Challenges and Strategies for Assessing Specialised Knowledge for Teaching

Chandra Hawley Orrill
University of Massachusetts Dartmouth

Ok-Kyeong Kim
Western Michigan University

Susan A. Peters
University of Louisville

Alyson E. Lischka
Middle Tennessee State University

Cindy Jong
University of Kentucky

Wendy B. Sanchez
Kennesaw State University

Jennifer A. Eli
The University of Arizona

Received: 27 April, 2014 / Revised: 30 October, 2014 / Accepted: 9 December, 2014

Developing and writing assessment items that measure teachers' knowledge is an intricate and complex undertaking. In this paper, we begin with an overview of what is known about measuring teacher knowledge. We then highlight the challenges inherent in creating assessment items that focus specifically on measuring teachers' specialised knowledge for teaching. We offer insights into three practices we have found valuable towards overcoming challenges in our own cross-disciplinary work to create assessment items for measuring teachers' knowledge for teaching.

Keywords: teacher knowledge · mathematics' teacher knowledge · measurement · item development

Introduction

The knowledge necessary for effective teaching of mathematics is a topic of increasing interest for many parties (e.g., policy makers, grant funding agencies, mathematics teacher educators). A clear understanding of such knowledge and how to assess it, however, has been elusive. Simultaneous with the drive to understand the nature of teacher knowledge has been an emergence of requirements for measuring such knowledge. For example, for several years, *Math and Science Partnership* grants through the USA Department of Education have required measures of teacher learning for funded professional development. Because of these requirements, and because there are not many instruments readily available for use by researchers and professional developers, project personnel create their own measures of teacher knowledge, with little uniformity across the developed measures (Moyer-Packenham, Bolyard, Kitsantas, & Oh, 2008). Creating high-quality assessments is a complex and difficult task. Measuring teachers' knowledge is particularly challenging for a number of reasons. One set of challenges is related to the multidimensional nature of teacher knowledge as evidenced by the complex design of studies to investigate this knowledge such as the *Teacher Education and Development Study – Learning to Teach Mathematics* (TEDS-M) (Blömeke, Hsieh, Kaiser, & Schmidt, 2014) and *COACTIV: Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy* (Kunter, Baumert, Blum, Klusmann, Krauss, & Neubrand, 2013). These efforts and the research of others (e.g., Ball, Thames, & Phelps, 2008) demonstrate that knowing how to teach content includes not only knowing that content but also knowing how to teach, the kinds of problems students might face in learning, and understanding how one aspect of the content connects to another both within grades and across grade levels. Another set of challenges in creating assessments for specialised knowledge of mathematics is related to complexities of the applied nature of teacher knowledge. That is, if teacher knowledge is framed as a teacher's personal understanding for himself or herself, the resulting assessments will be fundamentally different from those developed based on a framing of teachers' knowledge in action (e.g., Kersting, 2008; Kersting, Givvin, Sotelo, & Stigler, 2010). In all cases, measuring teacher knowledge forces test developers and users to face pragmatic and philosophical challenges.

In this paper, we consider the ways in which the nature of mathematics teacher knowledge impacts item development. Then, we reflect on our own experiences as item developers to highlight particular challenges inherent to measuring teacher knowledge. Finally, we offer insights into three practices we have found to be valuable in overcoming the challenges of developing assessment items. We conclude with thoughts about the state of measurement of teacher knowledge.

Mathematics Teachers' Knowledge

Teacher knowledge has long been a topic of interest to scholars and policy makers. For example, in the United States, teacher education programs became prevalent in the late 19th century (Hansen, 2008) and were based on assumptions that effective teaching requires teachers to have specialised knowledge (Donaghue, 2003), including subject-matter knowledge. After all, it is logical that "student learning depends substantially on what teachers know and can do" (Darling-Hammond, 2000, p. 10).

Despite this focus and the logic of the argument that teacher knowledge matters for student learning, a host of studies in the second half of the 20th century produced contradictory conclusions about whether a relationship exists between teacher knowledge and student achievement. Some researchers found low or insignificant correlation between teacher knowledge of subject matter and student performance (e.g., Begle, 1972; Eisenberg, 1977). Others who relied on analyses of teachers' verbal knowledge rather than mathematics knowledge found positive correlations between teacher knowledge and student achievement (e.g., Boardman, Davis, & Sanday, 1977; Hanushek, 1972). Studies using proxy measures for teacher knowledge, such as the number of mathematics courses completed, also produced inconsistent results (e.g., Begle, 1979; Monk, 1994). Reasons for the inconsistencies may have included limited measures, statistical methods and technologies, and researchers' singular focus on teachers' knowledge of mathematics content rather that the specialised knowledge needed to teach that content.

Recognising that a focus solely on teachers' knowledge of mathematics fails to capture the complexity and multidimensional nature of knowledge needed for teaching, researchers began examining knowledge related to particular topics and knowledge specific to the work of teachers. Shulman's (1987) introduction of seven kinds of teacher knowledge fundamentally shaped our current conceptions of teacher knowledge. Most critical to mathematics education was his introduction of pedagogical content knowledge (PCK) (Shulman, 1986). Shulman's proposition that PCK was specialised knowledge necessary for teachers to teach their content laid the foundation for a variety of theoretical and empirical studies of the knowledge mathematics teachers need (e.g., Ball et al., 2008; Baumert, Kunter, Blum, Brunner, Voss, Jordan & Tsai, 2010; Callingham & Watson, 2011; Fennema & Franke, 1992; Silverman & Thompson, 2008; Thompson & Thompson, 1996).

Defining teachers' specialised knowledge has proven to be extremely difficult; which, in fact, directly impacts our ability to measure that knowledge. Since PCK's introduction, a variety of scholars have suggested additional components be included in PCK, such as discourse knowledge—knowledge about the "culturally embedded nature of inquiry and forms of communication in mathematics" (Hauk, Jackson, & Noblet, 2010, p. 2). Others offered conceptualisations and assessments for content knowledge that extend beyond common mathematical knowledge to deep understandings of the school curriculum content without considering specialised knowledge of mathematics used in teaching (e.g., Krauss, Baumert, & Blum, 2008; Linsell & Anakin, 2012; Maher & Muir, 2013).

