

The Influence of Teachers' Knowledge on Student Learning in Middle School Physical Science Classrooms

Philip M. Sadler
Gerhard Sonnert
Harold P. Coyle
Nancy Cook-Smith
Jaimie L. Miller

Harvard-Smithsonian Center for Astrophysics

This study examines the relationship between teacher knowledge and student learning for 9,556 students of 181 middle school physical science teachers. Assessment instruments based on the National Science Education Standards with 20 items in common were administered several times during the school year to both students and their teachers. For items that had a very popular wrong answer, the teachers who could identify this misconception had larger classroom gains, much larger than if the teachers knew only the correct answer. On items on which students did not exhibit

PHILIP M. SADLER is the director of the Science Education Department at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138; e-mail: psadler@cfa.harvard.edu. His research program includes assessment of students' scientific misconceptions, the transition to college of students who wish to pursue STEM careers, and the enhancement of the skills of teachers of science.

GERHARD SONNERT is a research associate at the Harvard-Smithsonian Center for Astrophysics; e-mail: gsonnert@cfa.harvard.edu. His research interests include gender in science, the sociology and history of science, and science education.

HAROLD P. COYLE is a project manager at the Harvard-Smithsonian Center for Astrophysics; e-mail: hcoyle@cfa.harvard.edu. His primary research interests are identifying the misconceptions that most significantly impede learning by K-12 students and teachers in the earth and physical sciences, and determining the most effective teaching strategies to address these misconceptions.

NANCY COOK-SMITH is a project psychometrician at the Harvard-Smithsonian Center for Astrophysics; e-mail: ncook@cfa.harvard.edu. Her research interests include test construction and validation in the sciences.

JAIMIE L. MILLER is a testing coordinator at the Harvard-Smithsonian Center for Astrophysics; e-mail: jlmiller@cfa.harvard.edu. Her research interests include the achievement gap in science and engineering between the genders and racial/ethnic groups.

misconceptions, teacher subject matter knowledge alone accounted for higher student gains. This finding suggests that a teacher's ability to identify students' most common wrong answer on multiple-choice items, a form of pedagogical content knowledge, is an additional measure of science teacher competence.

KEYWORDS: subject matter knowledge, pedagogical content knowledge, teacher, science education, misconceptions

Everybody wants teachers to be knowledgeable. Yet there is little agreement on exactly what kinds of knowledge are most important for teachers to possess. Should a teacher have a deep knowledge of the subject matter, gleaned from college study, additional graduate courses, or even research experience? Or is it better if the teacher has an understanding of what students think? Is there some optimal combination of different types of knowledge? Discussions of such issues, if they make use of data at all, are often based on indirect methods of gauging teacher knowledge. College degrees, courses taken, and grades achieved often serve as proxies for a teacher's subject matter knowledge (SMK). Teachers' awareness of the prior knowledge of students is harder to assess and is often revealed by the choices that teachers make in what to cover and how to cover a topic, which requires the time and judgment of a skilled observer to evaluate. Moreover, studies that rigorously investigate the relationship between the different kinds of teacher knowledge and student gains in understanding of science are rare (Baumert et al., 2010).

Beliefs about teacher knowledge shape both the policies regulating how teachers are prepared, certified, hired, and evaluated as well as programs that provide ongoing professional development for practicing teachers. Recent increases in funding for federal programs that prepare and enhance the abilities of mathematics and science teachers have been made with the objective of boosting our country's economic competitiveness.¹ The public investment in increasing teacher knowledge is certainly well spent if substantive gains in student achievement result but is poorly spent if improvements in the kinds of teacher knowledge promoted have little to do with student outcomes.

Our study applies the rarely used method of administering identical assessment items to both teachers and their own students. Developed to align with the National Science Education Standards (National Research Council [NRC], 1996), these five-option multiple-choice test items reflect advances by cognitive science in that many items require a choice between accepted scientific concepts and misconceptions that have been well documented in the science education literature (Sadler, 1998; Schoon, 1988; Treagust, 1986).

Relevant Research

The knowledge that teachers must possess to be effective has historically been a topic of scholarly interest. A review of the existing literature reveals no shortage of opinions and philosophical essays concerning the kinds of knowledge that are essential to good teaching. However, rigorous empirical studies are few. Wilson, Floden, and Ferini-Mundy (2002) claimed that studies of teacher effectiveness too commonly rely on proxies for teacher SMK (e.g., college major, courses taken), teacher self-reports, test scores from overly broad exams (e.g., National Teachers Examination), and area of teacher certification (for which requirements vary by state). Such measures of teacher competence historically have been found to be poor predictors of student achievement, particularly standardized exam scores (D'Agostino & Powers, 2009; Mitchell, 1985) and affective traits (Haney, Madaus, & Kreitzer, 1987; Hawley & Rosenholtz, 1984). We use measures more directly related to the specific knowledge that a teacher needs to effectively teach a specific subject at a particular grade level. One should keep in mind that these are not the only aspects of teacher knowledge that may be important, but they are measures that, despite their highly plausible relation to student gains, have only seldom been studied.

Subject Matter Knowledge (SMK)

As Ball (1991a) succinctly stated, “Teachers cannot help children learn things they themselves do not understand” (p. 5). SMK is defined as the general conceptual understanding of a subject area possessed by a teacher, which is obtained by completing the required coursework (Shulman, 1986). While No Child Left Behind unleashed an unprecedented wave of tests of students’ knowledge, there has been relatively little testing of teachers, save at the very start of their careers. Attempts to directly measure the SMK most needed to teach a particular course for a particular age level are uncommon; instead, researchers and administrators examine teachers’ backgrounds to quantify their education (coursework, grades, degrees) or certification (number of certifications, certified subject areas) or use standardized test scores (see, e.g., Boardman, Davis, & Sanday, 1977; Ferguson, 1991; Greenwald, Hedges, & Laine, 1996a, 1996b; Hanushek, 1972, 1996; Harbison & Hanushek, 1992; Mullens, Murnane, & Willett, 1996; Rowan, Chiang, & Miller, 1997; Strauss & Sawyer, 1986; Tatto, Nielsen, Cummings, Kularatna, & Dharmadasa, 1993). While these measures all may be related to teachers’ SMK for courses they teach, they are still proxy variables. They do not directly test for understanding of the particular science concepts, facts, and skills that teachers are charged with conveying to students in a specific science course. A direct measurement would require that teachers take and perform well on tests tailored to a particular course’s content—for instance, on the same tests that are given to their students.

Within the educational research community, studies have examined how individual teachers differ in their SMK (using concept maps: Hoz, Tomer, & Tamir, 1990; Shymansky et al., 2006). Other studies have found that tasks requiring structuring of SMK are particularly difficult for novice teachers (Lederman, Gess-Newsome, & Latz, 1994) and that SMK can increase over time (Arzi & White, 2008). In related research, studies have shown that teaching mathematics requires specialized knowledge that the average adult would not have (Ball, 1988, 1990, 1991a; Borko et al., 1992; Leinhardt & Smith, 1985). Comparison studies have examined differences between the SMK of U.S. teachers and those of other countries (Harbison & Hanushek, 1992; Ma, 1999), and of new teachers and veterans (Ball, 1991b; Baturu & Nason, 1996; Brickhouse & Bodner, 1992; Clermont, Borko, & Krajcik, 1994; Czerniak & Lumpe, 1996; Even, 1993; Leinhardt & Smith, 1985), while other studies have focused on new teachers alone (Ball, 1990). Whereas these studies are informative in fleshing out aspects of SMK, there are few tools for easily quantifying the amount of teacher SMK, nor has the relationship between science teachers' SMK and their students' learning gains been established.

We must look outside of the United States to find studies in which teachers take the same tests as students to ascertain SMK. The performance of third grade teachers in Belize on a primary school mathematics test was associated with their own students' yearly mathematics gains on a similar test (Mullens et al., 1996). In Brazil, Harbison and Hanushek (1992) found that teacher scores on the same fourth grade mathematics test taken by their students were far from perfect, but a significant predictor of their students' achievement. In the United States, Lockheed and Longford (1991) found that proxy measures of SMK—teachers' years of experience and postsecondary coursework—were unrelated to student learning gains; this result may have been caused by these proxy variables being poor substitutes for the knowledge and skills that teachers need to help students learn (Hill, Rowan, & Ball, 2005). Byrne's (1983) review of 30 studies relating teachers' scores on tests of SMK to student achievement found mixed results. He noted that because there was so little variation in teachers' test scores, the fact that results were not statistically significant should not be surprising. Another explanation is that SMK impact on student learning diminishes beyond a basic level of teacher knowledge (Darling-Hammond, 2000).

