

The Impact of High School Life Science Teachers' Subject Matter Knowledge and Knowledge of Student Misconceptions on Students' Learning

Chen Chen,* Gerhard Sonnert, Philip M. Sadler, and Susan Sunbury

Science Education Department, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138

ABSTRACT

One of the foundational assumptions in education is that greater teacher knowledge contributes to greater gains in student knowledge, but empirical evidence in support of this assumption is scarce. Using a U.S. sample of 79 biology teachers and their 2749 high school students, we investigate whether teachers' subject matter knowledge (SMK) and knowledge of students' misconceptions (KOSM) in high school life science are associated with students' posttest performance on multiple-choice test items designed to reveal student misconceptions, after controlling for their pretest scores. We found that students were more likely to answer an item on the posttest correctly if their teachers could answer the question correctly, themselves (SMK). Teachers' ability to predict students' most common wrong answer (KOSM) for an item predicted even better student performance. Items for which a particular wrong answer rose above others in popularity saw an even greater benefit for teacher KOSM.

INTRODUCTION

In 1987, Shulman and colleagues famously conceptualized teacher knowledge (Shulman, 1987; Wilson *et al.*, 1987). In the following 30 years, their formulation went through many generations of extension and modification (Hill *et al.*, 2004; Hill and Chin, 2018; Rowland *et al.*, 2005; Ball *et al.*, 2008), yet the essence has remained largely constant: a knowledgeable teacher should be equipped with both *subject matter knowledge* (SMK) and *pedagogical content knowledge* (PCK). SMK includes possessing the factual knowledge contained in the course material and organizing the material in a curriculum (Ball and McDiarmid, 1989), as well as knowing the correct answer to problems (Ball, 1991), the methods of inquiry (Kennedy, 1990), and the social history root of the subject (McDiarmid, 1988; Baturu and Nason, 1996). PCK consists of a teacher's understanding of students' learning sequence (Clark and Peterson, 1986), anticipation of students' responses to different pedagogies (Stein *et al.*, 2008), and the *knowledge of student misconceptions* (KOSM; e.g., Hill *et al.*, 2005; Darling-Hammond *et al.*, 2009). Both practitioners (National Board for Professional Teaching Standards, 1989; Council of Chief School Officers, 2011) and academics (Shulman, 1986, 1987; Even, 1993; Cohen *et al.*, 2003) agree that teachers' knowledge, a combination of SMK and PCK, is probably the most important teacher quality in the teaching of science, technology, engineering, and mathematics (STEM).

The belief that higher teacher SMK and PCK brings about higher gains in student performance is one of the core theoretical foundations of teacher training and certification policies and professional development programs (Tamir, 1988; Blank *et al.*, 2007; Yoon *et al.*, 2007; Dimitrov, 2009; Zhang *et al.*, 2010; Salinas, 2010; Moyer-Packenham *et al.*, 2011; Wilson, 2013). However, direct empirical support for such a belief is surprisingly scarce (Borasi, 1994; Jacobs *et al.*, 2007; Baumert *et al.*, 2010; Heller, 2010). Moreover, the influence of teachers' SMK and PCK on student learning

James J. Smith, *Monitoring Editor*

Submitted Sep 5, 2019; Revised Dec 13, 2019;
Accepted Dec 23, 2019

CBE Life Sci Educ March 1, 2020 19:ar9

DOI:10.1187/cbe.19-08-0164

*Address correspondence to: Chen Chen
(chen_chen@mail.harvard.edu).

© 2020 C. Chen *et al.* CBE—Life Sciences Education © 2020 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

outcomes is commonly believed to be mitigated by the attractiveness of the misconceptions about specific topics or contexts—the misconception strength (Fisher, 1985; Shulman, 1986; Grossman, 1990; McCrory, 2008; Mavhunga and Rollnick, 2012; Park and Chen, 2012; Sadler *et al.*, 2013b). Yet this assumption is even less examined empirically, because researchers have found it difficult to operationalize and to quantify misconception strength (Duckworth, 1987; Ferrara and Duncan, 2011; Hill and Chin, 2018).

In this study, we ask whether teachers' SMK and KOSM (one component of PCK) in high school life science is associated with students' performance at the end of a school year (posttest) on misconception-driven multiple-choice test items, after controlling for students' scores at the beginning of the school year (pretest). In addition, we ask whether the effects of teacher SMK and KOSM vary by the extent to which the topics invoke frequently held student misconceptions (the misconception strength).

Subject Matter Knowledge

As stated by Ball (1991, p. 5), "Teachers cannot help children learn things they themselves do not understand." Education policies (e.g., No Child Left Behind) and many teacher preparation programs have stressed SMK by requiring teachers to obtain at least undergraduate, or equivalent, degrees in the subject of teaching (e.g., California Teacher Program: teaching credential along with the bachelor degree; UTeach: flexible plan for 4-year or equivalent degree; Ohio State University College of Education and Human Ecology: bachelor of science in education; Teach for America: bachelor degree with a minimum GPA of 2.5; Darling-Hammond *et al.*, 2005; Heilig and Jez, 2014). However, studies examining the relationship between teacher SMK and student outcomes have yielded mixed results in the past four decades (Byrne, 1983; Darling-Hammond, 2000; Wilson *et al.*, 2001; Ahn and Choi, 2004). Some studies found that the amount of course work teachers had completed was not correlated with students' learning gains (Lockheed and Longford, 1991; Monk and King, 1994). Using subject content knowledge tests to directly measure teachers' SMK, some researchers reported no significant relationship between teachers' SMK and teachers' instruction quality (Delaney, 2012) or students' learning achievement (Cauet *et al.*, 2015). Nevertheless, studies have shown that SMK does matter to teaching quality and student outcomes (e.g., Goldhaber and Brewer, 2001). Sanders *et al.* (1993) found that teachers were more capable of organizing and sequencing their presentation logically when they taught within, rather than outside, their subject area expertise. Teachers with more content knowledge expertise were found to make more connections between SMK and the real world (Gess-Newsome and Lederman, 1993), and making connections to the real world in turn was found to enhance student learning (Berlin and Hillen, 1994; Bouillion and Gomez, 2001). Researchers also found that teachers' scores attained on a middle school physical science test were positively predictive of their students' scores on the same test. (Sadler *et al.*, 2013b). Analogous results were obtained in mathematics (Harbison and Hanushek, 1992; Hill *et al.*, 2005, Hill, 2007). Moreover, these researchers emphasized the necessity of considering the effect of SMK in combination with PCK, because teachers with high SMK may still lack the ability to predict

student responses and succumb to "expert blind spot" (Nathan and Petrosino, 2003), that is, a teacher basing instruction on his or her own SMK rather than taking into account a more student-centered perspective reflecting student difficulties and needs (Hellman and Nuckles, 2013).

Knowledge of Student Misconceptions

According to Magnusson *et al.* (1999), PCK consists of knowledge of student's understanding, instructional strategies, curriculum, and assessment. This study focuses on one component under the umbrella of knowledge of student's understanding: teacher KOSM.

Although most teachers are, at an abstract level, familiar with the notion that students bring their own preconceptions (which in many cases are misconceptions, commonly held popular conceptions at odds with accepted science) to formal learning, many teachers are unaware of the students' misconceptions in spite of their own teaching experiences (Morrison and Lederman, 2003; Meyer, 2004; Davis *et al.*, 2006; Gomez-Zwiep, 2008; Otero and Nathan, 2008). In a survey of 30 elementary teachers, Gomez-Zwiep (2008) found that, when interviewed, one-third were unable to provide a single example of student misconceptions from their own classes. Studies have also shown that teachers prefer to adhere to scientifically normative ideas (Otero and Nathan, 2008) and to address students' misconceptions only by reteaching the accurate information (Davis *et al.*, 2006) or the solution (Peterson *et al.*, 1989). The assumption behind such practices is that SMK alone is sufficient for effective teaching. However, as mentioned earlier, such an assumption might trap teachers in an "expert blind spot." On top of delivering the SMK, teachers may need to make additional efforts to assume the students' points of view so as to be able to move the students away from inaccurate preconceptions. Without KOSM, teachers, even with proficient SMK, may be absorbed in their own scientifically accurate points of view and miss the chance to address students' preconceptions.

Many researchers have proposed that student misconceptions afford valuable opportunities for meaningful inquiries to promote conceptual change and should be assiduously addressed by teachers who have adequate KOSM (Smith *et al.*, 1994; Elby, 2000; Hammer *et al.*, 2005; Scott *et al.*, 2007; Larkin, 2012; Delgado and Lucero, 2015). Indeed, Peterson *et al.* (1989) found that first-grade teachers with an average teaching experience of 8.2 years paid more attention to student responses and adopted more inquiry-based probing if they had a better understanding of their students' perspectives. Several studies have shown PCK (including KOSM) to predict teaching quality (Peterson *et al.*, 1989; Hill *et al.*, 2005; Windschitl *et al.*, 2011) and student performance (Peterson *et al.*, 1989; Hill *et al.*, 2005; Delaney, 2012; Sadler *et al.*, 2013b; Ergönenç *et al.*, 2014; Keller *et al.*, 2017). Baumert *et al.* (2010) further showed that the effect of SMK on teaching quality is mediated by PCK.

Misconception Strength

To have a misconception means to hold and believe in a mental model that appears to explain a certain phenomenon or observation but is inherently flawed and unscientific (Vosniadou, 2012). This is different from lacking knowledge, in which case learners hold neither the scientific mental model nor a popular erroneous mental model (few learners have a complete lack of

knowledge, everyone has preconceptions; but some preconceptions—the misconceptions—are more common and are reinforced in the informal learning environment; Spiro *et al.*, 1989; Chew, 2006). A lack of knowledge can be solved by learning the basics, but holding attractive misconceptions often requires “unlearning” and inhibiting the existing mental model, which can increase learners' frustration. For example, our recent study (Chen *et al.*, 2019) showed that students holding misconceptions were more likely than students who lacked knowledge to drop out of an online astronomy course.

In many multiple-choice questions, students' wrong answers are not evenly distributed among all possible wrong choices. There is often one wrong answer, also known as the distractor, that is most frequently selected by the students, and such a distractor often reflects students' common misconceptions. Conversely, students who lack knowledge but do not hold misconceptions would have an about equal probability of selecting (or guessing) any of the choices. Whereas the presence of a strong distractor is commonly deemed undesirable in many academic tests, psychometricians have purposefully developed distractor-driven multiple-choice tests (e.g., Hermann-Abell and DeBoer, 2011; Wind and Gale, 2015) that contain (at least) one popular misconception per item to probe students' misconceptions in science subjects. Studies have shown that students' misconceptions are topic specific and elicited by the context of instruction (Chen *et al.*, 2016; Auerbach *et al.*, 2018). Likewise, teachers' PCK, or KOSM, is also considered to be topic or domain specific (Blömeke *et al.*, 2015), situated in varying contexts (Ball *et al.*, 2008; Lee and Luft, 2008; Borowski *et al.*, 2010; Depaepe *et al.*, 2013; Gess-Newsome, 2015; Hayden and Eades Baird, 2018). Depaepe *et al.* (2015) showed that students had lower scores on items that required a higher level of PCK of their teachers, compared with items that required only teachers' SMK. Regarding the combination of teacher SMK and PCK about a specific test item, Sadler *et al.* (2013b) distinguished three categories: 1) teachers who have neither SMK nor KOSM; 2) those who have only SMK, but not KOSM; and 3) those who have both SMK and KOSM. Teachers who had KOSM but not SMK occurred extremely infrequently. Sadler *et al.* (2013b) further showed that, when test items did not activate student misconceptions, teachers' SMK alone was sufficient to generate student gains, but when question items did invoke strong misconceptions (items that contained a distractor choice that reflected commonly held student misconceptions), only teachers who had both SMK and KOSM could help students achieve significant gains.