Further, studies are beginning to suggest that multiple aspects of teachers' knowledge are interwoven in nature (e.g., Blomeke, Houang, & Suhl, 2011; Shechtman, Roschelle, Haertel, & Knudsen, 2010), which further complicates our ability to understand and measure this knowledge. One well-known and widely measured construct in this domain, mathematical knowledge for teaching (MKT) (Ball, Lubienski, & Mewborn, 2001; Ball et al., 2008), focuses on both the subject matter knowledge necessary to teach and the knowledge necessary to teach the subject matter including knowledge of teaching mathematics, knowledge of students' mathematical learning, and knowledge of curriculum (Ball et al., 2008). Similarly, the unsettled debate about whether the knowledge that matters is in the head or knowledge enacted in context (e.g., Fennema & Franke, 1992; Hodgen, 2011; Petrou & Goulding, 2011) has significant implications for the study and measure of this knowledge. This lack of a single conception of the specialised knowledge of teaching has made studying and measuring teacher knowledge very difficult.

Despite the difficulties of measuring teacher knowledge, there are studies that suggest further consideration of specialised knowledge is promising. For example, in one study of elementary teachers' MKT, a significant relationship was shown between teacher knowledge and student performance on assessments (Hill, Rowan, & Ball, 2005). In their ongoing video analysis work, the *Learning Mathematics for Teaching Project* [LMT] researchers (2011) have also found a direct relationship between teachers' performance on written assessments of MKT and observations of their teaching practices. In a different line of research, Baumert et al. (2010) found that PCK was not only distinguishable from regular content knowledge (CK), but also that higher levels of PCK correlated to higher levels of student performance on assessments. Studies like these suggest that, as a community, we are beginning to understand what may be important to measure if we want to understand the relationship between teachers' knowledge and student performance. However, there is still much to learn.

Issues Around Developing Items to Measure Teacher Knowledge

As evidenced by our discussion thus far, we assert that particular difficulties in developing teacher assessments arise from distinct characteristics of, and limitations in, the field's understanding of teacher knowledge. To this end, we briefly explore four aspects of teacher knowledge in relation to item construction: the multidimensional nature of teacher knowledge; the characteristics of knowledge worth understanding (e.g., knowledge in action); the general lack of understanding of teacher knowledge in many areas of the domain; and the lack of a theoretical or empirical developmental trajectory to support the development of assessments that would successfully capture change over time.

Teacher knowledge is necessarily multidimensional and interrelated. As indicated by various models of teacher knowledge—such as the MKT framework (e.g., Ball et al., 2008), the knowledge quartet (e.g., Rowland, Huckstep, & Thwaites, 2005), and the conceptual framework utilised by TEDS-M (e.g., Beswick & Goos, 2012; Blömeke et al., 2014)—to successfully teach mathematics, a teacher needs to know far more than just the content. For example, the teacher needs to know the content in a way that supports the engagement of students in learning the content, including multiple strategies for supporting a wide variety of learners as well as the mathematical foundation necessary to interpret the developmental work of those learners. Teacher knowledge also extends into supporting students in communicating about mathematics. This is more than PCK (Shulman, 1986), as indicated by a small body of research that suggests that the nature of discourse in the mathematics classroom fundamentally impacts student learning (e.g., Kazemi & Stipek, 2001; Wood, Williams, & McNeal, 2006). There are many concepts to know and measure.

For assessment development, the multidimensional nature of teacher knowledge raises two important issues. The first issue is *what* to measure. Is it adequate to measure content knowledge? Is it adequate to measure content knowledge as demonstrated through interpretation of student responses? Which kind of knowledge is most important?

The multidimensional nature of knowledge also raises issues about *how* to measure knowledge. Is it more important to measure knowledge in situ as the teacher is working with students, in a situation that tries to mimic the actual classroom (e.g., Kersting et al., 2010), or to use paper and pencil assessment of mathematics knowledge (e.g., LMT, 2011)? Or, are paper and pencil assessments of knowledge adequate when they are developed to measure particular aspects of the specialised knowledge teachers need? In the past decade, a number of such paper and pencil instruments have been produced through large-scale assessment development efforts including the *Learning Mathematics for Teaching* (LMT) instruments (LMT, 2011); *Diagnostic Teacher Assessments in Mathematics and Science* (DTAMS) (Saderholm, Ronau, Brown, & Collins, 2010); and *Knowledge of Algebra for Teaching* assessments (Ferrini-Mundy, McCrory, & Senk, 2006; McCrory, Floden, Ferrini-Mundy, Reckase, & Senk, 2012).

The second aspect of teacher knowledge to consider in item construction is intrinsically tied to multidimensionality. It is consideration of what aspects of knowledge are the most worth measuring. Although most recent assessments of teacher knowledge have chosen to focus on knowledge that can be measured with pencil and paper (e.g., Bradshaw, Izsák, Templin, & Jacobson, 2014; Ferrini-Mundy et al., 2006; Hill, Ball, & Schilling, 2008; Izsák, Orrill, Cohen, & Brown, 2010; Kim & Remillard, 2011; Saderholm et al., 2010; Shechtman et al., 2010), there are some compelling arguments for other kinds of measures. For example, one strong argument supports only measuring the knowledge teachers use in action. This argument is based in the idea that the only knowledge a teacher has that matters to student learning is the knowledge enacted in the classroom (Kersting, 2008; Kersting et al., 2010; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012). These arguments are supported by studies that engage teachers in more authentic activities such as analysing student thinking.

The question of what knowledge matters is critical for assessment development because the format of the assessment and the questions on which it focuses are intrinsically intertwined with the knowledge to be measured (e.g., Orrill & Cohen, in press). For example, if we assert that the only teacher knowledge worth measuring is the knowledge teachers draw upon in their classroom teaching, then a paper and pencil assessment of mathematical skills may not help answer questions about the knowledge teachers need in practice. Some knowledge may remain inert in a traditional assessment, whereas something more like a performance assessment may invoke critical understandings (e.g., Kersting, Givvin, Sotelo, & Stigler, 2002).