The most promising work in measuring teacher knowledge has been carried out in the field of mathematics by a research group led by Deborah Ball and Heather Hill, who created a sophisticated tool for measuring SMK for elementary level mathematics and then used this instrument in a larger study (Ball & Bass, 2000, 2003). They reported on the specific "mathematical knowledge used in teaching" (Hill et al., 2005, p. 377), as measured by a survey filled out by 699 first and third grade teachers. Using a common standardized test to gauge student gains in mathematics, they found that

teachers with more of this kind of knowledge had significantly larger student gains in their classrooms. However, the effects of this variable were small (0.05 *SD* student gain for each 1.0 *SD* increase in teacher knowledge). Quantitative, empirical large-scale evidence that teacher SMK influences student learning gains in science is conspicuously absent from the literature, as is a simple, quantifiable direct measurement of science teacher SMK (Wayne & Youngs, 2003). The data we collected provide the opportunity to investigate the results of teachers answering the same test items as their students because an item-level analysis can link the particular knowledge that a teacher possesses to any student gains on that particular item.

Knowledge of Student Misconceptions (KOSM)

The ideas that students bring to the science classroom are well documented in the research literature across scientific domains (Driver, Squires, Rushworth, & Wood-Robinson, 1994).² This misconception research has spanned qualitative and quantitative methods (Wandersee, Mintzes, & Novak, 1994), using one-on-one interviews (Keuthe, 1963; Nussbaum & Novak, 1976; Piaget & Inhelder, 1929), open-ended written instruments (Freyberg & Osborne, 1985), multiple-choice tests (Halloun & Hestenes, 1985), two-tiered multiple-choice tests with written justifications (Treagust, 1986; Tsai & Chou, 2002), and large-scale multiple-choice tests (as used in this study).

A teacher's knowledge of the common student misconceptions that make learning a concept difficult is hypothesized to be crucial to effective teaching (Ausubel, Novak, & Hanesian, 1978). While some researchers advocate that teachers should know common student misconceptions for the topics that they teach (Carlsen, 1999; Loughran, Berry, & Mulhall, 2006), others advocate that teachers should develop interviewing skills (Duckworth, 1987) or tests (Treagust, 1986) to reveal student preconceptions in their classrooms. Yet the research literature falls short in assessing science teachers' knowledge of particular student misconceptions and the impact of this knowledge on student learning.

KOSM is a part of Shulman's (1986) construct of pedagogical content knowledge (PCK), which he defines as "the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations" (p. 9). Shulman describes the importance of a teacher's knowledge of

the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons. If those preconceptions are misconceptions, which they so often are, teachers need knowledge of the strategies most likely to be fruitful in reorganizing the understanding of learners, because those learners are unlikely to appear before them as blank slates. (pp. 9–10)

Such a view recognizes that learning science is as much about unlearning old ideas as it is about learning new ones. Learners struggle to change their misconceptions, ideas that make sense to them.

Around the same time that Shulman created the term PCK, Ball (1988) and Grossman (1990) sought to expand the concepts of teacher SMK to encompass not just knowing the subject but also knowing the subject matter for teaching. Hill et al. (2005) included in their "content knowledge for teaching" (p. 387) a teacher's knowledge of "learners' typical errors and misconceptions" (Hill, Schilling, & Ball, 2004, pp. 12–13). Grossman (1990) found a need for teachers to examine concepts from the perspective of students, paying particular attention to "potential student difficulties" (p. 59).

Study Goals

This study assesses teacher SMK and the knowledge of students' misconceptions component of PCK (referred to as KOSM in this article) in the context of the key concepts defined by the national standards and measures their relationship to student learning. It bears repeating that the concept of PCK, as defined by Shulman, is multifaceted and includes more elements than only misconception knowledge. To accomplish our task, we employed the novel approach of administering the same multiple-choice items, which were designed to test the NRC standards, to both students and teachers. This method allowed us to simultaneously evaluate the teachers' SMK and KOSM and to examine if these teacher measures predict student gains in middle school physical science classrooms.

We realize that many educators question the value of tests composed of multiple-choice items. However, when items are written to include popular misconceptions as distractors, they function well in diagnosing misconceptions that impede the learning of science concepts (Sadler, 1998), as first suggested by Treagust (1986). Science learners often struggle with misconceptions, which are well documented in the science education literature, but remain difficult to alter. Good examples concern the causes of the seasons and of the phases of the moon. In a popular video *A Private Universe* (Schneps & Sadler, 1987), bright and articulate graduating college seniors, some with science majors, revealed their misunderstandings of these common middle school science topics. If teachers hold such misconceptions themselves or simply are unaware that their students have such ideas, their attempts at teaching important concepts may be compromised. Simple tests that teachers could take to ascertain their mastery level of certain elements of SMK and KOSM would be a boon for teachers' self-assessment. Such tests could also serve as an aid in certification or for the evaluation of professional development (PD) experiences.

The goal of this article is to test two hypotheses regarding teacher knowledge in middle school physical science courses:

1. Teachers' knowledge of a particular science concept that they are teaching predicts student gains on that concept.
2. Teachers' knowledge of the common student misconceptions related to a particular science concept that they are teaching predicts student gains on that concept.

We measured gains on key concepts by assessing students several times during a 1-year middle school physical science course. Numerous control variables were employed to account for differences between students that could be expected to affect their learning. Using a hierarchical logistic regression model, we assess, at the level of individual test items, how strongly teachers' SMK and KOSM are associated with student gain.

Sample

Recruiting a nationally representative sample of middle school physical science teachers who would both administer tests to their students and fill out assessments themselves proved to be a daunting task. Ours was a "low-stakes" test, neither counted for academic credit nor required by the state. In this light, dedicating three additional class meetings during the school year to testing could appear burdensome if no benefit would accrue to the teacher or the students. While we could offer no financial incentive to participating teachers, our recruitment materials explained the unusual nature of our test, in both covering all of the concepts contained in the relevant NRC standards and measuring common misconceptions. Offering to report back to teachers the aggregate scores of their students and associated student gains in comparison with our national sample proved attractive to many. We reported these data to the teachers at the end of the school year so that the information would not affect teachers' efforts during the study period.

As a part of a larger recruitment effort, 620 teachers of seventh and eighth grade physical science at 589 schools responded to a blanket, nationwide direct mailing to middle school science teachers. Of schools represented, 91% were public, 4% were private non-Catholic religious schools, 3% were private nonreligious schools, and 2% were private Catholic schools. All 620 teachers were sent pretest forms for their classrooms; 219 teachers returned 24,654 physical science pretests and became part of the study. Of these participants, 181 teachers also returned either midyear tests or posttests (or both) in addition to their pretests. Students' birth date (MMDDYY) and gender were used for matching student tests in each classroom (we did not collect student names to keep students anonymous), allowing gains in knowledge to be calculated by individual student. In all, roughly half of the starting students could be matched with one other test taken, for a total of 12,642 participants (with reduction in participants resulting from student illness, other absences, noncompliance, changing schools, recording errors,

blanks, etc.). Eighth grade students made up 78% of our sample; 22% were seventh graders. Of these students, 75% submitted complete demographic information, with all questions answered. In all, 9,556 participants had sufficient information (i.e., two or three tests and demographics) and could be used for the statistical analysis presented here.

Teachers volunteering to be part of this study were quite experienced, with a mean time teaching of 15.6 years ($SD = 7.0$) and a mean time teaching middle school physical science of 10.4 years ($SD = 7.8$). The teachers had a range of undergraduate preparation: 17% with a degree in the physical sciences, 25% with a degree in another science, 36% with a science education degree, 23% with a nonscience education degree, and 9% with a degree from another field. Multiple undergraduate degrees were held by 8% of teachers. Of the total sample, 56% held a graduate degree in education and 14% held a graduate degree in science.

One concern in studying classrooms of teachers who were not randomly selected but had volunteered was that their students may have borne little resemblance to the population of students who take middle school physical science. To understand and account for differences in student demographics, we asked students for their race/ethnicity, home language spoken, and parents' education. Comparative statistics are in Table 1.