Yet misconception strength is never binary; rather it should be seen as a spectrum. Different questions elicit different misconceptions that have different prevalence in the population. Thus, we modified the definition of misconception strength to be the percentage of the student population that chose the most popular wrong answer (therefore the most distracting wrong answer) among all wrong answers of a misconception-driven multiple-choice item.

Research Rationale

Based on our literature review, it appears fair to hypothesize that the difficulty of an item is a function of, among many other factors, its misconception strength. It is also reasonable to assume that teachers with KOSM in specific topics are more

likely to address misconceptions surrounding these topics with their students. Therefore, items with strong misconceptions may appear less difficult to students of teachers who possess KOSM about such items than to students whose teachers lack the relevant KOSM. For items with weaker misconception strength (the misconception is less obvious), the teachers' SMK may be much more important than their KOSM for student learning, because the teacher cannot effectively detect and address a misconception if the misconception is weak.

Measuring both the students' misconceptions and teachers' SMK and KOSM on the same topic, while establishing the misconception strength in the population (not just in the sample) can be very challenging (see earlier efforts made by Carpenter *et al.*, 1982; Ball and Bass, 2000, 2003; Ball *et al.*, 2008; Hill *et al.*, 2004, 2005, 2008; Bell *et al.*, 2010; Krauss *et al.*, 2008; Sadler *et al.*, 2013b). In this study, we adhered to Sadler *et al.*'s (2013b) earlier approach of measuring item-level student gain and teachers' SMK and KOSM, using misconception-driven multiple-choice items. This approach has been shown to be valid and reliable in measuring both student and teacher knowledge (Sadler *et al.*, 2013a; Hill and Chin, 2018). This study differs from Sadler *et al.*'s earlier study (2013b) in two respects. First, Sadler *et al.* (2013b) studied middle school physical science; while this study examines older students studying a high school life science. Second, we expanded on our previous binary definition of misconception strength and instead used a continuous definition of misconception strength. Each item was drafted and tested to probe a disciplinary core idea (DCI) within grade 9–12 *Next Generation Science Standards for Life Science* (NGSS Lead States, 2013). When possible, each item contained one dominant misconception, but the misconception strength varied from item to item. For details about designing, testing, and dissemination of the item bank, please see Sadler *et al.* (2013a).¹

We defined teachers' SMK of an item as answering the item correctly, and teachers' KOSM of an item as correctly identifying the most popular wrong answer in the high school student population. At the beginning of the school year, students answered 29 multiple-choice questions to establish their pretest scores. At the same time, for each question, teachers indicated which answer was correct and also which incorrect answer would be selected most often by their students to establish their SMK and KOSM scores. All tests were returned to the researchers, and teachers were instructed to teach as usual, not to teach or discuss the particular test items with their students. Teachers could teach the content probed by the questions if it was included in their business-as-usual plans. At the end of the year, students answered the same items again (posttest).

Research Question

In this study, we asked:

RQ1: Does teachers' possession of SMK and KOSM of an item predict students' likelihood of answering that item correctly on the posttest, after controlling for students' pretest scores and other demographic information, such as grade, gender, race/ethnicity, and parental education?²

¹Test items available at <https://osf.io/8d6r7>.

²These control variables are commonly observed to strongly predict students' science performance.

We hypothesized that (H1), on average, students of teachers with both SMK and KOSM of an item would have the highest likelihood of answering that item correctly on the posttest; students of teachers having only SMK would rank in the middle; and students of teachers without SMK would rank the lowest.³

RQ2: Does the effect of teachers' SMK and KOSM on students' performance in the posttest vary by that item's misconception strength (interaction effect), after controlling for the same covariates listed in RQ1?

We hypothesized that (H2), for items with low misconception strength, only teachers' SMK, not their KOSM, would matter, so that students of teachers with both SMK and KOSM and students of teachers with only SMK would, on average, have a similar performance, but students who had teachers without SMK would do worse than the two former groups of students. For items with high misconception strength, however, teachers' KOSM would become highly relevant; therefore, students of teachers with both SMK and KOSM would perform better than students of teachers with only SMK, and students of teachers without SMK would rank the lowest. In other words, we hypothesized that the effect of teachers' KOSM on students' performance depends on the misconception strength of an item.

DATA AND METHODS

Pre- and Posttest Development

The 29-item assessment was developed by the MOSART HS-LS project drawing from an original test bank containing 543 high school life science items (see items in the Supplemental File). These were created by a team of science education researchers with expertise in teaching life science. The item development process took 18 months and was completed 3 months before the reported study began. Each multiple-choice item was designed to address a single NGSS DCI and was composed of an item stem, one correct answer, and four incorrect answers. One of the incorrect answer choices was a common misconception among high school students, which was the most frequently selected answer among all of the four wrong answers, as identified by previous research studies of misconception on topics such as cell biology, ecology, genetics, and biological evolution (Haslam and Treagust, 1987; Mann and Treagust, 1998; Odom and Barrow, 1995; Anderson *et al.*, 2002; Lin *et al.*, 2004; Baum *et al.*, 2005; Knight and Wood, 2005; Garvin-Doxas *et al.*, 2007; Bowling *et al.*, 2008; Nadelson and Southerland, 2010; Shi *et al.*, 2010; Tsui and Treagust, 2010). Item writers followed a structure that was nearly identical for each of the 29 NGSS DCIs. They first met as a group to discuss each DCI and its meaning, reviewing relevant literature related to misconceptions and examining existing standardized test items or those used in research studies. During the next 1 to 2 weeks, each member created 10–15 original items. The group then met to vet all items, selecting those that could most definitively differentiate between students who held a particular misconception and those who understood the science behind the DCI. Often, independently written items turned out to be quite similar and could be combined by the group into a new, more streamlined form. Each item was subsequently reviewed independently by three biologists external to the project to assess whether the correct answer agreed with accepted science.

³There was no group in which teacher had KOSM, but not SMK; see explanation in the *Sample* section.

For the purpose of selecting well-performing anchor items for use to connect multiple field-test forms, a pilot test was conducted using an online crowd-sourcing website, Amazon Mechanical Turk (Sadler *et al.*, 2016). With a minimum of 1000 crowd-sourced subjects taking each item, six anchor items were identified having high discrimination and a range of difficulty for use on the field-test forms based on a three-parameter item response theory model. Concurrently, each item was rated by an external reading specialist for appropriateness (none higher than eighth-grade reading level). Of the 543 original items, 523 were deemed to have an appropriate reading level and acceptable correct answers.

Field testing was carried out with the 9740 high school life science students of 187 teachers recruited using a national mailing. Twenty-two test forms were composed of six anchors, 24 additional items, and several demographic questions (e.g., gender, grade level, parental education level). Results were analyzed using classical test theory (CTT) and item response theory, resulting in item parameters for: difficulty, discrimination, guessing, gender bias, and misconception strength. With this information, 29 items (one for each DCI, spread across four standards—cell biology, ecology, genetics and biological evolution—in order to cover as many DCIs as possible while keeping the test as short as possible) were selected for the pre/posttest. As a group, these 29 items exhibited: high unidimensionality (pretest eigenvalue of first factor = 6.34 all others <0.62, Cronbach alpha = 0.88), a range of difficulty (from 0.25 to 0.84, mean of 0.56), high discrimination, low gender bias, and a range of misconception strength (from 0.3 to 0.9, mean of 0.54). Each item was assigned a misconception strength value, which was defined as the percentage of choosing the (most frequent) misconception answer among all wrong answers on the field test. The correlation between misconception strength and item difficulty was not significant.

In the example shown below, choice “C” is the correct answer, and choice “E” is the dominant misconception answer; therefore, the misconception strength in this item is $17/(9 + 2 + 3 + 17) = 0.53$. The mean misconception strength of the 29 items in the pretest was 0.51 (SD 0.17) with a high of 0.92 and a low of 0.28, similar to the field-test results.

The nucleus of a cell

- a) is defined by protons and neutrons 9%
- b) has a positive charge 2%
- c) contains DNA 68%
- d) is defined by electrons 3%
- e) is located in the center 17%

Item Characteristics:

- CTT difficulty (i.e., easiness) = 0.68
- CTT discrimination = 0.48
- CTT misconception strength = 0.53

Sample

In all, 2749 high school students and their 78 life science teachers in whose classes the students were nested were sampled. The 78 teachers were located at 78 different high schools throughout the United States. The median class size was 28. The research reported was approved and determined to be exempt from review by the institution's Institutional Review Board.

TABLE 1. Descriptive statistics of variables

Variable	Mean	SD	Min	Max
Pretest	0.56	0.17	0	1
Posttest	0.64	0.19	0	1
Parent education	3.72	1.08	0	5
Student grade	9.76	0.97	7	12
MS (item misconception strength)	0.53	0.49	0.40	0.90
Teacher's average knowledge				
SMK	0.93	0.06	0.66	1
KOSM	0.31	0.05	0.10	0.41
Asian	0.09			
Black	0.04			
Hispanic	0.14			
Other race	0.12			
White	0.59			
Student gender (M vs. F)	0.45			

In the first weeks of the life science course that the students enrolled in during the Fall semester 2017, each student and each teacher answered 29 high school life science multiple-choice items. In the last session of the course at the end of the school year, each student answered the same 29 items again. In the pretest, teachers were told that students would be tested again (they were not told that the tests were the same).