The above aspects are exacerbated by the general lack of understanding of teacher knowledge across a wide variety of domains within mathematics. This lack of understanding is particularly noticeable outside of number concepts in K–8 and algebra in high school. Further, even in domains such as rational numbers, for which a large knowledge base exists, the research has tended to focus on misconceptions and missing understandings (e.g., Ma, 1999; Post, Harel, Behr, & Lesh, 1988; Riley, 2010). As a field, we know little about how teachers understand the content they teach, how and whether they understand its placement in the larger body of mathematics, and what ideas they know well. One of the reasons for assessing teacher knowledge is to drive professional development; yet, as a field, we need a better and deeper understanding of teacher knowledge to provide sufficient information to support professional development.

The fourth aspect of teacher knowledge that makes its measurement particularly difficult is the lack of a clear trajectory of development for teacher knowledge. Because the field does not have an established theory of what a reasonable development of knowledge might look like for teachers, we are limited in the kinds of assessments we can create. For example, without a clear image of what kinds of knowledge develop at various stages of a teaching career, we can only look at a snapshot of the teacher in time as compared with a hypothetical "best" teacher. This leaves us without solid tools to position teachers within a trajectory of development and limits our abilities to support the teachers in further developing their understandings over time.

The four aspects of teacher knowledge discussed above highlight the need for mathematics educators to conduct additional studies to guide the development of measures of teacher knowledge. Such knowledge is one factor in a complex system that influences what happens in classrooms. Curriculum, teacher beliefs, discourse, affect, student knowledge, motivation, culture, and many other factors combine with teacher knowledge to impact student learning in classrooms (cf. Shechtman et al., 2010; Simon, 1997).

As researchers seek understanding of connections between knowledge and practice, well-developed measures of teacher knowledge are a necessary component of that research.

Creating Items for Assessments of Teacher Knowledge

At the heart of teacher knowledge assessments are the individual items that comprise the assessments. It is in these items that test developers embed their beliefs about the knowledge that matters and it is through these items that teachers demonstrate what they know about aspects of teaching mathematics. In this section, we introduce some of the key challenges of writing items to measure mathematics teachers' knowledge, based on the analysis of failed items as well as good items.

Too often, items are not made available for examination by other researchers. This is reasonable given that releasing items in scholarly writing limits the lifespan of the items due to the target audience of the assessment gaining access to the items. Items are necessarily kept secure to be used and reused. The lack of item disclosure, however, creates a unique challenge as it leaves the field without templates for item development and without the benefit of learning from the work of others. We present a variety of items with analyses to help fill this gap in available items. This section concludes with suggestions for overcoming some of the challenges in item development.

Challenges in Item Development

One useful way to frame thinking about item writing is through examination of items that were not successful. Items fail for various reasons, including proving to be too easy or too hard, measuring something other than the target construct, lacking one or more clearly correct answers, lacking mathematical precision, or incorporating vague language. One or more of these reasons can result in a measure with low validity, meaning that the items do not help predict whether a teacher is knowledgeable on a particular target construct. Teachers with deep understandings of the target construct may answer the items incorrectly, whereas teachers with insufficient knowledge may answer correctly. In the end, these items do not contribute to accomplishing the purpose of the overall measure.

Based on our experience as item writers and our knowledge of item-writing efforts, we have identified five main challenges in writing assessment items intended to measure specialised knowledge for teaching mathematics:

1. creating items with appropriate difficulty levels;
2. creating items for the target constructs;
3. using precise language;
4. incorporating pedagogical concerns appropriately; and
5. writing clear stems and distracters.

We elaborate on each of these challenges below in order to address issues and complexities of developing items.

These challenges, however, are closely interrelated and do not arise in isolation. Item development inevitably involves multiple inter-related challenges; we focus on item-level issues in this paper as one step towards informing the design of well-developed measures of teacher knowledge. Challenges of designing a measure (a set of items) involve much more complexity.

Creating items with appropriate difficulty levels. Creating items at the appropriate difficulty level is a challenge when measuring teachers' knowledge. After all, when we measure teachers' knowledge of the mathematics they teach, we are often measuring elementary or middle school mathematics. If items are too easy or too hard, they fail to discriminate among teachers in terms of the traits being measured. For example, if an item is too easy, most teachers will answer correctly thus obscuring the item's ability to distinguish variability in teachers' knowledge. Creating items that are too easy has been a common issue for assessment development projects (e.g., Hill, Schilling, & Ball, 2004; Kim & Remillard, 2011). We assert that this is related to the lack of clear understanding of the construct of teacher knowledge and the lack of a clear learning trajectory for teacher knowledge. Thus, finding ways to make K–12 content appropriate for teacher assessments is one key challenge.

The item below presents an item from Geometry Assessments for Secondary Teachers (GAST), which was designed with an intention to assess teachers' knowledge for teaching geometry (see Figure 1). This item was quite easy for teachers as evidenced by the fact that 95% of the responses were correct.

> A teacher is preparing to teach a unit on the similarity of triangles. She wants to create an assessment to make sure students have the necessary prerequisite knowledge for learning about similarity in triangles.
>
> Which mathematics topic is critical for understanding relationships among similar triangles?

A. The sum of the angles of a triangle is 180 degrees
B. Parallel lines intersect lines proportionally
C. The formula for the area of a triangle
D. The meaning of ratios and proportions

Figure 1. Sample item that is easy.

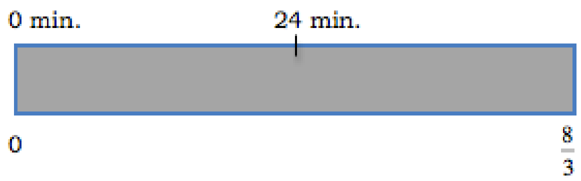
Moreover, this item produced an item-total correlation of -0.022, which means that it did not correlate strongly with teachers' overall scores on the assessments that were comprised of an overall set of harder questions. In short, this item was failing to contribute to the measurement of the target construct (i.e., whether the teacher recognises, describes, and assesses critical prerequisite knowledge) in the overall assessment.