Table 1 shows our sample characteristics alongside the appropriate national statistics. Compared with K–8 students in the U.S. population, our sample appears to underrepresent Black and Hispanic students and overrepresent students with parents with college degrees (Aud et al., 2012, pp. 140, 148, 163). More classroom teachers have degrees in the appropriate field (physical science or science education), compared with the national statistics (National Science Board, 2008). A larger fraction of public schools participated than represented nationally (Snyder, 2012, Table 5). Thus, the sample cannot be considered fully representative of the national population. However, its large size likely captured the range (albeit not in population proportions) of the existing variation in the relevant variables so that they could be controlled for in a hierarchical model. Inferences about the relationship between dependent and independent variables can still be made. Further note that the alternative methods of studying teacher knowledge that require a much larger amount of effort, time, and money are typically, and almost necessarily, restrained to much smaller samples.

Study Design

Physical science curricula vary in time allocation during the middle school years. While some schools devote an entire academic year to the subject, other schools include physical science within a general science sequence that includes earth and space science and life science. The design allowed an interval shorter than an entire school year for pre-post data

Table 1
Comparison of Survey Sample to National Statistics

Level	Group		Study Sample (%)	National (%)
Student	Race/ethnicity	White	62	54
		Black/African American	10	15
		Asian/Pacific Islander	5	4
		Hispanic	14	23
		Other/multiracial	9	4
	Language spoken at home	English	75	79
		Other than English	25	21
	Highest parent education level	< high school	6	11
		High school diploma	21	23
		< 4 years of college	15	29
Teacher	Undergraduate degree	≥ 4 years of college	60	39
		Physical science or science education (in field)	53	37
		Other science	25	35
	Graduate degree	Nonscience education	23	24
		Other field	8	4
School	Control	Education or science	70	46
		Public	91	75
		Private	9	25

collection from those teaching physical science for a shorter period. While all participating teachers returned pretests at the beginning of the school year, some returned only the midyear set of tests, while others returned only the year-end test and another group returned all three (pretest, midyear test, year-end test). The study included all classes that returned a minimum of two sets of tests. In this data set, 34% of teachers chose to participate for only one semester. These classrooms contain 25% of the students in the sample. In the 75% of the classrooms participating for two semesters, students who took two tests account for 36% of the sample, and those who took all three tests account for 39% of the sample.

The use of this time-series testing approach had an additional advantage over other design options. We were concerned that the initial science achievement of participating classrooms might obscure any changes in student achievement during the school year. For example, it may be that, compared to their less experienced colleagues, more experienced or expert teachers are assigned students who have shown higher prior achievement.

The pre-post design enabled us to control for the students' initial knowledge level.

Instrument

The authors developed the study instrument for a National Science Foundation–funded project that sought to construct multiple-choice items that reflect the content of the NRC Grades 5–8 physical science standards.³ This was an extensive development project, with the goal of producing a set of assessment instruments to be used for diagnostic purposes in middle school classrooms teaching physical science. Intending to generate a valid short test, the project staff first created a 110-item test bank. Three pilot tests were developed to identify 6 well-performing items that could be included as anchors on all future tests. Six field tests were then constructed to measure the parameters of all developed items and were administered to 6,994 students of 85 teachers in 31 states. (Teachers were recruited from a nationwide mailing similar to that used for this study, but in smaller numbers.) Results from the field test were used to create pretests, midyear tests, and year-end tests of 31 items each. A total of 20 items were common to all three final tests—and these are the 20 items that we use for analysis in this article, because our analysis is at the item level. These final tests included well-performing items with a range of difficulties (for both teachers and students) and high item discrimination to comprehensively measure the range of physical science concepts at the middle school level as defined by the NRC science content standards. While we are constrained from publishing the actual item wording on this standardized instrument because it is in wide use by PD evaluators nationally, we can provide a list of the concepts addressed.⁴

NRC Standard I: Properties and Changes in Properties of Matter: A substance has characteristic properties: fixed boiling point (1 item). Substances react chemically in characteristic ways with other substances to form new substances: chemical reactions can produce invisible gases (2 items), mass is conserved in chemical reactions (2 items). An element can vary in outward appearance. An element can exist as a solid, liquid, or gas (1 item).

NRC Standard II: Motions and Forces: The motion of an object can be represented in a variety of ways including position versus time, velocity versus time (2 item). An object's position, direction of motion, and speed are interrelated (1 item). Unbalanced forces will cause change in the speed or direction of an object's motion (2 items).

NRC Standard III: Transfer of Energy: Energy is conserved and can do work (2 items). Heat flows from higher temperature objects to lower temperature objects (3 items). Light travels in a straight line at a constant speed until it interacts with matter (1 item). Electrical circuits provide a means of transferring electrical energy when heat, light, sound, and chemical changes are produced (2 items).

Energy can be transferred into or out of a system in a variety of forms, including heat, light, mechanical motion, and electricity (1 item).

Earlier analysis had determined that content questions fell into two categories with respect to the relative popularity of the wrong answers. Eight questions were classified as having “weak” or no evident misconceptions, with the most common wrong answer chosen by fewer than half of the students who gave incorrect responses. The results for Master Item 38 are given as an example. While 38% of students answered this question correctly (Option d, underlined), a corresponding 62% answered incorrectly; 26% of all students responded with Option b, and thus 42% (i.e., 26%/62%) of the incorrect responses were Option b. While b was the most popular wrong answer, it does not meet the criterion of more than half of the students who answer incorrectly choosing it. Hence, the item is considered not to have an identifiable misconception.

38. A scientist is doing experiments with mercury. He heats up some mercury until it turns into a gas. Which of the following do you agree with most?

- a. The mercury changes into air. 12%
- b. Some of the mercury changes into carbon dioxide. 26%
- c. The mercury changes into steam. 14%
- d. The gas is still mercury. 38%
- e. The mercury is completely destroyed when heated. 10%

A total of 12 questions were classified as having “strong” misconceptions, with 50% or more students who chose a wrong answer preferring one particular distractor. The example given is Master Item 13, in which only 17% of students answered the question correctly (Option a, underlined) and a corresponding 83% answered incorrectly. A very large fraction (59%) of students chose one particular wrong answer, d. Hence, of the students choosing an incorrect answer, 71% (i.e., 59%/83%) preferred this single distractor. This response indicates a strong misconception, which is generally the most popular wrong answer in all teachers’ classrooms.

13. Eric is watching a burning candle very carefully. After all of the candle has burned, he wonders what happened to the wax. He has a number of ideas; which one do you agree with most?

- a. The candle wax has turned into invisible gases. 17%
- b. The candle wax is invisible and still in the air. 6%
- c. The candle wax has been completely destroyed after burning. 8%
- d. All of the wax has melted and dripped to the bottom of the candle holder. 59%
- e. The candle wax has turned into energy. 10%

Influence of Teachers' Knowledge on Student Learning

The Kuder–Richardson 20 (KR-20) scores for the common 20-item component of the pretests, midyear tests, and posttests are, respectively, 0.53, 0.64, and 0.71. As expected, reliability increased with later administration of each test because student knowledge increased. This is within the range of acceptable internal consistency for tests with multiple latent variables—in this case, the three different NRC Grades 5–8 physical science standards—in which performance between standards may not be highly correlated. Classroom coverage of the content represented by the test items was near universal. Only eight teachers reported that they did not cover the content tested by one particular item, and two teachers reported that they did not cover the content in two items.