Table 1 presents the descriptive statistics of the sample. Of the students, 44.8% were male, 55.2% were female. On average, the educational level of the students' parents (we used the highest among the two parents, following Ermisch and Francesconi, 2001) was between three (some college) and four (a 4-year degree). The average correct rate was 56.0% in the pretest (in line with our predetermined average item difficulty of 0.56 when we selected the items) and 64.5% in the posttest (small improvement, suggesting low inflation—teachers did not teach to the test). This indicated that the test was not subject to floor or ceiling effects. On average, for an item answered incorrectly in the pretest, the correct rate in the posttest was 47.0%, and for an item answered correctly in the pretest, the correct rate in the posttest was 77.6%. Aggregating the 29 items for each teacher, teachers on average had SMK for 93% of the 29 items and had PCK for 31% of the 29 items. As mentioned earlier, our definition of teachers' SMK or KOSM is item specific. A teacher possessed SMK of an item if he/she answered the item correctly, and a teacher had KOSM of an item if he/she correctly identified the (most frequent) misconception answer among high school students on the item. To prepare for analysis, we organized the data into the long format. Each row (case) contained the correctness of the posttest of item i answered by student j , and SMK and KOSM of student j 's teacher, as well as other covariates. For each item answered by each student, we further categorized the corresponding teachers' knowledge into combinations of SMK and KOSM. In 6% of the cases, teachers had neither SMK nor KOSM (no-SMK condition); in 46.4% of the cases, teachers had SMK but not KOSM (SMK-only condition); and in 46.7% of the cases, teachers had both SMK and KOSM (SMK&KOSM condition). In only 0.9% of the cases did teachers have no SMK, but correct KOSM, which was a bizarre combination and an extremely small subsample. In this combination, the teacher neither knew the correct answer nor

believed the common misconception, but embraced an eccentric wrong answer. In this case, there is little value in knowing the common misconception. Prior studies also showed teachers' SMK to be a prerequisite and precursor of their pedagogical knowledge (Banilower *et al.*, 2007; Heck *et al.*, 2008; Rollnick and Mavhunga, 2014; Tajudin *et al.*, 2017). Thus, we excluded this subsample of cases, and retained only the above three categories of teacher knowledge for each item answered by each student. It is worth noting that, in the rest of the paper, when we mention a "no-SMK teacher" or "no-SMK condition," we mean that the teacher had no SMK for a particular item. We do not mean a teacher with no SMK for all items; such teachers did not exist in our sample. The same applies to "SMK-only teacher" and "SMK&KOSM teacher."

Analysis

The analysis consisted of three increasingly sophisticated analytical steps. In the first step, we examined the effect of teacher knowledge on student gains on strong and weak misconception items. For this, we aggregated scores across all items for each teacher, while aggregating scores across each item type for each student. Thus, we established overall knowledge scores for each teacher and overall gain scores, separately for strong and for weak misconception items, for each student. In the second and third steps, we proceeded to the smaller "grain size" of item-level models.

Step 1. Following our previous approach (Sadler *et al.*, 2013b), we dichotomized the items into two groups: strong misconception items (misconception strength > 0.5) and weak misconception items (misconception strength ≤ 0.5). We calculated each teacher's mean scores (correct rate) in answering all items (teacher's correct rate in SMK) and in identifying students' misconception answers in all items (teacher's correct rate in KOSM). We set 1 SD below the mean to be the threshold for a low correct rate in SMK ($0.93 - 0.06 = 0.87$) or a low correct rate in KOSM ($0.33 - 0.05 = 0.26$). Based on this threshold, we categorized teachers' correct rates into three categories: the SMK&KOSM category (both SMK and KOSM correct rates were above the threshold), the SMK-only category (only the SMK correct rate was above the threshold), and the no-SMK category (both SMK and KOSM were below the threshold). As reasoned earlier, we excluded the small subsample that had a KOSM correct rate above the threshold but an SMK correct rate below the threshold. We calculated each individual student's mean score (correct rate) in both the pre- and posttest, separately for strong misconception and weak misconception items, and then calculated the student's gain in the correct rate by subtracting the pretest mean score from the posttest mean score for each type of item. We built a multilevel regression model (students clustered in teachers) in which students' gain in the correct rate was the outcome variable, and teachers' knowledge categories and the item misconception strength dummy variable were the key predictors. In this model, we also specified an interaction effect between teachers' knowledge categories and the item misconception strength dummy variable and further controlled for student's gender, age, race/ethnicity, and parental education. To set this analysis apart from the item-level modeling that followed, we named the first-step model the aggregated-level model.

The shortcoming of the aggregate-level model was that the thresholds for misconception strength and for teachers' correct rates in SMK and KOSM were arbitrary. As argued earlier, we hypothesized that students' response to an item would be a function of item misconception strength, and each item's misconception strength should be located on a spectrum of misconception strength, rather than being simply be grouped into a dichotomy. We also considered a teacher's SMK or KOSM to be item specific, that is, we hypothesized that it was a teacher's SMK or KOSM on a specific item that influenced student answers on that specific item. For these reasons, we built and discussed item-level models in steps 2 and 3.

Step 2. Because we were interested in examining factors that affected the correctness of the student posttest response for each item, we had a binary dependent variable (1 = correct answer; 0 = wrong answer) and therefore adopted a logistic approach. More specifically, because students were nested within teachers, we chose a multilevel modeling approach, using a two-level logistic regression. At the first level, the log odds of the correctness of each item i , answered by each student j , in the posttest (POST) were a linear function of the misconception strength (MS) of the item i , the student j 's correctness of item i on the pretest (PRE) and other attributes of student j (including gender, grade, parental education, and race/ethnicity). The fact that students were nested in teachers was signified by the added subscript k , indicating teachers, for the student-level variables. The second-level predictors included teacher k 's SMK and KOSM (two dummy variables for three categories: SMK_{Only} and SMK&KOSM vs. no-SMK) for item i . There were also cross-level interaction terms: SMK_{Only} × MS and SMK&KOSM × MS, which allowed the slope of MS to vary by teachers' knowledge. The formal specification of the model is shown here:

$$\begin{aligned} \text{Logit} \{ \Pr(\text{POST}_{ijk} = 1) \} = & \beta_{00} + \beta_{01} \text{MS}_i + \beta_{02} \text{PRE}_{ijk} \\ & + \beta_{03} \text{MALE}_{jk} \dots + \beta_{0n} \text{Grade}_{jk} + \beta_{11} \text{SMK}_{\text{Only}_{ik}} \\ & + \beta_{12} \text{SMK\&KOSM}_{ik} + \beta_{21} \text{SMK}_{\text{Only}_{ik}} \times \text{MS}_i \\ & + \beta_{22} \text{SMK\&KOSM}_{ik} \times \text{MS}_i + \mu_{0k} + \mu_{1k} \text{MS}_i \end{aligned}$$

β_{00} was the fixed intercept, μ_{0k} was the teacher (class)-level random intercept, β_{01} was the fixed slope of misconception strength (average overall effect of MS), μ_{1k} was the random slope at the second level (deviation of the teacher-specific slope from the fixed slope). β_{02} was the fixed effect of the student pretest; β_{03} , β_{04} , etc., were fixed effects for student attributes, such as gender, grade, race/ethnicity, and parental education. β_{11} and β_{12} were fixed effects of teachers' SMK_{Only} and SMK&KOSM conditions, with no-SMK serving as the baseline. Finally, β_{21} and β_{22} were coefficients for the cross-level interactions of SMK_{Only} × MS and SMK&KOSM × MS. We first built a model without interaction effects (M1) and then added interaction effects in the second model (M2).

Step 3. The item-level models built in the second step only investigated the effects of the various predictors on item-level posttest correctness, but they did not explicitly examine item-

level student gains. At the item level, the trajectory of a student response from pre- to posttest could assume only one of three patterns: to gain (change from 0 to 1), to lose (change from 1 to 0), or to maintain (go from 0 to 0, or from 1 to 1). When pretest = 1, a student could only maintain the correct answer or lose, dropping to a value of posttest = 0; when pretest = 0, a student could only maintain or gain. The probability of these patterns could be shown by separately predicting and plotting the posttest correctness, while holding the pretest correctness to 1 or 0, respectively. When pretest correctness was fixed at 1, the plot would show the probability of maintaining the correctness as a function of teacher knowledge and item misconception strength; when pretest correctness was fixed at 0, the plot would show the probability of gaining correctness.

RESULTS

Aggregate-Level Model

Table 2 presents the results from the aggregate-level model. There was a significant main effect of the teacher knowledge category and a significant interaction effect between the teacher knowledge category and the item misconception strength dummy variable.

For items that had a weak misconception strength, students with a teacher in the SMK&KOSM category achieved equivalent gains with students who had a teacher in the SMK-only category (post hoc test $F(1, 2743) = 0.40, p = 0.52$), and students from both these groups had roughly double the gains compared with students with a teacher in the no-SMK category. For items that had a high misconception strength, students with a teacher in the SMK&KOSM category had roughly twice the gains than did students with teachers in SMK-only or in the no-SMK categories, and there was no significant difference between the SMK-only and no-SMK categories ($F(1, 2743) = 0.27, p = 0.60$). This relationship is shown in Figure 1.

Item-Level Model

Table 3 presents the parameters of the main effects model (M1) and the model including interaction effects (M2). The coefficients in each model were exponentiated to odds ratios.

Examining M1, we could see that, most importantly, students had higher odds of answering an item correctly on the posttest if their teachers had both SMK (answered the item correctly) and KOSM (knew the most popular misconception choice among all wrong answers) of the item, compared with students whose teacher only had SMK of the item (post hoc test $\chi^2(1) = 53.40, p < 0.001$). Further, students in both conditions had higher odds of answering an item correctly than did students whose teachers had no SMK (answered the item incorrectly) in the item. To briefly focus on the control variables, we found, first (unsurprisingly), that items that had been answered correctly in the pretest had higher odds of being answered correctly in the posttest by the same person. Second, students who were male, in higher grades, and had higher parental education had higher odds of answering an item correctly in the posttest. Third, Asian students had the highest correctness; white and Hispanic students ranked second, with no significant difference between them; Black students ranked the lowest.

Shifting to M2, we found an interaction effect between misconception strength and teachers' knowledge. In combination,

TABLE 2. Aggregate-level model predicting student gain in correct rate

	Estimate	SE	
Fixed effects			
(Intercept)	0.203	0.052	***
Item misconception strength dummy			
Strong-MS vs. weak-MS	0.019	0.014	
Teacher correct rate category			
SMK+KOSM vs. no-SMK	0.048	0.011	***
SMK-only vs. no-SMK	0.039	0.016	*
Male	0.012	0.016	
Age	-0.010	0.003	**
Parent education	-0.002	0.003	
Asian	-0.022	0.010	
Black	-0.026	0.013	
Hispanic	-0.003	0.008	
Other race	-0.001	0.009	
Interaction effect			
Strong-MS × SMK+KOSM	-0.023	0.016	
Strong-MS × SMK-only	-0.045	0.022	*
Random effects			
SD (teacher)	0.04		
Residual	0.20		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

the model estimated that the three teachers' knowledge groups had different slopes of misconception strength. The no-SMK group had a negative slope of misconception strength, with a

0.1 increase in misconception strength reducing the odds of answering a posttest item correctly by 8.6% [$\exp(-0.899 \times 0.1) - 1 = -0.086$]. The SMK-only group also had a negative slope of misconception strength, with a 0.1 increase in misconception strength reducing the odds of answering a posttest item correctly by 3.7% ($\exp[(-0.899 + 0.525) \times 0.1] - 1 = -0.037$). The SMK&KOSM group was the only group that had a positive slope of misconception strength, with a 0.1 increase in misconception strength increasing the odds of answering a posttest item correctly by 4.3% ($\exp[(-0.899 + 1.322) \times 0.1] - 1 = 0.043$). The interaction is best illustrated on a probability scale in Figure 2, which shows the estimated probability (with 95% confidence interval) of answering an item correctly as a function of the item's misconception strength, teacher knowledge, and the interaction between the two, while controlling other covariates at their means. We can see that, on average, the SMK&KOSM group had a higher probability of answering a posttest item correctly than did the SMK-only group, which, in turn, had a higher probability of correctness than did the no-SMK group. Zooming into the low end of misconception strength (0.4), the SMK-only and SMK&KOSM conditions could not be distinguished from each other ($\chi^2(1) = 0.07, p = 0.79$), and both conditions had higher probability of correctness than the no-SMK condition by a marginal probability of about 0.08 (SMK-only vs. no-SMK: $\chi^2(1) = 18.98, p < 0.001$; SMK&KOSM vs. no-SMK: $\chi^2(1) = 20.17, p < 0.001$), that is, in terms of effect size, 0.38 of an SD in the pretest score. More interestingly, the SMK&KOSM group had a positive slope of misconception strength, whereas the SMK-only group had a negative slope, which led the SMK&KOSM