Creating items for target construct. As we have already established, there are fundamental challenges in defining the construct of teacher knowledge in ways that can be measured. This is compounded by the difficulty of developing an item that measures some aspect of that construct after it is defined. Conceptualising the knowledge construct and developing items to measure it are interrelated, iterative processes (e.g., Bradshaw, Izsák, Templin, & Jacobson, 2014; Kim & Remillard, 2011).

Mr. Vargas is teaching some challenging problems to his students. One problem asks:

> Mr. Compton drives the same route from home to work every morning. One week he had to go to a training course that required him to drive 8/3 of the distance that he usually drives to work. He noticed that 24 minutes had passed when he had driven half way to the training course. How long does it take Mr. Compton to drive to his regular work place? Assume that for all trips he travels at the same constant speed.

One of Mr. Vargas's students has modeled the situation by drawing two number lines side-by-side, as shown below, but tells Mr. Vargas that he is stuck. Can you determine the time that Ms. Compton usually drives to work?



Choices:

A 6 minutes.

B 9 minutes.

C 18 minutes.

D 16 minutes.

E Cannot be determined.

Figure 2. Sample item that does not measure target construct.

Depending on the statistical model being used, each item should be designed to measure one or more particular constructs. When the target construct is based on a web of a few interrelated subcomponents, however, creating an item with a definite correct answer is extremely hard. For example, the *Improving Curriculum Use for Better Teaching* (ICUBiT) project team members developed an assessment to measure teacher knowledge of mathematics embedded in curriculum materials (Kim & Remillard, 2011). The assessment measured four major constructs including *surrounding knowledge*—knowledge of how a particular mathematical goal is situated within a set of foundational and future concepts. When an item is about prerequisite or prior concepts needed to learn a particular concept, determining why one answer choice is definitely correct and others are not is difficult because learning is not linear and various previous concepts and ideas are related to future learning. In fact, more than half of the items that were created to target this construct failed. Most of them yielded item-total correlations that were negative or very close to zero.

Even when the target construct is very clear, items can measure an unintended concept. A teacher can choose a correct answer for reasons other than those anticipated. For example, in the *Diagnosing Teachers' Multiplicative Reasoning* (DTMR) project, one key construct being measured was proportional reasoning. Items were developed with the aim of engaging teachers in reasoning about quantities. Yet, teachers unexpectedly applied algorithmic thinking to situations such as the one presented in Figure 2, which obscured the item's ability to measure reasoning about proportions. Rather than comparing the quantities, some participants set up an equation such as $\frac{8}{3d} = 48$ minutes and used algebraic rules to solve for d. Although this an acceptable path to the correct response, it did not highlight the teacher's reasoning about relationships between quantities, which was the goal of the assessment. When focusing o teachers, it is difficult to create items in which algorithms can be used because those algorithms con obscure the understandings of interest to test developers.

Teachers can also select correct answers using reasoning that reveals incomplete knowledge. For example, in the ICUBiT project, an item asking for a division story problem in which the product and the number of groups are known and the number in each group is unknown (i.e., partitive/sharing division), teachers tended to choose a correct answer based on familiarity with the context or situation given in the distracters. A common reason why a choice was not selected was "This doesn't look like what we usually see in the textbook." By contrast, one of the reasons why teachers chose a correct answer was "This context looks very familiar to my students". In such cases, items need to have an equal number of familiar or unfamiliar choices, or all very similar choices in order to truly measure the intended construct. The distracters for a division story problem using a *partitive* meaning were finally modified to those shown in Figure 3.

(a) If there were 46 jellybeans to share equally between 3 friends, how many would each friend have?
(b) There are 46 jellybeans and they were placed into equal-sized piles. How many jellybeans are in each pile?
(c) There are 46 jellybeans. Each student will receive 3 jellybeans. How many students will receive 3 jellybeans?
(d) There are 46 jellybeans to be distributed evenly amongst a group of friends. If each person gets 3 jellybeans, how many friends get jellybeans and how many jellybeans are left over?

Figure 3. Choices for a division story problem using a partitive meaning.

All answer choices include jellybeans in an equal group context, which reduces the possibility of relying on context familiarity to answer the item correctly. Note that choices (c) and (d) are in the context of measurement division. Although both choices (a) and (b) use a partitive division context, choice (b) does not specify the number of groups. The intention was not to include a specific number of groups in choice (b) in order to make it different from choice (a) while having the same context of partitive division as in item (a).

Items that allow multiple interpretations also fail. Diverse interpretations support various answers and cause lack of one clearly correct answer. For example, in the assessment developed by the ICUBiT project, the item displayed in Figure 4 asks for a mathematical reason to use 1,800 after finding the longest multiplication combination (i.e., a multiplication string with the prime factors of a number) of 180. All of the choices have plausible support in some way, even though the expected answer was (c). Moreover, all the choices are related to each other. Such reasons make it hard to argue that (c) is the one clearly correct answer.

> Of the following, which is the best reason for choosing the number 1,800 in the next task (as a follow-up task of finding the longest multiplication combinations of 180) students are asked to work on?
>
> (a) To see if students understand multiplying by tens.
> (b) Because 1,800 is a multiple of 10.
> (c) Because 1,800 is ten times 180.
> (d) To see how the longest multiplication combinations of multiples are related.
> (e) 1,800 will be accessible to students because it is a round number.

Figure 4. Sample item that involves multiple interpretations.

Teachers' diverse interpretations and perceptions are more prevalent when items embed pedagogical situations in them. This will be further discussed later when we address using pedagogical concerns in items.

Using precise language. Items that measure teacher knowledge simultaneously involve the three distinct languages: of mathematics; psychometric, and pedagogy. Items are written in various combinations of language, and incorporating these three languages into items properly is another challenge for item writers.

Here, we particularly focus on the language of mathematics and later discuss the language of pedagogy. We do not address psychometric language in this paper because issues related to psychometric knowledge apply to the general development of assessment items and not just items for measuring teacher knowledge. (For discussions about some of the issues surrounding psychometric language, see, for example, Frey, Petersen, Edwards, Pedrotti, and Peyton, 2005; or see Haladyna, Downing, and Rodriguez, 2002.)