The concepts addressed by the 20 test items that appeared on all three administrations are broken down by standard. Common misconceptions are noted in italics with a citation to a relevant early study:

I. Properties and Changes of Properties of Matter.

- a. A substance has characteristic properties; *boiling point varies with the amount of material (Andersson, 1980).*
- b. Substances react chemically in characteristic ways with other substances to form new substances; *burning produces no invisible gases (BouJaoude, 1991).*
- c. All substances are composed of one or more elements, *matter is not conserved (Driver, 1985).*

II. Motions and Forces

- a. Position can be used to represent an object's motion; *objects that are speeding up cover the same distance per unit time (Mori, Kojima, & Deno, 1976).*
- b. An object's position, direction of motion, and speed are interrelated; *graphs of motion versus time are similar to physical path followed by the object (McDermott, Rosenquist, & van Zee, 1987).*
- c. Forces can act in the direction opposite to an object's motion; *force is always in the direction of an object's motion (Clement, 1982).*

III. Transfer of Energy

- a. Objects come to the temperature of their surroundings; *some materials are intrinsically cold (Grimellini-Tomasini & Pecori Balanda, 1987).*
- b. Light propagates and interacts with matter and it is passively detected; *light travels in a straight line even when it interacts with matter (Huang & Chiu, 1993).*
- c. Electrical circuits provide a means of transferring electrical energy when heat, light, sound, and chemical changes are produced; *electricity behaves in the same way as a fluid (Pruem, 1985).*

Inferential Modeling

The primary goal of this study is to determine the relationship between teacher knowledge and student learning. The dependent variable in the statistical models that we estimate for this purpose is student posttest item score. If we have a year-end score, it serves as the posttest score; if we have only a midyear score, that score is used instead. (The timing of the posttest is accounted for through a control variable.) Our data are hierarchical: Students are grouped within teachers' classrooms, and for each student we have more than one score to predict. Hence, the basic hierarchical structure of our statistical models is scores within students within classrooms. To control for factors that might affect student learning, we include a range of student demographic and background variables (grade, gender, highest level of parental education, race, half or full year between pre- and posttest) in the model, along with students' performance on the math and reading items, and the student pretest score as a baseline for student knowledge. All analyses separate out two kinds of test items, those items without a dominant misconception (non-misconception items) and those items where more than half of incorrect students chose one single wrong answer (misconception items). Our key independent variables are teacher SMK and teacher KOSM scores because we are primarily interested in how teacher performance predicts student performance.

The "grain size" of this analysis is at the item level, which probes the effect of teachers' SMK and KOSM about particular scientific concepts. Analysis of student learning at the item level is quite rare in the education literature. A probable reason for this is that, in most studies, researchers are interested in the students' overall improvement, based on certain teacher qualities or decisions (e.g., use of labs, amount of homework). Each item in an assessment is seen as contributing to an estimate of some underlying latent variable. However, we are interested in accounting for the impact of teachers having any specific "holes" in knowledge, be they in SMK or KOSM. Item-level analysis is sensitive to particular differences in teacher knowledge. The item-level analysis uses each teacher's individual SMK and KOSM scores for each item to model the score of each of their students on each test item. The item-level analysis is able to account for both the teachers' knowledge of each item and each item's difficulty level. A test-level analysis aggregates this information and does not account for differences between teachers in their knowledge of individual items.

To carry out a conventional test-level analysis, a hierarchical linear regression model would typically be used because the outcome variable, student posttest total score, is normally distributed (skewness = 0.500, kurtosis = -0.0249). However, the item-level analysis requires a less common statistical method because the student posttest item score can have only one of two values, either incorrect or correct (0 or 1). Hence, a hierarchical logistic

regression model is appropriate (Wong & Mason, 1985), with 20 item scores predicted for each student rather than a posttest total for each of two kinds of items, misconception (12 total) and non-misconception (8 total). The hierarchical logistic model was implemented through PROC GLIMMIX of the 9.2 release of the SAS statistical package. For this model, we experimented with including "item" as a random factor in addition to "classroom" and "student nested within classroom." However, models of this type failed to compute (even on the Harvard-MIT Data Center cluster). As an alternative, we included the variable "item difficulty" (the overall proportion of correct responses to each item) in the hierarchical logistic models.

Establishing how well a hierarchical logistic model fits the data is rather tricky. While for ordinary least square multiple regression, the R^2 statistic serves as the widely accepted measure for the goodness of fit of the estimated model (intuitively interpretable as the proportion of variance in the dependent variable that is explained by the model), measuring goodness of fit becomes more complicated for hierarchical models that partition the overall variance (Singer & Willett, 2003). The situation is even more opaque when it comes to logistic regression (let alone hierarchical logistic regression). Here, in the logistic realm, statisticians have developed a whole range of goodness-of-fit measures, and debate continues about their relative strengths and weaknesses (Allen & Le, 2008; Hosmer & Lemeshow, 1989; Menard, 2000). We chose a pseudo- R^2 that consists of the squared correlation between the observed and the predicted values (Singer & Willett, 2003, p. 102). Because our interest focuses on teachers' knowledge of individual items, we took, for each teacher, the observed and predicted classroom posttest scores for each item and then squared their correlation. Summary results are expressed as an effect size (gain in units of standard deviation of the 20-item pretest score) so that the effects of the levels of teacher knowledge can be compared. The levels of teacher knowledge for each item are captured by simple dummy variables: Teachers either have the SMK for each item or not, corresponding to a 0 or 1 value, and they have the KOSM for each misconception item or not, also corresponding to a 0 or 1 value.

Results

To "get a feel" for the data collected, we first calculated descriptive statistics, particularly of how students and teachers performed on each test item.

Descriptive Statistics

Student scores on these assessments were relatively low, indicating that the items were difficult. The mean prescore across all items (both those without misconceptions and those with misconceptions) was 37.7%. Scores on the final administration of the test were higher, 44.8% for items without

Table 2
Student Performance on Pretest and Posttest

	Non-Misconception Items				Misconception Items			
	Pre	Post	Gain	ES	Pre	Post	Gain	ES
<i>M</i>	0.378	0.448	0.071	0.380	0.376	0.417	0.042	0.278
<i>SD</i>	0.186	0.200	0.200	1.075	0.150	0.169	0.170	1.133
<i>SE</i>	0.002	0.002	0.002	0.010	0.001	0.002	0.002	0.010

Note. $N = 9,556$ students. ES = effect size. Mean pretest scores for non-misconception and misconception items are almost identical, but gains are larger for non-misconception items. Effect size is gain in units of standard deviation of the pretest.

misconceptions and 41.7% for those with misconceptions. Such posttest results are not uncommon when test items include popular misconceptions as distractors and teachers use traditional instructional methods (Hake, 1998). For example, Tekkaya (2003) found that middle school science students' posttest scores on a diffusion and osmosis task employing misconceptions in an experimental group were 54.1% correct, while a control group attained only 38.7% (pretest scores were 22.5% and 19.1%, respectively). Baser (2006) found that preservice second grade teachers exposed to a cognitive-conflict physics curriculum scored 60.7% correct on a posttest dealing with understanding heat and temperature, while a control group attained only 40.9% (pretest scores were 32.2% and 30.2%, respectively). While gains appear relatively small (Table 2), it is more useful to express them as effect size, or gain calculated in units of standard deviation of the pretest score (Cohen, 1969). For non-misconception items, the effect size was 0.380 *SD*, while the effect size for misconception items was smaller: 0.278 *SD*. We conclude that students had an easier time learning the content for which there appeared to be no dominant misconception.

Items used to gauge each student's effort on this "low-stakes" test were included on the midyear test (two mathematics items) and the posttest (two reading items). The two reading items were constructed to represent students' literal and inferential comprehension of a science-related text. The first reading item required that the students comprehend the actual text, while the second required them to infer from the text. Similarly, of the two mathematics items, one required a well-defined arithmetic operation, while the second required students to identify the relevant features of a word problem before responding. Mean reading and math scores were both 58%. The items were used to construct a composite variable. Students with less than half of the nonscience content items they encountered on the midyear test and/or posttest correct (27% of participants) were tagged as "low nonscience." This index allows us to examine gains

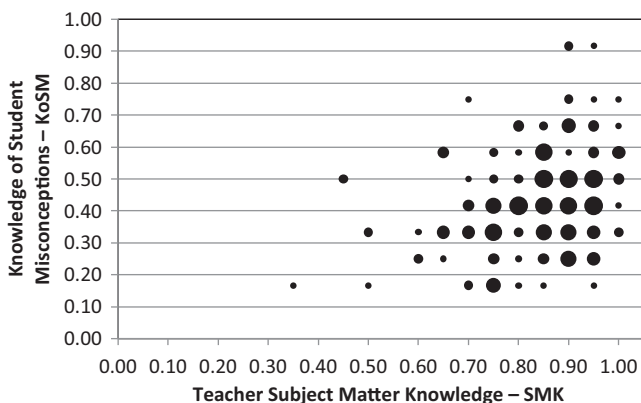


Figure 1. Teachers' mean subject matter knowledge versus mean knowledge of student misconceptions.

Note. $N = 181$ teachers. KOSM = knowledge of student misconceptions; SMK = subject matter knowledge. This figure illustrates the correlation between teachers' KOSM and SMK. Symbol area is proportional to the number of teachers with a particular combination of SMK and KOSM. Note how few teachers have high KOSM and low SMK (upper-left sector).

for each group separately. We hypothesize that students who performed in the low nonscience range in reading or doing simple math would have had difficulty answering the science questions on the test or would simply not have given the test their best effort.