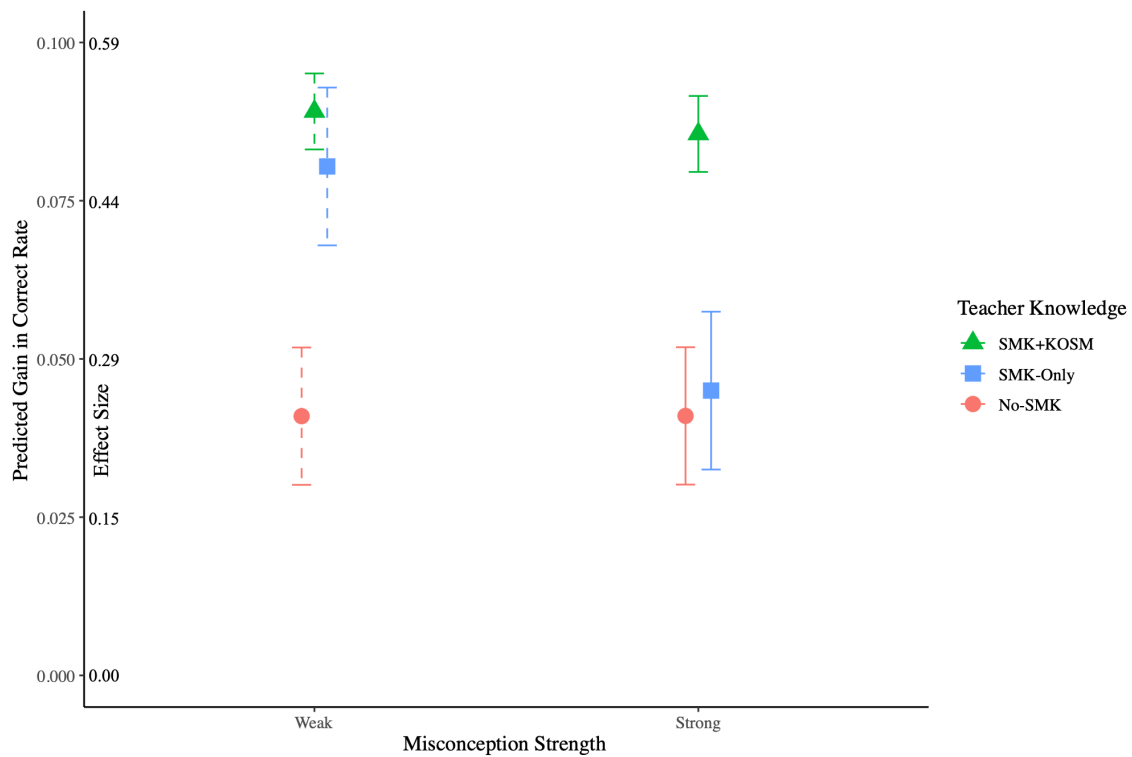


FIGURE 1. Predicted gain in correct rate (also converted to effect size where 1 SD is 0.17 correct rate according to pretest score) by misconception strength groups and by teacher knowledge groups, based on the aggregated model. The error bars indicated ± 1 SE. The sizes of the shapes were proportional to the sample size under each condition.

TABLE 3. Multilevel logistic regression predicting correctness in posttest items

	M1			M2				
	Coefficient	SE	Odds ratio	Coefficient	SE	Odds ratio		
Fixed effects								
Teacher's knowledge								
SMK&KOSM vs. no-SMK	0.261	0.033	***	1.195	-0.346	0.113	***	0.685
SMK-only vs. no-SMK	0.178	0.032	***	1.298	-0.057	0.114		0.945
MS								
(item misconception strength)	0.083	0.050		1.087	-0.899	0.224	***	0.407
Interaction effects								
SMK&KOSM × MS	—	—	—	—	1.322	0.233	***	3.751
SMK-only × MS	—	—	—	—	0.525	0.239		1.690
Controlled variables								
Pretest	1.248	0.017	***	3.483	1.242	0.017	***	3.463
Student gender (M vs. F)	0.023	0.025	*	1.023	0.023	0.025	*	1.023
Parent education	0.138	0.012	***	1.148	0.138	0.012	***	1.148
Student grade	0.138	-0.015	***	0.877	0.136	0.014	***	1.146
Asian	0.133	0.055	*	1.142	0.134	0.015	***	1.143
Black	-0.132	0.069	*	0.876	-0.129	0.069	***	0.879
Hispanic	-0.071	0.050		0.931	-0.068	0.050	***	0.934
Other race	-0.002	0.039		0.998	-0.001	0.040	***	0.999
Intercept	-2.614	0.162	***		-2.145	0.191	***	
Random effects								
SD(MS)	0.280	0.075			0.323	0.072		
SD(Teacher)	0.349	0.043			0.359	0.043		
Corr(MS, Teacher)	0.103	0.289			-0.019	0.242		

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; after false discovery rate (FDR) adjustment.

group to outperform the SMK-only group when misconception strength was equal to, or larger than, 0.45 ($\chi^2(1) = 4.10$, $p = 0.04$). At the high end of misconception strength, the SMK&KOSM group had a higher probability than did the SMK-only group by a margin of 0.10 ($\chi^2(1) = 72.99$, $p < 0.001$, effect size = 0.48) and a higher probability than did the no-SMK group by a margin of 0.20 ($\chi^2(1) = 58.65$, $p < 0.001$, effect size = 0.95).

Table 3 and Figure 2 showed the comparison of posttest scores (controlling for pretest) between predictor conditions. They did not explicitly present the item-level gain. Hence, as described for step 3 in the *Analysis* section, we fixed pretest to 1 or 0 and separately predicted the posttest correctness as a function of item misconception strength and teacher knowledge, based on model M2. This relationship is presented in Figure 3. The top panel of Figure 3 shows that, when an item was answered correctly in the pretest, students did not always maintain the correct answer in the posttest. Instead, students under each condition regressed, as a group, in the probability of giving a correct answer (under the SMK&KOSM or SMK-only conditions, they regressed by about 20%, and under the no-SMK conditions, they regressed by about 30%). For students under the SMK-only or no-SMK conditions, the stronger the item misconception, the less likely they were to maintain the correct answer. Yet for students under the SMK&KOSM condition, the stronger the misconception, the more likely they were to maintain the correct answer. The bottom panel of Figure 3 shows a similar pattern. When an item was answered incorrectly in the pretest, students under each condition gained, as a group, in correctness in the posttest. For students under the SMK-only or no-SMK condition, the stronger the misconception of an item,

the less gain these students achieved, on average, whereas for those under the SMK&KOSM condition, the stronger the misconception strength, the more the students gained.

To summarize our key finding, which has shown up consistently across three different analytical approaches: The stronger the misconception strength of an item, the more difficult the item was in a posttest for students whose teachers had no SMK or only SMK without KOSM on that item. By contrast, for students whose teachers had both SMK and KOSM on the item, the stronger the misconception strength of the item, the easier it became in the posttest.

DISCUSSION

The answer to RQ1 (Does teachers' possession of SMK and KOSM of an item predict students' likelihood of answering that item correctly on the posttest, after controlling for students' pretest scores and other demographic information?) is yes. On average, after controlling for student pretest scores and other demographic information, such as gender, grade, parental education, and race/ethnicity, students were more likely to answer an item correctly if their teachers had both KOSM and SMK of the item, compared with students whose teachers had only SMK of the item. Similarly, students whose teachers had only SMK of an item were more likely to answer the item correctly than did students' whose teachers had no SMK of it (recall that teachers who had no SMK did not have KOSM either). In short, our finding supported our hypothesis (H1) that, in terms of students' posttest performance, teachers' SMK&KOSM > SMK-only > no-SMK. Each condition was significantly different from another, but the odds ratio was small (based on the effect size

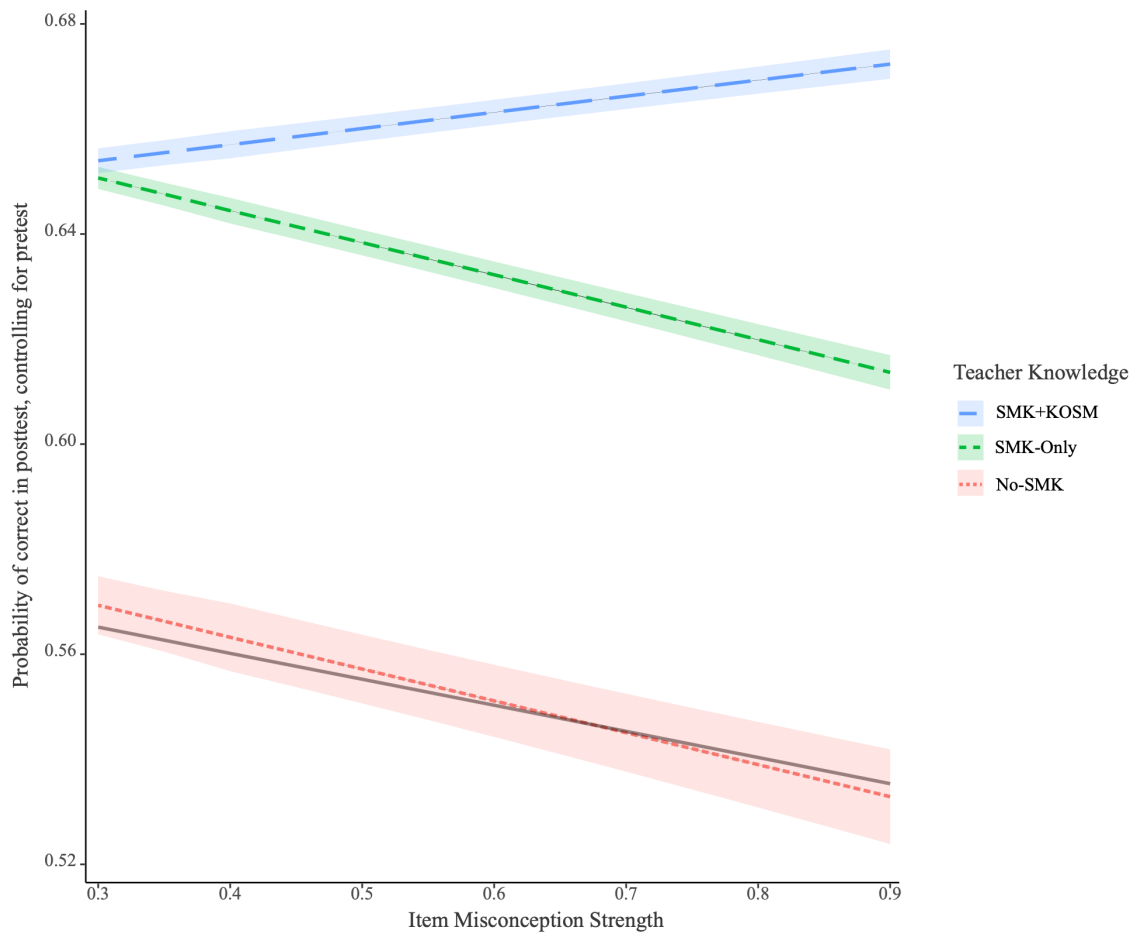


FIGURE 2. The estimated probability (with ± 1 SE confidence interval) of answering an item correctly as a function of the item's misconception strength, teacher knowledge, and the interaction between the two, while controlling other covariates at their means. For reference, the heavy black line is the average pretest score by misconception strength.

criterion for odd ratios, which considers an odd ratio < 1.5 a small effect, suggested by Chen *et al.*, 2010).