Maintaining mathematical precision. In developing items to measure teacher knowledge, we have found that there is a need to maintain mathematical precision, but that precision sometimes becomes problematic for accurately measuring teacher knowledge. Clearly, there is a need for items designed to measure knowledge for teaching mathematics to use precise mathematical language. For example, in an item about selecting the correct longest multiplication combination for the expression $924x^2$□ from among various multiplication combinations, the item is not mathematically accurate or rigorous without specifying that x and y are prime numbers. Otherwise, all the options provided may not be appropriate. Another example of the need to use precise language can be seen in the difference between statements such as "If the longest multiplication combination of a number has a 3 or 5, then it is an odd number" and "If and only if the longest multiplication combination of a number has a 3 or 5, then it is an odd number." These subtle language differences can create opportunities for assessing different aspects of teacher knowledge.

A related problem arises when item developers want to pursue mathematical ideas that are not typical. For example, in developing the DTMR assessments, one of the goals was to uncover how teachers understand the relationship between fractions and ratios (Izsák, Lobato, Orrill, & Jacobson, 2010). To this end, the developers created items that used addition of ratios to explore one aspect that differentiates fractions and ratios. When they asked mathematicians about this, the mathematicians were extremely uncomfortable with using fraction notation to show addition of ratios (for example, $\frac{2}{4}+\frac{3}{7}=\frac{5}{11}$). This opened up an interesting dialogue around how to be mathematically accurate and acceptable while still pursuing the attribute of interest.

Another issue with precision arises when teachers rely on their interpretation of students' understandings and their considerations for how they teach to shape the decisions they make on assessments. For example, teachers may use both "Division by zero is undefined" and "It is not possible to divide by zero" in their teaching, thus obscuring their ability to differentiate between the two phrases on an assessment task. Similarly, we found that teachers were not comfortable with statements using "all," "always," "only," and "never." For example, for items from the *Does it Work* project (Orrill, Izsák, & Cohen, 2006), using such words led teachers to reject correct answers simply because they teach their students to be wary of statements about things being "always" correct or incorrect. As a result, teachers were attending to an understanding of mathematics that was not the construct of interest.

In *Does it Work*, about one third of interviewed teachers selected an answer that they knew was mathematically less precise to explain why cross multiplication works. Their rationale for this decision was that they would never teach the more precise definition to their middle school students, but they would teach those students the less precise explanation. This raises significant challenges for measuring teachers' understanding, as teachers may clearly understand the mathematics but use other priorities to select the answers on assessments.

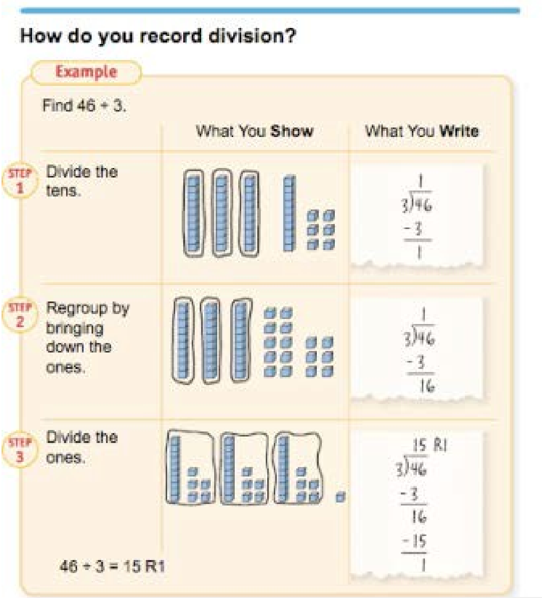
Using mathematically incorrect statements as distracters can be effective for measuring teacher knowledge. For example, in the ICUBiT project and the DTMR project, some distracters were generated from teacher responses in pilot interviews or made to look more mathematical through the use of mathematical terms in distracters. However, mathematically incorrect statements need to be used carefully in order to ensure the instrument is capable of producing valid measures. Item developers want to ensure that teachers are selecting responses—whether correct or incorrect—for reasons aligned with item intentions and not because teachers were lured or tricked into particular responses. For example, in the ICUBiT project, a statement that addition and multiplication are inverse operations was created as a distracter along with other statements about inverse relationship between addition and subtraction and inverse relationship between multiplication and division. After a careful examination and discussion in the team, the choice was removed because the statement was not mathematically accurate, although it would potentially be attractive to teachers with low knowledge and thus help to discriminate knowledge among teachers. The rationale was to have precise mathematics on the test, regardless of a correct or incorrect choice.

Thus language and symbols can play an important role in shaping a teacher's response. Attending to language that is accessible and acceptable to teachers is as important as mathematical precision in creating an instrument with reliable measures. Attending to the work that teachers do, and not just the mathematics, is also important for developing items that accurately measure teacher knowledge.

Incorporating pedagogical concerns appropriately. As we have discussed thus far, measuring knowledge for teaching mathematics is different from measuring traditional content knowledge. Teacher knowledge is situated in teaching and thus measuring teacher knowledge logically requires a wide variety of pedagogical considerations to be incorporated in assessments. For example, LMT measures of MKT include items that require teachers to examine various student strategies that are not commonly used and determine whether the strategies are mathematically sound (e.g., Hill, Schilling, & Ball, 2004). To assess mathematical understandings for teaching content, the ICUBiT project also uses excerpts from a range of curriculum programs that vary in terms of their pedagogical stances.

The DTMR project found, in the development of their proportional reasoning assessment, that using contextualised situations requiring teachers to analyse student thinking or respond to peers helped focus teachers on the specific construct of interest because sample student work could engage teachers in particular kinds of mathematical reasoning that were of interest to the research team. Thus, incorporating pedagogical aspects such as analysing student thinking fundamentally shaped the kinds of questions asked and the ways in which teachers reasoned about the mathematics of interest.

As mentioned earlier, however, it is difficult to coordinate pedagogical issues, situations, and representations in items primarily because these situations often are not uniformly interpreted and may allow multiple perspectives. There are also concerns about the universality of particular pedagogical approaches. For example, the LMT developers avoided including items in their assessments that could be biased toward teachers who use inquiry strategies in their teaching (Hill et al., 2005). In contrast, in the DiW assessment, the developers explicitly used particular drawn representations uncommon in current teaching materials because they believed that those representations provided particular insights into the ways in which teachers understood content deeply (Izsák et al., 2010).