Teacher SMK performance on the pretest was strong, with 84.5% ($SD = 13.7\%$) correct on non-misconception items and 82.5% ($SD = 13.8\%$) on misconception items. Hence, on average, teachers missed only 3 out of 20 items. Teachers' KOSM, the ability to identify the most common wrong answer on misconception items, was weak, with an average score of 42.7% ($SD = 16.6\%$) identified, averaging only 5 out of the 12 items with strong misconceptions. If teachers simply guessed the most common incorrect response, the accuracy of prediction would be 3 out of 12 items, on average (for those teachers who knew the correct answer). Teachers' mean SMK and KOSM are graphed in Figure 1.

At the item level of analysis, teachers' performance on each of the eight non-misconception items falls into one of two categories:

- SMK (teacher answered correctly)—84.6% of responses
- No SMK (teacher answered incorrectly)—15.4% of responses

As expected, the majority of teachers were competent in their SMK, especially when the item did not include a strong misconception among its distractors.

Teachers' performance on each of the 12 misconception items falls into one of four possible categories:

- Both SMK and KOSM (teacher answered correctly and knew the most common wrong student answer)—40.7% of responses
- SMK, no KOSM (teacher answered correctly, but did not know the most common wrong student answer)—41.8% of responses
- No SMK, KOSM (teacher answered incorrectly, but knew the most common wrong student answer)—2.0% of responses
- No SMK, no KOSM (teacher answered incorrectly and did not know the most common wrong student answer)—15.5% of responses

In the case of teachers not knowing the science (i.e., getting the item wrong), most selected the dominant student misconception as their “correct” answer. We decided to combine the third and fourth categories into one, because teachers in both categories did not possess the relevant SMK for that item. Moreover, it is hard to interpret what the very small (2.0%) “no SMK, KOSM” category really means, because in these cases the teachers evidently have rather eccentric views. If the teachers themselves hold an uncommon misconception, why would any particular benefit for the students derive from those teachers correctly identifying the students' common misconception? Moreover, the number of teachers with high KOSM (as measured by knowledge of misconceptions) and low SMK was also very small, as shown in the upper-left quadrant of Figure 1. Teacher SMK and KOSM thus appear related, rather than independent from each other (Kind, 2009). Whereas McEwan and Bull (1991) argued that there are no formal differences between types of teacher knowledge, it seems that SMK, at least in the form that we measure, should be considered a necessary, but not sufficient, precondition of KOSM.

Inferential Analysis

The item-level hierarchical logistic regression model results (Table 3) reveal that student gains are related to their teacher's knowledge level. The coefficients in this model represent the log odds of a particular student answering an item correctly. All control variables are significant in the logistic model. For instance, students with high nonscience levels (i.e., those who correctly answered at least 50% of the four reading and math items) and those who scored correct on the pretest item had better odds of answering the posttest correctly. As one would expect, item difficulty is also significant and has the largest coefficient.

Table 3
Item-Level Hierarchical Logistic Regression Model Results

	Variable	DoF	F Value	Prob.	Categories	Logistic Coeff.	t Value	Prob.	
Student	Intercept					-3.338 (0.143)	-23.34	0.0001	
	Grade level	1	14.12	0.0002	Year in school	0.064 (0.017)	3.76	0.0002	
	Gender	1	282.54	< .0001	Male, Female	0.206 (0.012)	16.81	< .0001	
	Highest parental ed	1	57.97	< .0001	Years of education	0.041 (0.005)	7.61	< .0001	
	Race	1	28.02	< .0001	White, non-White	0.079 (0.015)	5.29	< .0001	
	Fraction of year	1	137.91	< .0001	1/2,1	0.351 (0.030)	11.74	< .0001	
	Reading and math score	1	90.19	< .0001	Low or high	0.029 (0.010)	2.76	< .0001	
	Pretest score by item	1	956.20	< .0001	Incorrect, correct	0.928 (0.054)	17.19	< .0001	
	Teacher	SMK & KOSM by item	4	12.38	< .0001	SMK, KOSM	0.047 (0.051)	0.92	0.3577
						SMK, no KOSM	-0.069 (0.051)	-1.37	0.1720
No SMK, no KOSM						-0.034 (0.057)	-0.59	0.5521	
No dist, SMK						0.122 (0.049)	2.46	0.0141	
Item level Interactions	Item difficulty SMK & KOSM by item	4	11.58	< .0001	No dist, no SMK	4.004 (0.042)	95.94	< .0001	
					Item difficulty SMK, KOSM	0.140 (0.055)	2.54	0.0112	
Cases Pseudo-R2	Pretest score by item X SMK&KOSM by item	4	17.52	< .0001	SMK, no KOSM	0.058 (0.055)	1.06	0.2898	
					No SMK, no KOSM	-0.061 (0.062)	-0.98	0.3276	
					No dist, SMK	0.185 (0.054)	3.42	0.0006	
					No dist, no SMK SMK, KOSM	0.008 (0.051)	0.16	0.8729	
Cases Pseudo-R2	Pretest score X reading and math score	1	236.34	< .0001	SMK, no KOSM	0.185 (0.051)	3.65	0.0003	
					No SMK, no KOSM	0.268 (0.056)	4.75	< .0001	
Cases Pseudo-R2	Item difficulty X pretest score by item	1	122.79	< .0001	No dist, SMK	0.090 (0.050)	1.81	0.0700	
					No dist, no SMK	0.374 (0.024)	15.37	< .0001	
						-0.720 (0.065)	-11.08	< .0001	

Note. N = 9,556 at the student level; N = 181 at the teacher level. KOSM = knowledge of student misconceptions; SMK = subject matter knowledge.

The following interactions are significant: Students who both had a high nonscience level and scored correct on the pretest item received an extra boost; and the benefits of having a more knowledgeable teacher, with SMK and KOSM, overproportionately accrued to the high nonscience students. In addition, an interaction between the pretest score and the SMK and KOSM category indicates that students' posttest results were more sensitive to levels of teacher knowledge when the students checked a wrong answer in the pretest than when they selected the correct answer on the pretest. This pattern is to be expected because the binary nature of the student score in the item-level model exerts a ceiling effect for those with a correct pretest score; they could not improve on that item, they could only score worse on the posttest on that item. Finally, there is an interaction between item difficulty and pretest score. On the easy items, there is relatively little difference in the log odds of answering the posttest correctly, regardless of whether the students had the pretest correct or not; on the difficult items, that difference is larger.

The log odds of the logistic model require transformation to become more intuitive and interpretable. Hence, we convert the coefficients of the logistic regression into odds and then into probabilities of a correct response on the posttest for the categories of interest. We then use these probabilities arising from the logistic model to calculate the gain between pre- and posttest scores for each category of interest in units of standard deviation of the pretest scores. Calculating the effect size of a year of instruction requires that student-estimated probabilities of answering the posttest correctly be aggregated across the two relevant item categories.

The effects of teacher knowledge can be clearly seen (Figure 2). There are large differences between students of low and high nonscience levels. For students at the low nonscience level, there was no significant gain on a particular non-misconception item if their teacher did not have the SMK for that item, but there was a significant effect if a teacher did have the SMK. Furthermore, for low nonscience students, there was no significant gain on misconception items for any of the categories of teacher knowledge: no SMK, SMK only, or SMK plus KOSM. For high nonscience students, gains were much larger. Even for students with teachers who did not have the SMK for an item, medium-sized gains can be seen. Large gains accrued to high nonscience students with teachers having SMK on non-misconception items. The most interesting results are seen for high nonscience students on misconception items: Students of teachers who had only SMK on an item had gains that were not significantly different from those of students who had teachers without SMK. Only when teachers had both SMK and KOSM were student gains significantly larger. While teacher SMK is a strong predictor for gains on non-misconception items, teacher SMK alone is not sufficient to predict higher gains (relative to the absence of teacher SMK) for high nonscience students on items with strong misconceptions.