Had we estimated only the main effects of teachers' knowledge, we would have concluded that teacher SMK was helpful, and teacher KOSM in addition to SMK was even more helpful, perhaps because the KOSM manifested teachers' PCK, which would be reflected in their lesson design and teaching quality in general. It was only when the interaction effect between teacher knowledge and item misconception strength was included that it became apparent that KOSM was primarily effective for items with stronger misconceptions. This was found in our exploration of RQ2 (Does the effect of teachers' SMK and KOSM on students' performance in the posttest vary by that item's misconception strength [interaction effect]?), where the answer was again yes.

Specifically, at the lower end of the misconception strength scale in Figure 2, SMK-only and SMK&KOSM groups were more similar to each other, performing higher than the no-SMK group. As the misconception strength of items increased, the correct rate in the no-SMK and SMK-only groups both dropped steeply, whereas the correct rate of the SMK&KOSM group increased. At the higher end of misconception strength, the SMK-only group outperformed the no-SMK group, while the

SMK&KOSM group exhibited an even higher correct rate. A plausible explanation is that teachers with SMK can explain and reiterate a target concept correctly, whereas teachers without SMK may have trouble doing so. Because teachers in either condition do not have KOSM of the target concept, they cannot easily "get inside a student's head" to select activities and provide evidence that helps students question and reconstruct the way they think about a concept. For misconceptions that are very popular, students are even more likely to maintain their way of thinking. While teachers with only SMK of an item can at least deliver a correct explanation of the concept, their students have a better chance of identifying the correct answer than do students with no-SMK teachers.

When concepts elicited strong misconceptions, teachers who did not know the correct answer probably held the misconceptions themselves. Without knowing their answers to be incorrect, they were thus likely to actively and confidently teach the wrong ideas to their students. No-SMK teachers may not only have neglected to counteract their students' existing or potential misconceptions (as did the SMK-only teachers), but they may also have actively reinforced the misconceptions, exacerbating the problem, sometimes undoing what a student knew to be true before. This conjecture would explain why the

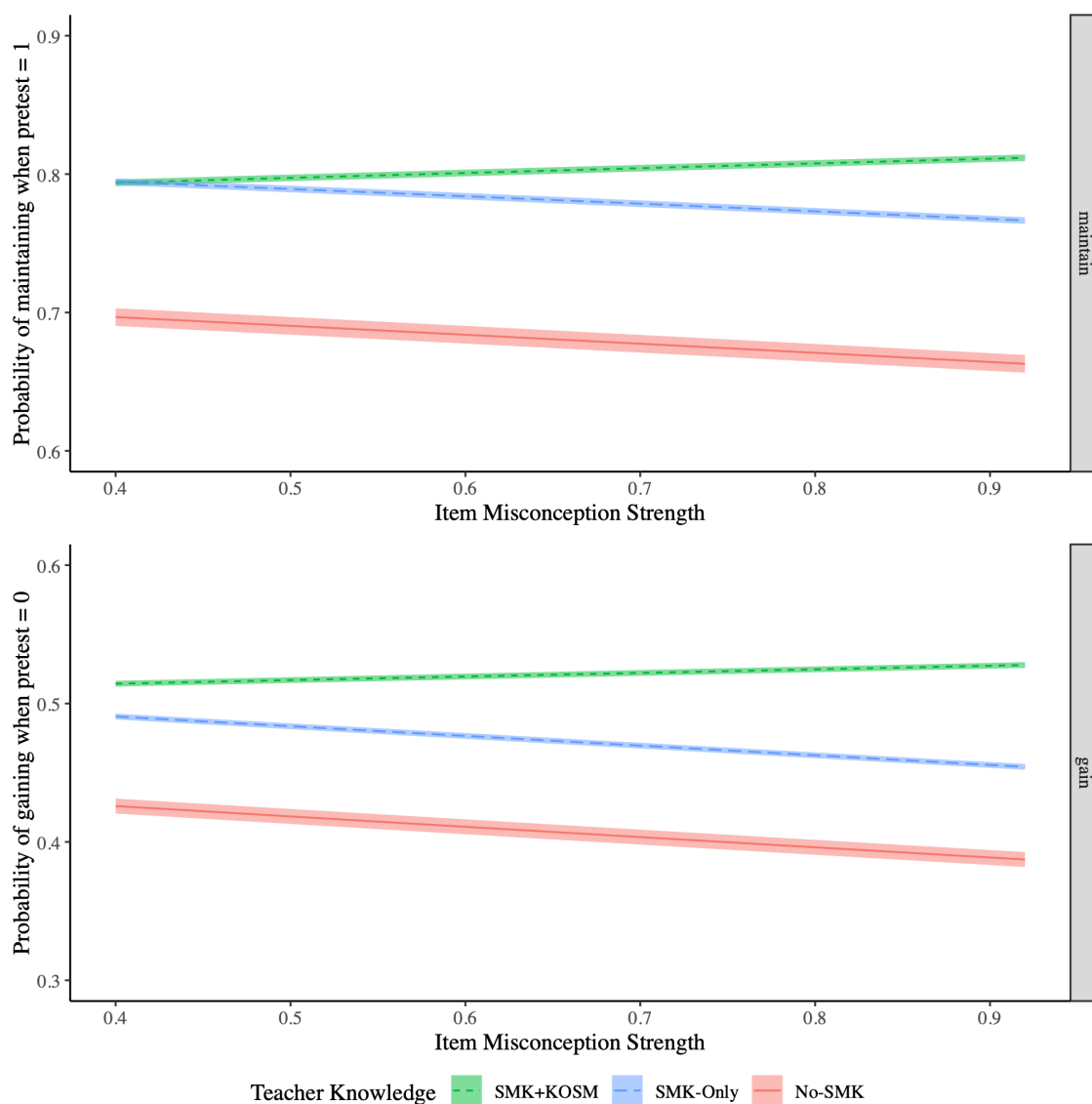


FIGURE 3. The estimated probability (with ± 1 SE confidence interval) of maintaining correctness when pretest = 1 (44% of the student answers) or gaining correctness when pretest = 0 (56% of the student answers) as a function of the item's misconception strength, teacher knowledge, and the interaction between the two, while controlling other covariates at their means.

no-SMK condition had a steeper negative slope for misconception strength than did the SMK-only condition—with the result that student performance suffered most on items with strong misconceptions in the no-SMK condition.

Comparing SMK&KOSM and SMK-only teachers, we found an interesting contrast in the direction of the slope as a function of misconception strength. Whereas the slope was negative for students with SMK-only teachers (and no-SMK teachers as well), it was positive for students with SMK&KOSM teachers for both the gain and maintain groups. In other words, when the teachers did not have KOSM in an item, the stronger the misconception included in an item, the more difficult the item was for the students. But when the teacher did have KOSM, items with stronger misconceptions became easier for the students. A plausible explanation for this pattern is that teachers with KOSM can accurately anticipate students' initial ideas. They

probably not only delivered a scientifically correct explanation, but also structured lessons to address popular misconceptions associated with the target concepts. Therefore, when students encounter an item with strong misconception, they may be able quickly to identify the misconception answer as the wrong answer. What has been the most attractive wrong answer would have become the most obvious wrong answer to be immediately excluded from the multiple choices available to students of SMK&KOSM teachers.

An increasing number of studies have shown that learners do not easily replace misconceptions with the correct conceptions in an information-acquisition manner (e.g., see Pintrich *et al.*, 1993; Murphy and Mason, 2006). Successfully resolving misconceptions involves many affective and situational factors (Gregoire, 2003; Sinatra and Mason, 2008), such as motivation (Taasobshirazi and Sinatra, 2011), emotion

(Broughton *et al.*, 2013), self-efficacy (Pintrich, 1999), and self-regulation (Sinatra and Taasobshirazi, 2011). These findings call for a shift from information-acquisition pedagogies to holistic and contextual pedagogies (Pintrich *et al.*, 1993; Sinatra, 2005), such as restructuring the argument and presentation (Diakidoy *et al.*, 2003) or adopting multiple perspectives (Duit and Treagust, 2003; Chen *et al.*, 2016). Moreover, conceptual change theories have been transitioning from emphasizing a grand mental paradigm shift to focusing on a more fine-grained transformation of personal experience and insights (Vosniadou and Ioannides, 1998; Pugh and Girod, 2007). For example, Heddy and Sinatra (2013) showed that teachers promote stronger conceptual growth in the learning of biological evolution when they relate to students' prior experience and transform that experience, using scientific models, than when they directly refute the misconceptions. It is possible that teachers with KOSM may adopt a variety of pedagogical strategies to effectively address the misconceptions once they correctly predict and identify these misconceptions. Nevertheless, when misconceptions are not obvious, teachers with KOSM may not identify them and end up adopting traditional information-acquisition pedagogies, just like SMK-only teachers. In other words, concepts that elicit stronger misconceptions may give teachers with KOSM more room (and obvious opportunity) to exercise their pedagogical skills.

Recent studies showed that learners not only need to reconstruct their conceptual understanding, but also need to actively inhibit their misconceptions (Brookman-Byrne *et al.*, 2018; Mason *et al.*, 2019), because many misconceptions are difficult to eradicate and keep coexisting with newly acquired conceptions (Gelman, 2011; Legare and Visala, 2011; Shtulman and Lombrozo, 2016). For example, under time pressure, even professional scientists start to reveal intuitive misconceptions (Goldberg and Thompson-Schill, 2009; Kelemen *et al.*, 2013; Shtulman and Harrington, 2016). Functional magnetic resonance images of learners' and experts' brains indicate that there is a need to (subconsciously) inhibit a tendency to revert to prior misconceptions in order to give the correct answer (Foisy *et al.*, 2015; Lubin *et al.*, 2016; Mareschal, 2016; Nenciovici *et al.*, 2018, 2019; Wang, 2018; Mason *et al.*, 2019). It is possible that teachers with KOSM paid more attention to existing and potential misconceptions with their students, and once their students elucidated and found their ideas unproductive, they could more easily inhibit them and start to reason out of their SMK (which is when teachers' SMK starts to have effects). Pushing the scenario to the extreme, even if a student had no SMK of a concept and had to guess randomly, once the most attractive misconception was inhibited, he/she only needed to guess one out of four (with three wrong choices that were unpopular in the population) rather than one out of five (with four wrong choices, one of which was very popular in the population). For items with weaker misconceptions, those misconceptions are not obvious. In these cases, teachers with SMK&KOSM might not activate their misconception identification and resolution skills that benefit their students on items with obvious misconceptions. This may be why such items appeared to be more difficult than items with strong misconceptions for students with SMK&KOSM teachers.