Which of the following best describes the role of place-value blocks in this model of division?

(a) Place-value blocks show how many groups of 3 can be made from 46.

(b) Place-value blocks make it easier to count the groups of 10 in the dividend.

(c) Place-value blocks show where the remainder comes from in the long division algorithm.

(d) Place-value blocks show that the division algorithm begins grouping the largest place and proceeds to the smaller places.

Figure 5. Sample item that includes unclear pedagogical concern.

One issue with the inclusion of pedagogical aspects in assessments of mathematics teacher knowledge is that it can make determining one correct answer among all the given choices difficult. Consider the sample item from the ICUBiT project shown in Figure 5. The item asks which choice best describes the role of the model that is used to illustrate the represented mathematical idea in the item (i.e., partitive interpretation of division – 46 divided into three groups).

In Figure 5, all answer choices except choice (a) describe ways in which the blocks could possibly be used. In fact, there was a debate among the project members when creating this item. The intention of an item may be clear to the designers, but clarity is not certain until item analysis based on teacher responses ensures the appropriateness of the item. However, if an item such as that in Figure 4, includes controversial issues at the development stage, it is not likely that the item will survive the validation process (Messick, 1989).

In an attempt to communicate the intention of the item designers clearly, the item in Figure 5 was revised to require teachers to think specifically about how place-value blocks can be used to divide 46 into three equal groups. Although various ways of using the manipulatives in the division context are available, the focus of the item was on how the place-value blocks represent the overall approach of the long division. It was found in the field test, however, that the change was not successful, which demonstrates the difficulty of designing an item incorporating pedagogical concerns appropriately.

Writing clear stems and distracters. All the challenges addressed previously are related to designing both a good item stem and a set of solid distracters. Good stems and distracters are the products of efforts to coordinate critical factors of designing items (e.g., Frey et al., 2005; Haladyna et al., 2002). A good stem is a clearly stated question that provides information necessary to answer the question. Determining the appropriate amount of information and context given in the item is important because it offers the context of the question and sets a boundary for teachers' thinking and approaches that may be used in answering the question. This involves choice of wording, mathematical focus, pedagogical context in which the mathematics is embedded, and psychometric strategies, to list a few.

Next, items need good distracters, particularly those that look correct, but are not. Creating good distracters requires a detailed examination of content and a range of related factors. For example, an item from the ICUBiT project relates two solution methods, $a(b+c)$ and $ab+ac$ (see Figure 6). The item is about the distributive property of multiplication over addition, which is not mentioned in the given excerpt. Yet, the excerpt includes solution methods that produce the same answer, and the distributive property relates the two methods (i.e. $a(b+c) = ab+ac$).

What fundamental mathematical idea provides the basis for why the two solution methods produce the same answer?

(a) Commutative property
(b) Relationship between addition and multiplication
(c) Multi-step problem solving
(d) Distributive property
(e) Order of operation

Figure 6. Sample item in progress of refinement.

The original stem was "Which of the following is the most fundamental mathematical idea that the above excerpt illustrates?" The stem did not make clear whether the question intended to illustrate the mathematical idea to the teacher (who answers the question) or to a student. The item does not provide any additional information on whether students will explore the property after examining the two methods. Therefore, the basis for deciding mathematical importance is not clearly indicated in the stem. Such ambiguity led to the stem's revision to make it more specific about the relationship between the two solution methods among all other mathematical ideas, as shown in the figure.

Considering the distracters, choice (c) is not a mathematical idea. In fact, this was one of the popular choices selected by teachers during pilots, which indicated that teachers did not understand the focus of the item. Given the circumstances, that choice was eliminated.

Through the use of an iterative process that included teacher interviews on the items, the stem and distracters were refined.

Ways to Overcome the Challenges

To conclude our discussion on writing items for measuring teacher knowledge, we provide three important insights learned from across an array of assessment-writing efforts to help address the challenges above. These insights are aligned with best practices literature in measurement regarding the design of assessment items, including *Standards for Educational and Psychological Testing* (American Educational Research Association, 1999), *Knowing What Students Know: The Science and Design of Educational Assessment* (National Research Council, 2001), and *ETS International Principles for Fairness Review of Assessments* (Educational Testing Service, 2009). Our conclusions highlight how general guidelines for item writing apply in a mathematics teacher knowledge context.

When one looks across the five challenges above, it is clear that there is much to attend to in writing items. We have found that one of the most important ways to ensure attention to each of the five challenge areas is to create an interdisciplinary team to write the assessment. Item writing requires significant input from across a variety of fields in order to produce high-quality items (Manizade & Mason, 2011). Having input from across fields can help to address all five challenges above. The assessments described in this article were created by teams that included psychometricians, mathematics educators, mathematicians, and teachers. Although each expert brings a unique perspective, there is a need for the synergy of having input from all four disciplines. Psychometricians can help to shape the assessment by assisting the writing team in understanding particular needs for particular psychometric models. For example, some psychometric models allow multiple attributes of a domain to be measured by a single item while other models allow only one attribute per item. Psychometricians also know general principles for assessment item development, thus providing guidance on the development of stems and distractors. Mathematics educators have a rich understanding of teaching and learning, allowing them to focus on item difficulty, pedagogical concerns, and the items that can help measure a particular construct. Mathematicians have a deep vested interest in precision and can provide insights into vertical alignment, precision, norms, and other important content elements. Teachers can draw on their mathematics teaching experience at the target content level to bring a practitioner perspective for writing items with realistic classroom and pedagogical contexts in mind. Whether one or more teachers are on the development team, teachers need to be a part of the item development process to ensure item authenticity. Using a team of people with expertise in multiple domains can lead to better anticipation of possible problems, such as those of language use, before items are ever tested.