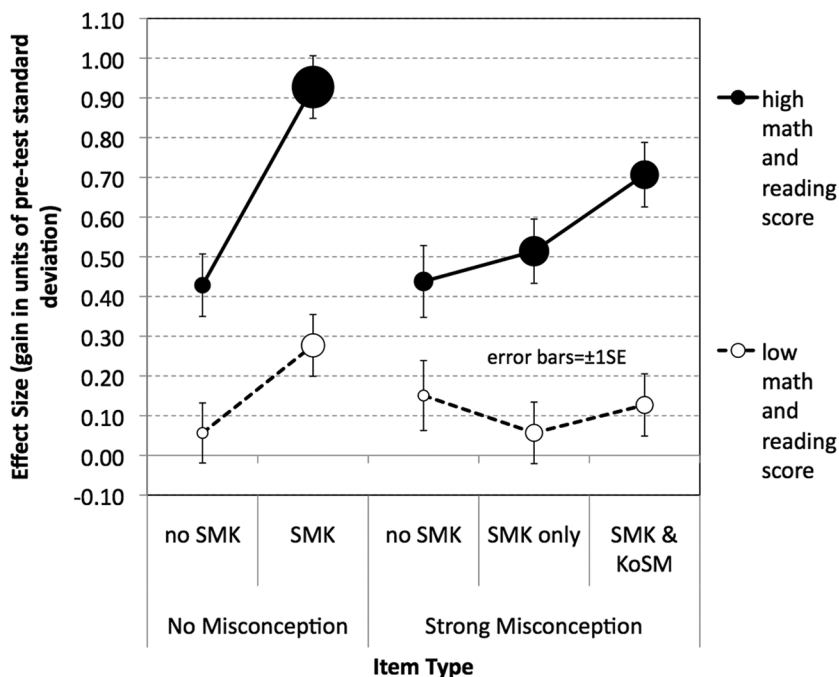


Figure 2. Yearly middle school physical science gains, estimated by hierarchical logistic regression.

Note. $N = 9,556$ for students and 181 for their teachers. KOSM = knowledge of student misconceptions; SMK = subject matter knowledge. This figure illustrates students' gains from pretest to posttest as estimated by hierarchical logistic regression, as a function of their teachers' SMK and KOSM and is scaled up to performance on the 20-item test. "No SMK" represents a teacher who gets no items correct. "SMK only" represents all SMK items correct. KOSM represents knowledge of all student misconceptions. Symbol areas represent the proportion of cases.

Discussion

Our analysis of students' knowledge of the concepts in the NRC content standards over the course of one or two semesters of middle school physical science shows moderate levels of gain. This is good news: Students are learning the content represented by the NRC standards, although mastery (performance at the 80% level or higher) is elusive for most. Analysis of teacher knowledge at the start of the year shows high levels of SMK, with some weaknesses, and rather moderate levels of KOSM, as measured by teachers' prediction of the most common wrong answers of their students. The item-level analysis uses each teacher's SMK and KOSM for each item

to predict their students' performance on that same item. Using a dichotomous teacher score increases the variation in teacher responses. While the mean value of these item-level KOSM and SMK responses stays the same, the standard deviation of these 0 and 1 responses increases to 37.2% and 48.2%, respectively (i.e., by a factor of 3 for SMK and of 5 for KOSM). By mapping teacher knowledge more precisely, this approach tends to "amplify" the differences in teacher knowledge and thus contributes to the sensitivity of the model. The item-level analysis keeps the "holes" in teacher knowledge in play, differences that may otherwise be obscured by a high average score.

This analysis shows significant differences in student gain associated with teacher knowledge and with student nonscience level (Figure 2). Students with high reading and math scores showed much larger gains than students who scored low on the nonscience items. These students, even if their teacher did not have the requisite SMK and KOSM, made moderate gains. There are many possible explanations for this result. For instance, these students may have found ways to gain knowledge from other sources, perhaps the textbook, homework, or discussion with other students. Having more knowledgeable teachers is associated with even larger gains for the high nonscience students than for the low nonscience students, bringing to mind the so-called "Matthew effect," coined by sociologist Robert Merton in 1968 (and adapted to an education model in 1986 by Keith Stanovich), which, loosely stated, says that those with an attribute in abundance (in this case, science knowledge) tend to gain more than those who start with less. Research has found that students with low reading levels exhibit lower gains in other subjects because much of the learning effort requires reading texts (Adams, 1990).

It also may be the case that students who answered the embedded reading and mathematics items incorrectly may simply not have taken this "low-stakes" test seriously. Those with low scores on these questions may have gotten these questions wrong because they were uninterested or uninvolved, and their performance on the 20 science items may have suffered in parallel. If this is the case, the findings for students of high nonscience levels (73% of the total) should be emphasized as more fairly reflecting the impact of teacher SMK and KOSM. However, a significant gain can be seen on non-misconception items for low nonscience students if they had a knowledgeable teacher, so at least some appear to have taken the tests seriously. It also appears that students with low reading and math scores were particularly dependent on the teacher's SMK, exhibiting no significant gain unless their teachers had the requisite SMK for these items (and the items had no misconceptions). The lack of gain on misconception items for these students, independent of the level SMK or KOSM, is particularly troubling. These items may simply have been misread, or they may be cognitively too sophisticated for these students at this point in their education,

or they may not have tried their hardest on a low-stakes test. O'Reilly and McNamara (2007) found that reading skill helped high school students earn higher scores on tests of science knowledge with a larger effect on students starting with more science knowledge. The significant and large positive coefficient for the interaction of *pretest score* \times *reading and math score* of 0.374 (0.024) lends support to this finding.

Among the students with high math and readings cores, our analysis reveals a clear relationship of teacher knowledge to student gains. For non-misconception items, student gains are nearly double if the teacher knows the correct answer. When items have a strong misconception, students whose teachers have KOSM are likely to gain more than do students of teachers who lack KOSM. Much of what happens in many science classrooms could be considered as simply a demonstration of the teacher's own SMK, without taking into account the learner's internal state. Without knowledge of misconceptions relevant to a particular science concept, it appears that students' success at learning will be limited.

The reason that many prior studies of the influence of teacher knowledge on student learning may not have found significant effects may lie, at least partially, in their painting with too broad a brush. The grain size of analysis of teachers' knowledge may be important. Our own initial analysis of total test scores (not shown) captured neither the nuances of a teacher's strengths and weaknesses nor the effects that these nuances have on student learning. Even when assessments of teacher knowledge have been carefully developed and rigorously analyzed (e.g., Hill et al., 2005), the magnitude of teacher knowledge effects is small. Our test-level analysis resulted in a similar relationship to teacher knowledge (0.05 to 0.10 *SD* between teachers low or high in overall measures of knowledge). However, moving to the item level, we find much stronger evidence that the knowledge that teachers need for teaching a particular science concept is both the SMK specific to that concept and, if there is a popular misconception among students about the concept, an awareness of this misconception. There appears to be little "transfer" of teacher SMK or KOSM between concepts, for example, a teacher's firm grasp of electrical circuits and relevant misconceptions appears to have little to do with the effective teaching of chemical reactions. Teachers who are generally well versed in physical science still may have holes that affect student learning of a particular concept. Our findings suggest that it is important to go to a smaller grain size and examine teacher knowledge surrounding particular concepts, because student performance at the item level is associated with teacher knowledge of a particular concept.

The relevance of this study for practitioners in teacher training and PD is to consider that an emphasis on identifying and remediating holes in the teachers' knowledge may be more helpful for the science teachers' effectiveness in the middle school classroom than developing a deep understanding

of only a few particular topics. Especially with the increasing emphasis on state testing, teachers must be prepared to teach all required topics well, and not just focus on a few of their favorites and avoid any topics in which they are only weakly prepared. Moreover, in teaching concepts for which students have misconceptions, knowledge of students' ideas may be the critical component that allows teachers to construct effective lessons. Because teacher KOSM is low, compared with their knowledge of the science content, PD programs that focus on this area are poised to effect substantial improvements.

A few caveats are in order. While this study has a large number of participants, it is not an experimental study. Its findings are correlational in nature. We have demonstrated that student learning is related to teacher knowledge, and the inclusion of several student-level variables allows us to account for alternative hypotheses that student background may contribute to the gains observed. However, our measures of teacher SMK and KOSM may simply be proxies for other variables not included in the model. One could imagine, for instance, that years of teaching experience is the key contributor to SMK and KOSM, and hence student gains. To explore if this might be the case, we investigated models using such variables: teachers' years of teaching school, years of teaching physical science, undergraduate degree (science, education, science and education, other), and graduate degree (science, education, science and education, other, none). None of these teacher variables reached a level of statistical significance ($p \leq .05$) when included with SMK and KOSM measures. We interpret this as meaning that our measures of SMK and KOSM are better at predicting student gains than the broader measures of teacher background for predicting student gains on the items tested.