In summary, our findings support our hypothesis (H2) that the effect of teachers' KOSM was context specific, where "context" in our study was defined as the strength of the misconceptions elicited by the question items. In the main effects model, misconception strength appeared to be unrelated to the correct rate, but once we broke down our sample by the teachers' KOSM (and SMK), we found a bifurcated relationship between misconception and correct rate. Items with stronger misconceptions appeared to be more difficult for students of teachers who lack KOSM, but they appeared to be easier for students with teachers possessing KOSM. This suggests that KOSM should be a focus of science teacher preparation and professional development.

Limitations

Readers should be reminded that this was a correlational study, and we could not make any causal claims, because we could not randomly assign students to teachers with different types of knowledge. To disentangle the causal relationship, one option for future study is to conduct randomized control trial interventions at teacher professional development programs, with the treatment group training teachers to improve their KOSM, and then to follow up and compare student outcomes between the treatment and control groups.

Another major limitation is that we did not measure what teachers with KOSM did differently in their classrooms. Thus, we cannot elucidate a particular mechanism or pedagogy at this stage leading from KOSM to classroom practice and to student gains. In the explanation of our findings offered earlier, we made the assumption that, when teachers had only SMK but not KOSM of a concept, they were less likely to address student preconceptions in their pedagogies by eliciting student ideas, providing evidence and activities that allow student to test their ideas, and only then, providing a correct explanation. Future study should examine teachers' pedagogical choices, particularly regarding misconception-related activities, to investigate the relevant mechanisms at work, a factor that remained a "black box" in our study.

Another important question for future study is to what extent the result of this study is transferable to the teaching and learning of other subjects. A lot of the theoretical literature that motivated this study came from the teaching of physics and mathematics. Therefore, we anticipate that we should find similar patterns in these STEM fields. Yet this hypothesis remains to be tested empirically.

Implications

The direct implication for classroom teaching is that it is not enough for teachers to only understand the science concepts that they teach, but they also must have a working knowledge of the ideas that students have when entering their classrooms. Teachers should make use of their KOSM by being able to imagine the target concepts they are attempting to teach from their students' perspectives, with the aim of addressing any misconceptions and achieving, in the end, deeper understanding. Teachers should be cognizant of the evidence that, by working with students on existing or potential misconceptions, they can make it easier for their students to reconceptualize their understanding of a concept. As discussed earlier, studies in the misconception literature have shown that people do not always

fully “replace” their misconceptions, but counteract these misconceptions by intentionally inhibiting them (e.g., Wang, 2018) as they solve problems or answer questions. Before the misconceptions can be inhibited, they need to be explicitly, and repeatedly, considered and reflected upon; and here help from the teachers is essential.

Previous studies have demonstrated a variety of pedagogical methods that can effectively address students’ misconceptions (e.g., Guzzetti, 2000; Eryilmaz, 2002; Tsai, 2003). Most begin with predicting student’s preconceptions or misconceptions, followed by lessons to probe students’ current knowledge and prior experiences. Nevertheless, not all teachers are familiar with these initial steps. In interviews with 30 science teachers in elementary school, Gomez-Zwiep (2008) found that 14 of the teachers could not recall any example of a misconception that their students expressed and that they never thought about misconceptions while planning or teaching classes. Among the 16 teachers who could recall at least one example of their students’ misconceptions, 11 had considered their students’ misconceptions when they were planning for class, and two of them explicitly reported that they tried to tap into students’ prior knowledge, make predictions of students’ understanding, ask students about what they knew, and use the comparison between teachers’ predictions and students’ actual responses to decide on the need for reteaching.

An important question is, of course, what factors or experiences predict teachers’ KOSM. Unfortunately, we could not answer this question in this study (another limitation of the study). Yet it is well documented that improving teachers’ SMK and PCK has been a common goal and feature among teacher professional development programs (Garet *et al.*, 2001; Bell *et al.*, 2010; Goldschmidt and Phelps, 2010; Kelcey and Phelps, 2013; Bausmith and Barry, 2011; Koellner and Jacobs, 2015; Lipowsky and Rzejak, 2015; Polly *et al.*, 2015). For example, the elementary mathematics teacher preparation program at Michigan State University has multiple courses (teaching lab and field instruction) that target the development of teachers’ SMK and KOSM (in the program’s own words, the teacher knowledge of “learners’ prior knowledge,” p. 13), particularly as related to lesson planning and student assessment. Before going through such a program, the preservice teachers were very vague in predicting students’ knowledge and often only provided one pedagogical approach to the teaching task at hand. After completing the program, they showed significant growth in their ability to anticipate the students’ responses and to provide multiple pedagogical approaches to the tasks (Wilson, 2014).

Nevertheless, very few professional development programs adopt such a targeted training or evaluation regarding teachers’ KOSM (Berry and Milroy, 2002). Even more worryingly, in a recent study with a large national sample of teacher professional programs, Doyle *et al.* (2018) found that almost none of the professional development activities, except for learning foundational knowledge in science concepts, had a positive impact on teachers’ KOSM, and that one additional year of teaching experience, on average, improved teachers’ KOSM by less than 1% of an SD in KOSM scores. In short, KOSM appears difficult to acquire in spite of the wide range in scores exhibited by teachers. Nevertheless, considering the importance of KOSM as shown in the findings of this study, teacher preparation and

professional development programming should make efforts in designing and implementing effective training activities that explicitly target KOSM among all types of teacher knowledge.

CONCLUSION

It is a common assumption and observation that strong misconceptions are tenacious and hard for students to resolve (e.g., Chi, 2005), but we found this is not the case for students with teachers who possess both SMK and KOSM about a concept. The combination of teacher SMK and KOSM provides a successful recipe for transmuting those concepts that are anticipated to be difficult due to strong misconceptions into concepts that appear to be easier for the students. Previous studies attributed the positive effect of teachers’ knowledge on students’ achievement to the teachers’ accurate delivery of content knowledge and appropriate adoption of pedagogy in general. This study showed that the effect of teachers’ knowledge is context specific and that KOSM works most effectively on concepts that elicit strong misconceptions—most likely because the more obvious the misconception, the more promptly a teacher can adopt appropriate pedagogies to (re)construct the knowledge with the students or to exhort them to inhibit these misconceptions. Future work should help teachers improve both their SMK and KOSM. It should also look into the development of KOSM during teachers’ professional training and investigate how teachers with both SMK and KOSM interact with students on topics that involve common and tenacious misconceptions.

ACKNOWLEDGMENTS

This work was carried out with support from the National Science Foundation’s grant for MOSART HS LS (EHR-1316645). Any opinions, findings, and conclusions in this article are the authors’ and do not necessarily reflect the views of the National Science Foundation. We thank the 70 scientists who reviewed and commented on the items in the development process. In addition to the authors, the project staff contributing to this research project comprised Annette Trenga, who handled data input and tracking of test forms; John Murray, who programmed and administered online tests; and Harold Coyle, who managed the project. Content consultants who developed items included Joel Mintzes (California State University, Chico) and Kimberly Tanner (San Francisco State University). Jimmy de la Torre (Hong Kong University) provided psychometric support. Andrea Kotula carried out reading-level analysis. Kerry Charles created technical illustrations. Horizon Research provided external evaluation. Our advisory board consisted of Michael Edgar (Milton Academy), Noel Michele Holbrook (Harvard University), Michael Klymkowsky (University of Colorado at Boulder), and Barbara Speziale (Clemson University). We appreciate the advice and support of our colleagues Charles Alcock and Wendy Berland of the Center for Astrophysics | Harvard & Smithsonian. Last but not least, we are grateful for the participation of the many high school life science teachers and their students, without whom this research would have been impossible.

REFERENCES

- Ahn, S., & Choi, J. (2004). Teachers’ subject matter knowledge as a teacher qualification: A synthesis of the quantitative literature on students’ mathematics achievement. Paper presented at: the American Educational Research Association (San Diego, CA). Retrieved August 2019 from <https://files.eric.ed.gov/fulltext/ED490006.pdf>

- Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*, 952–978.
- Auerbach, A. J., Higgins, M., Brickman, P., & Andrews, T. C. (2018). Teacher knowledge for active-learning instruction: Expert–novice comparison reveals differences. *CBE—Life Sciences Education*, *17*(1), ar12.
- Ball, D. L. (1991). Research on teaching mathematics: Making subject matter part of the equation. In Brophy, J. (Ed.), *Advances in research on teaching*. Vol. 2, *Teachers' knowledge of subject matter as it relates to their teaching practices* (pp. 1–48). Greenwich, CT: JAI.
- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In Boaler, J. (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 Kilpatrick, J., Martin, W. G., & Schifter, D. (Eds.), *A research companion to "Principles and Standards for School Mathematics"* (pp. 27–44). Reston, VA: National Council of Teachers of Mathematics.
- Ball, D. L., & McDiarmid, G. W. (1989). The subject-matter preparation of teachers. In Houston, W. R. (Ed.), *Handbook of research on teacher education* (pp. 437–465). New York: Macmillan.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389–407.
- 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.
- Baturo, A., & Nason, R. (1996). Student teachers' subject matter knowledge within the domain of area measurement. *Educational Studies in Mathematics*, *31*(3), 235–268.
- Baum, D. A., Smith, S. D., & Donovan, S. S. (2005). The tree-thinking challenge. *Science*, *310*(5750), 979–980.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133–180.
- Bausmith, J. M., & Barry, C. (2011). Revisiting professional learning communities to increase college readiness: The importance of pedagogical content knowledge. *Educational Researcher*, *40*(4), 175–178.
- Bell, C. A., Wilson, S. M., Higgins, T., & McCoach, D. B. (2010). Measuring the effects of professional development on teacher knowledge: The case of developing mathematical ideas. *Journal for Research in Mathematics Education*, *41*(5), 479–512.
- Berlin, D. F., & Hillen, J. A. (1994). Making connections in math and science: Identifying student outcomes. *School Science and Mathematics*, *94*(6), 283–290.
- Berry, A., & Milroy, P. (2002). Changes that matter. In Loughran, J., Mitchell, I., & Mitchell, J. (Eds.), *Learning from teacher research* (pp. 196–221). New York: Teachers College Press.
- Blank, R. K., de las Alas, N., & Smith, C. (2007). *Analysis of the quality of professional development programs for mathematics and science teachers: Findings from a cross-state study*. Washington, DC: Council of Chief State School Officers.
- Blömeke, S., Gustafsson, J., & Shavelson, R. (2015). Beyond dichotomies. Competence viewed as a continuum. *Zeitschrift für Psychologie*, *223*(1), 3–13.
- Borasi, R. (1994). Capitalizing on errors as "springboards for inquiry": A teaching experiment. *Journal for Research in Mathematics Education*, *25*(2), 166–208.
- Borowski, A., Neuhaus, B. J., Tepner, O., Wirth, J., Fischer, H. E., Leutner, D., ... & Sumfleth, E. (2010). Professionswissen von Lehrkräften in den Naturwissenschaften (ProwiN)—Kurzdarstellung des BMBF-Projekts. *Zeitschrift für Didaktik der Naturwissenschaften*, *16*, 341–349.
- Bouillion, L. M., & Gomez, L. M. (2001). Connecting school and community with science learning: Real world problems and school–community partnerships as contextual scaffolds. *Journal of Research in Science Teaching*, *38*(8), 878–898.
- Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., Moskalik, C. L., & Huether, C. A. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics*, *178*(1), 15–22.
- Brookman-Byrne, A., Mareschal, D., Tolmie, A. K., & Dumontheil, I. (2018). Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLoS ONE*, *13*(6), e0198973.
- Broughton, S. H., Sinatra, G. M., & Nussbaum, E. M. (2013). "Pluto has been a planet my whole life!" Emotions, attitudes, and conceptual change in elementary students' learning about Pluto's reclassification. *Research in Science Education*, *43*(2), 529–550.
- Byrne, C. J. (1983). Teacher knowledge and teacher effectiveness: A literature review, theoretical analysis and discussion of research strategy. Paper presented at: the 14th Annual Convention of the Northeastern Educational Research Association (Ellenville, NY).
- Carpenter, T. P., Moser, J. M., & Romberg, T. A. (Eds.) (1982). *Addition and subtraction: A cognitive perspective*. Hillsdale, NJ: Erlbaum.
- Cauet, E., Liepertz, S., Borowski, A., & Fischer, H. E. (2015). Does it matter what we measure? Domain-specific professional knowledge of physics teachers. *Schweizerische Zeitschrift für Bildungswissenschaften*, *37*(3), 462–479.
- Chen, C., Schneps, M. H., & Sonnert, G. (2016). Order matters: Sequencing scale-realistic versus simplified models to improve science learning. *Journal of Science Education and Technology*, *25*(5), 806–823.
- Chen, C., Sonnert, G., Sadler, P. M., Sasselov, D., & Fredericks, C. (2019). The impact of student misconceptions on student persistence in a MOOC. *Journal of Research in Science Teaching*. <https://doi.org/10.1002/tea.21616>
- Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation*, *39*(4), 860–864.
- Chew, S. L. (2006). Seldom in doubt but often wrong: Addressing tenacious student misconceptions. In Dunn, D. S., & Chew, S. L. (Eds.), *Best practices for teaching introduction to psychology* (pp. 211–223). Mahwah, NJ: Erlbaum.
- Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences*, *14*(2), 161–199.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In Wittrock, M. C. (Ed.), *Third handbook of research on teaching* (pp. 255–296). New York: Macmillan.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119–142.
- Council of Chief School Officers. (2011). *InTASC Model Core Teaching Standards*. Retrieved August 2019 from www.ccsso.org/resources/publications/intasc_model_core_teaching_standards_2011_ms_word_version.html
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, *8*(1), 1–46.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*, 42.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Dallas, TX: National Staff Development Council.
- Davis, E. A., Petish, D., & Smithy, J. (2006). Challenges new science teachers face. *Review of Educational Research*, *76*(4), 607–651.
- Delaney, S. (2012). A validation study of the use of mathematical knowledge for teaching measures in Ireland. *ZDM—The International Journal on Mathematics Education*, *44*(3), 427–441.
- Delgado, C., & Lucero, M. M. (2015). Scale construction for graphing: An investigation of students' resources. *Journal of Research in Science Teaching*, *52*(5), 633–658.
- Depaepe, F., De Corte, E., & Verschaffel, L. (2015). Students' non-realistic mathematical modeling as a drawback of teachers' beliefs about and approaches to word problem solving. In Pepin, B., & Roesken-Winter, B. (Eds.), *From beliefs to dynamic affect systems in mathematics education* (pp. 137–156). New York: Springer International.

- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12–25.
- Diakidoy, I. A. N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology, 28*(3), 335–356.
- Dimitrov, D. M. (2009). Intermediate trends in math and science partnership-related changes in student achievement with management information system data. *Journal of Educational Research & Policy Studies, 9*(2), 97–138.
- Doyle, J., Sonnert, G., & Sadler, P. (2018). How professional development program features impact the knowledge of science teachers. *Professional Development in Education, 1*–16. <https://doi.org/10.1080/19415257.2018.1561493>
- Duckworth, E. (1987). *The having of wonderful ideas and other essays on teaching and learning*. New York: Teachers College Press.
- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671–688.
- Elby, A. (2000). What students' learning of representations tells us about constructivism. *Journal of Mathematical Behavior, 19*(4), 481–502.
- Ergöneç, J., Neumann, K., & Fischer, H. E. (2014). The impact of pedagogical content knowledge on cognitive activation and student learning. In Fischer, H. E., Labudde, P., Neumann, K., & Viiri, J. (Eds.), *Quality of Instruction in physics* (pp. 145–160). Münster, Germany: Waxmann.
- Ermisch, J., & Francesconi, M. (2001). Family matters: Impacts of family background on educational attainments. *Economica, 68*(270), 137–156.
- Eryilmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching, 39*, 1001–1015.
- Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics Education, 24*(2), 94–116.
- Ferrara, S., & Duncan, T. (2011). Comparing science achievement constructs: Targeted and achieved. *Educational Forum, 75*(2), 143–156.
- Fisher, K. M. (1985). A misconception in biology: Amino acids and translation. *Journal of Research in Science Teaching, 22*(1), 53–62.
- Foisy, L. M. B., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education, 4*(1–2), 26–36.
- Garet, M. S., Porter, A. C., Desimone, L. M., Birman, B., & Yoon, K. S. (2001). What makes professional development effective? Analysis of a national sample of teachers. *American Educational Research Journal, 38*(3), 915–945.
- Garvin-Doxas, K., Klymkowsky, M., & Elrod, S. (2007). Building, using, and maximizing the impact of concept inventories in the biological sciences: Report on a National Science Foundation-sponsored conference on the construction of concept inventories in the biological sciences. *CBE—Life Sciences Education, 6*(4), 277–282.
- Gelman, S. A. (2011). When worlds collide—or do they? Implications of explanatory coexistence for conceptual development and change. *Human Development, 54*(3), 185–190.
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In Berry, P., Friedrichsen, P., & Loughran, J. (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 38–52). New York: Routledge.
- Gess-Newsome, J., & Lederman, N. G. (1993). Preservice biology teachers' knowledge structures as a function of professional teacher education: A year-long assessment. *Science Education, 77*(1), 25–45.
- Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science, 20*, 480–487.
- Goldhaber, D. D., & Brewer, D. J. (2001). Evaluating the evidence on teacher certification: A rejoinder. *Educational Evaluation and Policy Analysis, 23*(1), 79–86.
- Goldschmidt, P., & Phelps, G. (2010). Does teacher professional development affect content and pedagogical knowledge: How much and for how long? *Economics of Education Review, 29*(3), 432–439.
- Gomez-Zwiep, S. (2008). Elementary teachers' understanding of students' science misconceptions: Implications for practice and teacher education. *Journal of Science Teacher Education, 19*(5), 437–454.
- Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review, 15*(2), 147–179.
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. New York: Teachers College Press.
- Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading & Writing Quarterly, 16*(2), 89–98.
- Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. *Transfer of Learning from a Modern Multidisciplinary Perspective, 89*.
- Harbison, R. W., & Hanushek, E. A. (1992). *Educational performance of the poor: Lessons from rural northeast Brazil*. Oxford, UK: Oxford University Press.
- Haslam, F., & Treagust, D. F. (1987). Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. *Journal of Biological Education, 21*(3), 203–211.
- Hayden, H. E., & Eades Baird, M. (2018). Not a stale metaphor: The continued relevance of pedagogical content knowledge for science research and education. *Pedagogies: An International Journal, 13*(1), 36–55.
- Heck, D. J., Banilower, E. R., Weiss, I. R., & Rosenberg, S. L. (2008). Studying the effects of professional development: The case of the NSF's local systemic change through teacher enhancement initiative. *Journal for Research in Mathematics Education, 39*(2), 113–152.
- Heddy, B. C., & Sinatra, G. M. (2013). Transforming misconceptions: Using transformative experience to promote positive affect and conceptual change in students learning about biological evolution. *Science Education, 97*(5), 723–744.
- Heilig, J. V., & Jez, S. J. (2014). *Teach for America: A return to the evidence*. Boulder, CO: National Education Policy Center. Retrieved April 24, 2015, from <http://nepc.colorado.edu/publication/teach-for-america-return>
- Heller, J. I. (2010). *The impact of Math Pathways & Pitfalls on students' mathematics achievement and mathematical language development: A study conducted in schools with high concentrations of Latino/a students and English learners*. San Francisco, CA: WestEd.
- Hellman, K., & Nuckles, M. (2013). Expert blind spot in pre-service and in-service mathematics teachers: Task design moderates overestimation of novices' performance. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 35*(35), 2518–2523.
- Herrmann-Abell, C. F., & DeBoer, G. E. (2011). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chemistry Education Research and Practice, 12*(2), 184–192.
- Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis, 29*(2), 95–114.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.
- Hill, H. C., & Chin, M. (2018). Connections between teachers' knowledge of students, instruction, and achievement outcomes. *American Educational Research Journal, 55*(5), 1076–1112.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.
- Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105*(1), 11–30.
- Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education, 38*(3), 258–288.