A second critical approach to meeting the challenges outlined above is the use of clinical interviews. These are essential for establishing construct validity and they have been used in many development projects (e.g., Bradshaw et al., 2014; Hill et al., 2005; Izsák et al., 2010; Kim &

Remillard, 2011). Clinical interviews allow developers to understand whether an item is measuring the appropriate construct and whether it is at an appropriate difficulty level by providing insight into participants' thinking. Clinical interviews also provide important insights into the ways in which participants understand the stem and distracters. After a participant has completed potential items, the person is typically asked to explain each item and their reasoning process. Clinical interviews highlight whether participants: 1) interpret items in ways that were intended; and 2) use the reasoning intended to be measured. More informally, the clinical interviewing process can also illuminate item difficulty levels and whether extraneous concepts are measured. As mentioned earlier, clinical interviews can also provide the basis for developing distracters, as participants may introduce typical misconceptions or other errors in their explanations. To capitalise on the advantages of clinical interviews, some research teams begin with open-ended items and use clinical interviews to generate potential responses (e.g., ICUBiT and DTMR projects).

The third strategy for addressing the challenges of item writing is to adopt an iterative process to item writing. While this is time-consuming and requires patience and resources, it is an approach likely to yield a strong item pool. Items need to be piloted and revised several times, paying attention to the challenges mentioned above, to ensure a high-quality assessment. At various stages, the items should be systematically analysed by experts (for content validity) and should have their construct validity determined through clinical interviews. This ensures that the appropriate constructs are being measured.

In order to finish assessment development, it is also necessary to collect reliability data on the items and, in most cases, item difficulty data. The item difficulty data provides another tool for determining whether items are at the appropriate level of difficulty. Using matrices or other graphic organisers can help to illuminate the effectiveness and level of difficulty of an item even before there are a sufficient number of responses to perform statistical analyses. Such tools can provide an easy way to determine whether a response is correct with incorrect reasoning, which helps determine the effectiveness of an item and whether the distracters are viable. Allowing the time to develop the items using an iterative process substantially impacts the likelihood of the resulting assessment being an effective tool for measuring the desired body of knowledge.

Conclusion

Teacher knowledge matters and measuring it matters more than ever. The current era of standards and accountability, such as implementation of the *Common Core State Standards for Mathematics* (Council of Chief State Schools Officers (CCSSO), 2010) and the *Race to the Top* program in the United States, places student achievement at the centre of attention and has renewed efforts to establish the relationship between teacher knowledge and student achievement. However, there are still substantial hurdles to our development of such connections. As discussed here, there is still more work to be done in the area of conceptualising teacher knowledge. And, there are still many challenges to developing instruments that are capable of producing valid and reliable measures. In this article, we have pulled from a number of item development efforts to highlight some of the particular challenges to developing sound items.

Further work is needed in moving from item development to assessment construction. When considering an entire assessment, additional considerations compound. These range from questions about the scope of the instrument to the amount of time that it can take to the psychometric models that will be applied. Every aspect of measuring teacher knowledge clearly benefits from interdisciplinary collaboration.

Acknowledgments

The work reported here was funded, in part, through a conference grant from the National Science Foundation to Andrew Izsák, Janine Remillard, and Allan Cohen. The work reported herein expresses the views of the authors and may not necessarily reflect the views of the National Science Foundation. The authors wish to thank the conference organisers for hosting the *Interdisciplinary Conference on Assessment in K-12 Mathematics: Collaborations Between Mathematics Education and Psychometrics*. This conference brought the authors together around a common interest.

The authors also thank Mary Garner for her feedback on an earlier draft of this report.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Authors.

Ball, D. L., Lubienski, S., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). Washington, DC: American Educational Research Association.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*, 389–407.

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.

Begle, E. G. (1972). *Teacher knowledge and student achievement in algebra (School Mathematics Study Group Report No. 9)*. Palo Alto, CA: Stanford University.

Begle, E. G. (1979). *Critical variables in mathematics education: Findings from a survey of the empirical literature.* Washington, DC: Mathematical Association of America and National Council of Teachers of Mathematics.

Beswick, K., & Goos, M. (2012). Measuring preservice primary teachers' knowledge for teaching mathematics. *Mathematics Teacher Education and Development, 14*(2), 70–90.

Blömeke, S., Houang, R.T., & Shul, U. (2011). TEDS-M: Diagnosing teacher knowledge by applying multidimensional item response theory and multiple-group models. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 4*, 109–129.

Blömeke, S., Hsieh, F-J., Kaiser, G., & Schmidt, W. H. (Eds.) (2014). *International perspectives on teacher knowledge, beliefs, and opportunities to learn: TEDS-M results.* Dordrecht, The Netherlands: Springer.

Boardman, A. E., Davis, O. A., & Sanday, P. R. (1977). A simultaneous equations model of the educational process. *Journal of Public Economics, 7*, 23–49.

Bradshaw, L. Izsák, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understanding of rational numbers: Building a multidimensional test within the diagnostic classification framework. *Educational Measurement: Issues and Practices, 33*(1), 2–14.

Callingham, R., & Watson, J. (2011). Measuring levels of statistical pedagogical content knowledge. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics – Challenges for teaching and teacher education: A joint ICMI/IASE study.* Voorburg, The Netherlands: International Statistics Institute.

Council of Chief State Schools Officers (CCSSO) (2010). *Common core state standards for mathematics*. Retrieved June 2, 2010, from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 1–42.

Donoghue, E. F. (2003). The emergence of a profession: Mathematics education in the United States, 1890-1920. In G. M. A. Stanic, & J. Kilpatrick (Eds.), *A history of school mathematics* (Volume 1, pp. 159–193). Reston, VA: National Council of Teachers of Mathematics.

Educational Testing Service. (2009). *ETS international principles for fairness review of assessments: A manual for* *developing locally appropriate fairness review guidelines in various countries*. Lawrence Township, NJ: Educational Testing Service. Retrieved from www.ets.org/s/about/pdf/fairness_review_international.pdf

Eisenberg, T. A. (1977). Teacher knowledge and student achievement in algebra. *Journal for Research in Mathematics Education, 8*, 216–222.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147–164). Reston, VA: National Council of Teachers of Mathematics.

Ferrini-Mundy, J., McCrory, R., & Senk, S. (2006, March). *Knowledge of algebra teaching: Framework, item development, and pilot results*. Paper presented Research symposium at the research pre-session of NCTM annual meeting. St. Louis, MO.

Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education, 21*, 357–364.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*, 309–334.