Another concern is that participating teachers volunteered to join this project and that our results, therefore, may not be generalizable to other middle school physical science teachers. It may well be that our teachers were more confident in their abilities, or perhaps eager to be involved in the study because they felt their students would perform well. Or, they may be strongly motivated to be part of experimental programs, and this study may be but one of many in which they have been involved. Without a randomized selection of classrooms, one cannot definitively state the degree to which these classrooms are truly representative of the nation, but this limitation is mitigated by the presence of the whole gamut of teacher and student backgrounds. We hope that this study will form the foundation for one in which classrooms are selected randomly (although we are aware of the practical difficulties such a study would face).

Conclusion

A multiple-choice assessment instrument designed to measure student gains can be effectively "repurposed" to measure teacher SMK and KOSM.

Having teachers both select the correct answer and identify the incorrect answer most commonly chosen by their students fills a gap in the availability of instruments to measure science teachers' knowledge. This method is particularly well suited for gathering data across a variety of PD and teacher preparation programs (Moyer-Packenham, Bolyard, Kitsantas, & Oh, 2008) because such instruments are easy to administer and score.

SMK is an important predictor of student learning. That effective teachers must know the concepts they teach may sound like a truism, but empirical evidence has been rather elusive in prior studies. Attempts in the past to characterize teacher knowledge through global scores on written tests have failed to produce strong predictors of student learning. Hence, a finer grain size of analysis becomes essential here. While one may assume that the science content of middle school physical science is, in general, well understood by teachers, there are noticeable holes in their knowledge, and these weaknesses differ by teacher. It is not surprising that teachers with the proper SMK of a given concept can achieve larger gains with their students than can those lacking that SMK; a teacher without knowledge may teach the concept incorrectly, and students may end up with the same incorrect belief as their teacher. Effectiveness of middle school science teachers may thus have more to do with a mastery of *all* the concepts that they teach than with the depth of their knowledge in any particular topic. The increasing involvement of scientists (i.e., professors of science and research scientists) in teacher PD programs could have the impact of focusing those programs too narrowly on the scientists' special areas of expertise, which might boost participants' SMK only in a narrow set of topics. What might be more advantageous for PD is to conduct a diagnostic identification and remediation of teachers' knowledge "holes."

An intriguing finding of this study is that teachers who know their students' most common misconceptions are more effective than teachers who do not. This particular component of PCK may allow teachers to construct experiences, demonstrations, experiments, or discussions that make students commit to and then test their own ideas. A teacher knowing only the scientific "truth" appears to have limited effectiveness. It is better if a teacher also has a model of how students tend to learn a particular concept, particularly if there is a common belief that may make acceptance of the scientific view or model difficult. This finding, too, has practical implications. In PD programs, an emphasis on increasing teachers' SMK without sufficient attention to the preconceived mental models of middle school students (as well as those of the teachers) may be ineffective in ultimately improving their students' physical science knowledge.

PCK of teachers has been notoriously difficult to measure. This study demonstrates an easy way to measure a particular component of PCK, teachers' awareness of the mental models of their students, using a multiple-choice format assessment. This method requires the existence of a set of

items for a particular science field that captures the most prevalent misconceptions of major concepts. Efforts to develop such tests in the various science domains have produced such “misconception tests” or “concept inventories.” However, few are constructed that consist of a large number of items for which misconception strength is high and is identified in a scoring key. Our project constructed such tests based on the NRC science content standards in physical science (for Grades K–4 and 5–8), earth and space science (for Grades K–4, 5–8, and 9–12), high school chemistry, and high school physics. These instruments are available in two forms, one downloadable online at no cost for teachers and professional developers,⁵ and a second, secure form for researchers available on request from the authors. The measurement of PCK through this single aspect of identification of most prevalent misconceptions may be quite limited, compared with the range of pedagogical teacher knowledge needed to teach well, particularly how to conduct classroom discussions, sequence concepts, run lab sessions, and so on. Yet such a measurement may represent an easily obtainable and powerful indicator of the degree to which teachers are “student centered” and build classroom instruction around the capabilities and needs of their students.

Notes

We wish to thank many who contributed to this study. Cynthia Crockett coordinated recruitment of participating teachers and test administration for the Physical Science Assessment Project. Several scientists and educators aided in construction and revision of test items: Bruce Ward, Bruce Gregory, and Jennifer Grier. Annette Trenga scanned and verified all tests. Freeman Deutsch wrote the code for the online teacher survey data capture and oversaw survey data download and storage. Jean Darwick provided readability expertise for test items. Kerry Rasmussen provided illustrations for test items. Alison Plante and Lisa Portolese cocreated the online teacher survey. Janice M. Earle, Larry E. Suter, Elizabeth VanderPutten, and Kathleen B. Bergin of the National Science Foundation provided invaluable insight and support. Many scientists from around the country volunteered to anonymously review and provide feedback on the items as they were developed. We are indebted to the teachers who felt that this project was worth contributing a piece of their valuable class time to administer our tests and their students’ willingness to answer our questions. This work has been carried out with support from the National Science Foundation (Grants EHR-0454631, EHR-0412382, and EHR-0926272) and the Smithsonian Institution. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Smithsonian Institution.

¹The federal programs are the Math-Science Partnership (MSP) programs of the U.S. Department of Education and the National Science Foundation.

²A particularly common view, often held by adults, is that the seasons are caused by the earth’s elliptical orbit rather than the changing angle of the sun’s rays hitting the surface of the earth.

³For description of the rigorous development process used in creating assessment items and instruments in all fields of science, see Sadler et al. (2009).

⁴Comparable assessments available to researchers and teachers are available online at <http://www.cfa.harvard.edu/smgphp/mosart/>.

⁵Available at <http://www.cfa.harvard.edu/smgphp/mosart/>.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Allen, J., & Le, H. (2008). An additional measure of overall effect size for logistic regression models. *Journal of Educational and Behavioral Statistics*, 33, 416–441.
- Andersson, B. R. (1980). Some aspects of children's understanding of boiling point. In W. F. Archenhold, R. Driver, A. Orton, & C. Wood-Robinson (Eds.), *Cognitive development research in science and mathematics: Proceedings of an international seminar* (pp. 252–259). Leeds, UK: University of Leeds.
- Arzi, H., & White, R. T. (2008). Changes in teachers' knowledge of subject matter: A 17-year longitudinal study. *Science Education*, 9, 221–251.
- Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E. . . . , Zhang, J. (2012). *The condition of education 2011* (NCES 2012–045). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>
- Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view*. New York, NY: Holt, Rinehart & Winston.
- Ball, D. L. (1988). *Knowledge and reasoning in mathematical pedagogy: Examining what prospective teachers bring to teacher education* (Unpublished doctoral dissertation). Michigan State University, East Lansing.
- Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal*, 90, 449–466.
- Ball, D. L. (1991a). Research on teaching mathematics: Making subject matter knowledge part of the equation. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 2, pp. 1–47). Greenwich, CT: JAI.
- Ball, D. L. (1991b). Teaching mathematics for understanding: What do teachers need to know about subject matter? In M. Kennedy (Ed.), *Teaching non-science subjects to diverse learners* (pp. 63–83). New York, NY: Teachers College Press.
- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.
- Ball, D. L., & Bass, H. (2003). Making mathematics reasonable in school. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 27–44). Reston, VA: National Council of Teachers of Mathematics.
- Baser, M. (2006). Fostering conceptual change by cognitive conflict based instruction on students' understanding of heat and temperature concepts. *Eurasia Journal of Mathematics, Science and Technology Education*, 2, 96–114.
- Baturo, A., & Nason, R. (1996). Student teachers' subject matter knowledge within the domain of area measurement. *Educational Studies in Mathematics*, 31, 235–268.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., & . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180.
- Boardman, A. E., Davis, O. A., & Sanday, P. R. (1977). A simultaneous equations model of the educational process. *Journal of Public Economics*, 7, 23–49.
- Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their

