affects the information students receive and the teacher's ability to promote students' inquiry practices. Thus, more extensive transmission of knowledge from the teacher to students is not always better. The teacher's mastery of the content should not give way to long monologues that are too advanced for students to grasp, or that impede them from carrying out inquiry practices.

Rating Scale – Teacher Knowledge of Science Content

1. The teacher transmits multiple major inaccuracies (i.e., more than two) in the information presented to students, or makes statements that indicate a fundamental misunderstanding of the facts or processes involved.

2. The teacher transmits 1-2 major scientific inaccuracies during the lesson. His/her grasp of the science content is generally accurate, but shallow and/or tenuous. Uncertainties are not pursued with students as potential paths toward deeper understanding of the topic.

3. The teacher's knowledge of the lesson content is accurate. However, further queries by students, if they arise, are met with responses of "I don't know" or "That's not part of the lesson," with no discussion of how one might investigate further.

4. The teacher's knowledge of the lesson content is accurate. A couple of times during the lesson, the teacher responds to students' questions with appropriate information, OR offers resources for finding further information , OR spontaneously provides examples or analogies to further illustrate the content.

5. The teacher demonstrates knowledge of the topic that goes beyond the merely adequate, enriching the discussion with "extra" information at key points in the lesson. He/she enriches the lesson by providing deeper knowledge of the phenomena. He/she is able to link the topic to other phenomena or experiences known to students in accurate and relevant ways, allowing for deeper discussion.

Appendix C

Science Background

Please indicate the number of courses that you have taken at the undergraduate and/or graduate level in each of the following areas:

	0	*1*	*2*	*3*	*4*	*5*	*6 or more*
a. Methods of teaching science in elementary school	0	1	2	3	4	5	6
b. Methods of teaching science in secondary school	0	1	2	3	4	5	6
c. Physical science (physics, chemistry)	0	1	2	3	4	5	6
d. Earth/space science (astronomy, geology)	0	1	2	3	4	5	6
e. Life science (biology, ecology, environmental science, marine science)	0	1	2	3	4	5	6
f. Other (specify) ___________________	0	1	2	3	4	5	6