- Kelcey, B., & Phelps, G. (2013). Considerations for designing group randomized trials of professional development with teacher knowledge outcomes. *Educational Evaluation and Policy Analysis*, 35(3), 370–390.
- Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, 142, 1074–1083.
- Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching*, 54(5), 586–614.
- Kennedy, M. M. (1990). *A survey of recent literature on teachers' subject matter knowledge* (Vol. 90, No. 3). East Lansing, MI: National Center for Research on Teacher Education.
- Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, 4(4), 298–310.
- Koellner, K., & Jacobs, J. (2015). Distinguishing models of professional development: The case of an adaptive model's impact on teachers' knowledge, instruction, and student achievement. *Journal of Teacher Education*, 66(1), 51–67.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716.
- Larkin, D. (2012). Misconceptions about "misconceptions": Preservice secondary science teachers' views on the value and role of student ideas. *Science Education*, 96(5), 927–959.
- Lee, E., & Luft, J. A. (2008). Experienced secondary science teachers' representation of pedagogical content knowledge. *International Journal of Science Education*, 30(10), 1343–1363.
- Legare, C. H., & Visala, A. (2011). Between religion and science: Integrating psychological and philosophical accounts of explanatory coexistence. *Human Development*, 54(3), 169–184.
- Lin, J. W., Chiu, M. H., & Liang, J. C. (2004, April). Exploring mental models and causes of students' misconceptions in acids and bases. *National Association for Research in Science Teaching*.
- Lipowsky, F., & Rzejak, D. (2015). Key features of effective professional development programmes for teachers. *Ricercazione: Six-Monthly Journal on Learning Research and Innovation in Education*, 7(2), 27–53.
- Lockheed, M. E., & Longford, N. T. (1991). School effects on mathematics achievement gain in Thailand. In Raudenbush, S. W., & Williams, J. D. (Eds.), *Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective* (pp. 131–148). San Diego, CA: Academic.
- Lubin, A., Rossi, S., Lanoë, C., Vidal, J., Houdé, O., & Borst, G. (2016). Expertise, inhibitory control and arithmetic word problems: A negative priming study in mathematics experts. *Learning and Instruction*, 45, 40–48.
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources and development of pedagogical content knowledge for science teaching. In Gess-Newsome, J., & Lederman, N. G. (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95–132). Dordrecht, The Netherlands: Kluwer.
- Mann, M., & Treagust, D. F. (1998). A pencil and paper instrument to diagnose students' conceptions of breathing, gas exchange and respiration. *Australian Science Teachers' Journal*, 44(2), 55–60.
- Mareschal, D. (2016). The neuroscience of conceptual learning in science and mathematics. *Current Opinion in Behavioral Sciences*, 10, 114–118.
- Mason, L., Zaccoletti, S., Carretti, B., Scrimin, S., & Diakidoy, I. A. N. (2019). The role of inhibition in conceptual learning from refutation and standard expository texts. *International Journal of Science and Mathematics Education*, 17(3), 483–501.
- Mavhunga, E., & Rollnick, M. (2012). The development and validation of a tool for measuring topic specific PCK in chemical equilibrium. In Bruguère, C., Tiberghien, A., & Clément, P. (Eds.), *E-book proceedings of the ESERA 2011 Conference: Science learning and citizenship*. Part 3, Michelini, M., & Duit, R. (Co-Eds.) (pp. 122–128). Lyon, France: European Science Education Research Association.
- McCrorry, R. (2008). Science, technology, and teaching: The topic-specific challenges of TPCK in science. In American Association of Colleges for Teacher Education (AACTE) Committee on Innovation and Technology (Eds.), *Handbook of technological pedagogical content knowledge (TPCK) for educators* (pp. 193–206). New York: Routledge.
- McDiarmid, G. W. (1988). The liberal arts: Will more result in better subject matter understanding. *Theory into Practice*, 29(1), 21–29.
- Meyer, H. (2004). Novice and expert teachers' conceptions of learners' prior knowledge. *Science Education*, 88(6), 970–983.
- Monk, D. H., & King, J. (1994). Multi-level teacher resource effects on pupil performance in secondary mathematics and science. In Ehrenberg, R. G. (Ed.), *Contemporary policy issues: Choices and consequences in education* (pp. 29–58). Ithaca, NY: ILR Press.
- Morrison, J. A., & Lederman, N. G. (2003). Science teachers' diagnosis and understanding of students' preconceptions. *Science Education*, 87(6), 849–867.
- Moyer-Packenham, P. S., Bolyard, J. J., Oh, H., & Cerar, N. I. (2011). Common features of professional development activities for mathematics and science teachers. *Professional Development in Education*, 37(4), 571–589.
- Murphy, P. K., & Mason, L. (2006). Changing knowledge and beliefs. In Alexander, P. A., & Winne, P. H. (Eds.), *Handbook of educational psychology* (pp. 305–324). Mahwah, NJ: Erlbaum.
- Nadelson, L. S., & Southerland, S. A. (2010). Examining the interaction of acceptance and understanding: How does the relationship change with a focus on macroevolution? *Evolution: Education and Outreach*, 3(1), 82–88.
- Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal*, 40(4), 905–928.
- National Board for Professional Teaching Standards. (1989). *What teachers should know and be able to do*. Retrieved August 2019 from www.nbpts.org/sites/default/files/what_teachers_should_know.pdf
- Nenciovi, L., Allaire-Duquette, G., & Masson, S. (2019). Brain activations associated with scientific reasoning: A literature review. *Cognitive Processing*, 20(2), 139–161. <https://doi.org/10.1007/s10339-018-0896-z>
- Nenciovi, L., Brault Foisy, L. M., Allaire-Duquette, G., Potvin, P., Riopel, M., & Masson, S. (2018). Neural correlates associated with novices correcting errors in electricity and mechanics. *Mind, Brain, and Education*, 12(3), 120–139.
- NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press.
- Odom, A. L., & Barrow, L. H. (1995). Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *Journal of Research in Science Teaching*, 32(1), 45–61.
- Otero, V. K., & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. *Journal of Research in Science Teaching*, 45(4), 497–523.
- Park, S., & Chen, Y. C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, 49(7), 922–941.
- Peterson, P. L., Carpenter, T., & Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlational and case analyses. *Journal of Educational Psychology*, 81(4), 558.
- Pintrich, P. R. (1999). Motivational beliefs as resources for and constraints on conceptual change. In Schnotz, W., Vosniadou, S., & Carretero, M. (Eds.), *New perspectives on conceptual change* (pp. 33–50). Amsterdam: Pergamon.
- Pintrich, P. R., Marx, R. W., & Boyle, R. B. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167–199.
- Polly, D., McGee, J., Wang, C., Martin, C., Lambert, R., & Pugalee, D. K. (2015). Linking professional development, teacher outcomes, and student achievement: The case of a learner-centered mathematics program for elementary school teachers. *International Journal of Educational Research*, 72, 26–37.
- Pugh, K. J., & Girod, M. (2007). Science, art and experience: Constructing a science pedagogy from Dewey's aesthetics. *Journal of Science Teacher Education*, 18, 9–27.
- Rollnick, M., & Mavhunga, E. (2014). PCK of teaching electrochemistry in chemistry teachers: A case in Johannesburg, Gauteng Province, South Africa. *Educación Química*, 25(3), 354–362.

- Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, 8(3), 255–281.
- Sadler, P. M., Coyle, H., Smith, N. C., Miller, J., Mintzes, J., Tanner, K., & Murray, J. (2013a). Assessing the life science knowledge of students and teachers represented by the K–8 national science standards. *CBE—Life Sciences Education*, 12(3), 553–575.
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013b). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020–1049.
- Sadler, P. M., Sonnert, G., Coyle, H. P., & Miller, K. A. (2016). Identifying promising items: The use of crowdsourcing in the development of assessment instruments. *Educational Assessment*, 21(3), 196–214.
- Salinas, A. (2010). Investing in our teachers: What focus of professional development leads to the highest student gains in mathematics achievement? (Unpublished doctoral dissertation). Coral Gables, FL: University of Miami.
- Sanders, L. R., Borko, H., & Lockard, J. D. (1993). Secondary science teachers' knowledge base when teaching science courses in and out of their area of certification. *Journal of Research in Science Teaching*, 30(7), 723–736.
- Scott, P., Asoko, H., & Leach, J. (2007). Student conceptions and conceptual learning in science. In Abell, S. K., & Lederman, N. G. (Eds.), *Handbook of research on science education* (pp. 31–56). Mahwah, NJ: Erlbaum.
- Shi, J., Wood, W. B., Martin, J. M., Guild, N. A., Vicens, Q., & Knight, J. K. (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE—Life Sciences Education*, 9(4), 453–461.
- Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, 8, 118–137.
- Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In Barner, D., & Barron, A. S. (Eds.), *Core knowledge and conceptual change* (pp. 53–72). New York: Oxford University Press.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Sinatra, G. M. (2005). The "warming trend" in conceptual change research: The legacy of Paul R. Pintrich. *Educational Psychologist*, 40(2), 107–115.
- Sinatra, G. M., & Mason, L. (2008). Beyond knowledge: Learner characteristics influencing conceptual change. In Vosniadou, S. (Ed.), *International handbook of research on conceptual change* (pp. 560–582). Dordrecht, Netherlands: Springer.
- Sinatra, G. M., & Taasoobshirazi, G. (2011). Intentional conceptual change: The self-regulation of science learning. In Zimmerman, B., & Shunk, D. (Eds.), *Handbook of self-regulation of learning and performance* (pp. 203–216). New York: Routledge.
- Smith, J. P., III, Disessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, 3(2), 115–163.
- Spiro, R. J., Feltovich, P. J., Coulson, R. L., & Anderson, D. K. (1989). *Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition*. Melville, NY: Cambridge University Press.
- Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, 10(4), 313–340.
- Taasoobshirazi, G., & Sinatra, G. M. (2011). A structural equation model of conceptual change in physics. *Journal of Research in Science Teaching*, 48(8), 901–918.
- Tajudin, N. A. M., Chinnappan, M., & Saad, N. S. (2017). Relationship between mathematics teacher subject matter knowledge, pedagogical content knowledge and professional development needs. In *AIP Conference Proceedings* (Vol. 1847(1), 30001–30007). Melville, NY: AIP Publishing.
- Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, 4(2), 99–110.
- Tsai, C. C. (2003). Using a conflict map as an instructional tool to change student alternative conceptions in simple series electric-circuits. *International Journal of Science Education*, 25(3), 307–327.
- Tsui, C. Y., & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, 32(8), 1073–1098.
- Vosniadou, S. (2012). Reframing the classical approach to conceptual change: Preconceptions, misconceptions and synthetic models. In *Second international handbook of science education* (pp. 119–130). Dordrecht, Netherlands: Springer.
- Vosniadou, S., & Ioannides, C. (1998). From conceptual change to science education: A psychological point of view. *International Journal of Science Education*, 20(12), 1213–1230.
- Wang, J. Y. M. (2018). *Implicit learning in science: activating and suppressing scientific intuitions to enhance conceptual change* (Doctoral dissertation). University of Minnesota, Twin Cities.
- Wilson, S. M. (2013). Professional development for science teachers. *Science*, 340(6130), 310–313.
- Wilson, S. M. (2014). *Michigan State University teacher preparation program TEAC inquiry brief final report*. Retrieved November 2019 from <http://education.msu.edu/teacher-preparation/tpp-program-data/TEAC-Inquiry-Brief-final.pdf>
- Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. East Lansing: Michigan State University.
- Wilson, S. M., Shulman, L. S., & Richert, A. E. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In Calderhead, J. (Ed.), *Exploring teachers' thinking*. Sussex, UK: Holt, Rinehart, & Winston.
- Wind, S. A., & Gale, J. D. (2015). Diagnostic opportunities using Rasch measurement in the context of a misconceptions-based physical science assessment. *Science Education*, 99(4), 721–741.
- Windschitl, M., Thompson, J., & Braaten, M. (2011). Fostering ambitious pedagogy in novice teachers: The new role of tool-supported analyses of student work. *Teachers College Record*, 113(7), 1311–1360.
- Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, National Educational Laboratory Southwest.
- Zhang, X., McInerney, J., & Frechtling, J. (2010). Learning after you know it all: When STEM faculty teach teachers, who learns? *Change: The Magazine of Higher Learning*, 42(3), 24–28.