- instructors give up too easily? *Journal for Research in Mathematics Education*, 23, 194–222.
- BouJaoude, S. B. (1991). A study of the nature of students' understandings about the concept of burning. *Journal of Research in Science Teaching*, 28, 689–704.
- Brickhouse, N., & Bodner, G. M. (1992). The beginning science teacher: Classroom narratives of convictions and constraints. *Journal of Research in Science Teaching*, 29, 471–485.
- Byrne, C. (1983, October). *Teacher knowledge and teacher effectiveness: A literature review, theoretical analysis, and discussion of research strategy*. Paper presented at the 14th Annual Convocation of the Northeastern Educational Research Association, Ellenville, NY.
- Carlsen, W. S. (1999). Domains of teacher knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 133–144). Norwell, MA: Kluwer.
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50, 66–71.
- Clermont, C., Borko, H., & Krajcik, J. S. (1994). A comparative study of the pedagogical content knowledge of experienced and novice chemical demonstrators. *Journal of Research in Science Teaching*, 31, 419–441.
- Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. New York, NY: Non-science Press.
- Czerniak, C. M., & Lumpe, A. T. (1996). Relationship between teacher beliefs and science education reform. *Journal of Science Teacher Education*, 7, 247–266.
- 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.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8, 1–47.
- Driver, R. (1985). Beyond appearance: The conservation of matter under physical and chemical transformations. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), *Children's ideas in science* (pp. 145–169). Milton Keynes, UK: Open University Press.
- Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. London, UK: Routledge.
- Duckworth, E. (1987). *The having of wonderful ideas & other essays on teaching and learning*. New York, NY: Teachers College Press.
- Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics Education*, 24, 94–116.
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, 28, 458–498.
- Freyberg, P., & Osborne, R. J. (1985). Constructing a survey of alternative views. In R. J. Osborne & P. Freyberg (Eds.), *Learning in science: The implication of children's science* (pp. 166–167). Auckland, New Zealand: Heineman.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996a). The effect of school resources on student achievement. *Review of Educational Research*, 66, 361–396.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996b). Interpreting research on school resources and student achievement: A rejoinder to Hanushek. *Review of Educational Research*, 66, 411–416.
- Grimellini-Tomasini, N., & Pecori Balanda, B. (1987). Teaching strategies and children's science: An experiment on teaching about hot and cold. In J. Novak (Ed.), *Proceedings of the 2nd International Seminar, Misconceptions and*

Influence of Teachers' Knowledge on Student Learning

- Educational Strategies in Science and Mathematics* (Vol. 2, pp. 158–171). Ithaca, NY: Cornell University.
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York, NY: Teachers College Press.
- Hake, R. R. (1998). Interactive engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*, 66–74.
- Halloun, I. A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, *53*, 1043–1055.
- Haney, W., Madaus, G., & Kreitzer, A. (1987). Charms talismanic: Testing teachers for the improvement of American education. *Review of Research in Education*, *14*, 169–238.
- Hanushek, E. A. (1972). *Education and race: An analysis of the educational production process*. Lexington, MA: D. C. Heath.
- Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research*, *66*, 397–409.
- Harbison, R. W., & Hanushek, E. A. (1992). *Educational performance for the poor: Lessons from rural northeast Brazil*. Oxford, UK: Oxford University Press.
- Hawley, W. D., & Rosenholtz, S. (1984). Good schools: A synthesis of research on how schools influence student achievement [Special issue]. *Peabody Journal of Education*, *4*, 1–178.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371–406.
- Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematical knowledge for teaching. *Elementary School Journal*, *105*, 11–30.
- Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, NY: John Wiley.
- Hoz, R., Tomer, Y., & Tamir, P. (1990). The relations between disciplinary and pedagogical knowledge and the length of teaching experience of biology and geography teachers. *Journal of Research in Science Teaching*, *27*, 973–985.
- Huang, H.-W., & Chiu, Y.-J. (1993). Students' conceptual models about the nature and propagation of light. In J. Novak (Ed.), *Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics* (distributed electronically). Ithaca, NY: Cornell University.
- Keuthe, J. L. (1963). Science concepts: A study of sophisticated errors. *Science Education*, *47*, 361–364.
- Kind, V. (2009). Pedagogical content knowledge in science education: Perspectives and potential for progress. *Studies in Science Education*, *45*, 169–204.
- Lederman, N. G., Gess-Newsome, J., & Latz, M. S. (1994). The nature and development of preservice science teachers conceptions of subject matter and pedagogy. *Journal of Research in Science Teaching*, *31*, 129–146.
- Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology*, *77*, 247–271.
- Lockheed, M. E., & Longford, N. T. (1991). Effects on mathematics achievement gain in Thailand. In S. W. Randenbush & J. D. Willms (Eds.), *Schools, classrooms and pupils: International studies for schooling from a multilevel perspective* (pp. 131–148). San Diego, CA: Non-science Press.
- Loughran, J., Berry, A., & Mulhall, P. (2006). *Understanding and developing science teachers' pedagogical content knowledge*. Rotterdam, Netherlands: Sense.

- Ma, I. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum.
- McDermott, L. C., Rosenquist, M. L., & van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: examples from kinematics. *American Journal of Physics*, 55, 503–513.
- McEwan, H., & Bull, B. (1991). The pedagogic nature of subject matter knowledge. *American Education Research Journal*, 28, 316–334.
- Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *American Statistician*, 54, 17–24.
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.
- Mitchell, J. V. (1985). *Ninth mental measurement yearbook*. Lincoln: University of Nebraska, Buros Institute of Mental Measurements.
- Mori, I., Kojima, M., & Deno, T. (1976). A child's forming the concept of speed. *Science Education*, 60, 521–529.
- Moyer-Packenham, P. S., Bolyard, J. J., Kitsantas, A., & Oh, H. (2008). The assessment of mathematics and science teacher quality. *Peabody Journal of Education*, 83, 562–591.
- Mullens, J. E., Murnane, R. J., & Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review*, 40, 139–157.
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Science Board. (2008). *Science and engineering indicators 2008* (2 vols.). Arlington, VA: National Science Foundation.
- Nussbaum, J., & Novak, J. (1976). An assessment of children's concepts of the earth utilizing structured interviews. *Science Education* 60, 535–550.
- O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal*, 44(1), 161–196.
- Piaget, J., & Inhelder, B. (1929). *The child's conception of space* (F. J. Langdon & J. L. Lunzer, Trans.). New York, NY: Norton.
- Pruem, R. (1985). How do 12 year olds approach simple electric circuits? In R. Duit, W. Jung, & C. von Rhoebeck (Eds.), *Aspects of understanding electricity* (pp. 227–234). Kiel, Germany: Schmidt & Klaunig.
- Rowan, B., Chiang, F., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, 70, 256–284.
- Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, 35(3), 265–296.
- Sadler, P., Coyle, H., Miller, J., Cook-Smith, N., Dussault, M., & Gould, R. (2009). The astronomy and space science concept inventory: Development and validation of an assessment instrument aligned with the national standards. *Astronomy Education Review*, 8(1), 1–26.
- Schneps, M. H., & Sadler, P. M. (1987). *A private universe* [Video]. Washington, DC: Annenberg Media.
- Schoon, K. J. (1988). *Misconceptions in earth and space sciences: A cross-age study* (Unpublished doctoral dissertation). Loyola University Chicago, Chicago, IL.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4–14.

Influence of Teachers' Knowledge on Student Learning

- Shymansky, J. A., Woodworth, G., Norman, O., Dunkhase, J., Matthews, C., & Liu, C.-T. (2006). A study of changes in middle school teachers' understanding of selected ideas in science as a function of an in-service program focusing on student preconceptions. *Journal of Research in Science Teaching*, *30*, 737-755.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Snyder, T. (2012). *Digest of education statistics 2011* (NCES No. 2012-001). Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012001>
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360-407.
- Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review*, *5*, 41-48.
- Tatto, M. T., Nielsen, H. D., Cummings, W., Kularatna, N. G., & Dharmadasa, K. H. (1993). Comparing the effectiveness and costs of different approaches for educating primary school teachers in Sri Lanka. *Teaching and Teacher Education*, *9*, 41-64.
- Tekkaya, C. (2003). Remediating high school students' misconceptions concerning diffusion and osmosis through concept mapping and conceptual change text. *Research in Science & Technological Education*, *21*, 5-16.
- Treagust, D. F. (1986). Evaluating students' misconceptions by means of diagnostic multiple-choice items. *Research in Science Education*, *16*, 363-369.
- Tsai, C. C., & Chou, C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning*, *18*, 157-165.
- Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. Gabel (Ed.), *Handbook of research in science teaching and learning* (pp. 177-210). New York, NY: Simon & Schuster Macmillan.
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89-122.
- Wilson, S. M., Floden, R. E., & Ferini-Mundy, J. (2002). Teacher preparation research: An insider's view from outside. *Journal of Teacher Education*, *53*, 190-204.
- Wong, G. Y., & Mason, W. M. (1985). A hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, *80*, 513-524.

Manuscript received January 24, 2012

Final revision received December 19, 2012

Accepted January 12, 2013