Hansen, D. T. (2008). Values and purpose in teacher education. In M. Cochran-Smith, K. E. Demers, S. Feiman-Nemser, & D. J. McIntyre (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (3rd ed., pp. 10–26). New York: Routledge.

Hanushek, E. A. (1972). *Education and race: An analysis of the educational production process*. Lexington, MA: D. C. Heath.

Hauk, S., Jackson, B., & Noblet, K. (2010, February). No teacher left behind: Assessment of secondary mathematics teachers' pedagogical content knowledge. In S. Brown (Ed.), *Proceedings of the 13th conference on Research in Undergraduate Mathematics Education* held in Raleigh, NC. Electronic proceedings.

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*, 372–400.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105*, 11–30.

Hodgen, J. (2011). Knowing and identity: A situated theory of mathematics knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 27–42). Dordrecht, The Netherlands: Springer.

Izsák, A., Lobato, J., Orrill, C. H., Jacobson, E., (2010). *Diagnosing teachers' multiplicative reasoning attributes.* Unpublished report, Department of Mathematics and Science Education, University of Georgia, Athens, GA.

Izsák, A., Orrill, C. H., Cohen, A. S., & Brown, R. E. (2010). Measuring middle grades teachers' understanding of rational numbers with the mixture Rasch model. *The Elementary School Journal, 110*(3), 279-300.

Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *Elementary School Journal, 102*, 59-80.

Kersting, N. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2002). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education, 61*(1–2), 172–181.

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement, 68*, 845–861.

Kersting, N. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education, 61*, 172–181.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–589.

Kim, O-K., & Remillard, J. T. (2011, April). *Conceptualizing and assessing curriculum embedded mathematics knowledge*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM Mathematics Education, 40*, 873–892.

Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (Eds.), (2013). *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV Project*. New York: Springer.

Learning Mathematics for Teaching Project. (2011). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education, 14*, 25–47.

Linsell, C., & Anakin, M. (2012). Diagnostic assessment of preservice teachers' mathematical content knowledge. *Mathematics Teacher Education and Development, 14*(2), 4–27.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maher, N., & Muir, T. (2013). "I know you have to put down a zero, but I'm not sure why": Exploring the link between pre-service teachers' content and pedagogical content knowledge. *Mathematics Teacher Education & Development, 15*(1), 72–87.

Manizade, A. G., & Mason, M. M. (2011). Using Delphi methodology to design assessments of teachers' pedagogical content knowledge. *Educational Studies in Mathematics, 76,* 183–207.

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D., & Senk, S. L. (2012). Knowledge of algebra for teaching: A framework of knowledge and practice. *Journal for Research in Mathematics Education, 43*, 584–615.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5–11.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125–145.

Moyer-Packenham, P. S., Bolyard, J. J., Kitsantas, A., & Oh, H. (2008). The assessment of mathematics and science teacher quality. *Peabody Journal of Education, 83*, 562–591. doi: 10.1080/01619560802414940

National Research Council (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: The National Academies Press.

Orrill, C. H., & Cohen, A. (in press). Purpose and conceptualization: Examining assessment development questions through analysis of measures of teacher knowledge. *Journal for Research in Mathematics Education.*

Orrill, C. H., Izsák, A., & Cohen, A. (2006). *Does it work? Building methods for understanding effects of professional development.* [Funded grant proposal to the ARC Centre for Advancing Research and Communication].

Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 9–26). Dordrecht, The Netherlands: Springer.

Post, T. R., Harel, G., Behr, M. J., & Lesh, R. (1988). Intermediate teachers knowledge of rational number concepts. In Fennema, et al. (Eds.), *Papers from the first Wisconsin symposium for research on teaching and learning mathematics* (pp. 194–219). Madison, WI: Wisconsin Center for Educational Research.

Riley, K. R. (2010). Teachers' understanding of proportional reasoning. In P. Brosnan, D. B. Erchick, & L. Flevares (Eds.), *Proceedings of the 32nd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 1055–1061). Columbus, OH: The Ohio State University.

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*, 255–281.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS) instrument. *School Science and Mathematics, 110*, 180–192.

Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links from teacher knowledge, to classroom practices, to student learning in the instruction system of the middle-school mathematics classroom. *Cognition and Instruction, 28*(3), 317–359.

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), 1–22.
Silverman, J., & Thompson, P. W. (2008). Towards a framework for the development of mathematical knowledge for teaching. *Journal of Mathematics Teacher Education, 11*, 499–511.
Simon, M. A. (1997). Developing new models of mathematics teaching: An imperative for research on mathematics teacher development. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 55–86). Mahwah, NJ: Lawrence Erlbaum Associates.
Thompson, A. G., & Thompson, P. W. (1996). Talking about rates conceptually, part II: Mathematical knowledge for teaching. *Journal for Research in Mathematics Education, 27*, 2–24.
Wood, T., Williams, G., & McNeal, B. (2006). Children's mathematical thinking in different classroom cultures. *Journal for Research in Mathematics Education, 37*(3), 222–255.

Author details

Chandra Hawley Orrill, Department of STEM Education & Teacher Development, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747
Email: corrill@umassd.edu

Ok-Kyeong Kim, Department of Mathematics, Western Michigan University, 1903 W. Michigan Avenue, Kalamazoo, MI 49008-5248 USA
Email: ok-kyeong.kim@wmich.edu

Susan A. Peters, Department of Middle and Secondary Education, CEHD, University of Louisville, 1905 S. 1st St. Louisville, KY 40292
email: s.peters@louisville.edu

Alyson Lischka, Department of Mathematical Sciences, Middle Tennessee State University, Box 34, Murfreesboro, TN 37132
email: Alyson.Lischka@mtsu.edu

Cindy Jong, Department of STEM Education, University of Kentucky, 105 Taylor Education Building, Lexington, KY 40506
email: cindy.jong@uky.edu

Wendy Sanchez, Department of Mathematics, Kennesaw State University, 3201 Campus Loop Road, Kennesaw, GA 30144
email: wsanchez@kennesaw.edu

Jennifer A. Eli, Department of Mathematics, The University of Arizona, 617 N. Santa Rita Ave., Tucson, AZ 85721-0089
email: jeli@math.arizona.edu