CRESST REPORT 820

VALIDATING MEASURES OF ALGEBRA TEACHER SUBJECT MATTER KNOWLEDGE AND PEDAGOGICAL CONTENT KNOWLEDGE

SEPTEMBER, 2012

Rebecca E. Buschang

Gregory K.W.K. Chung

Girlie C. Delacruz

Eva L. Baker

National Center for Research on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

Validating Measures of Algebra Teacher Subject Matter Knowledge and Pedagogical Content Knowledge

CRESST Report 820

Rebecca E. Buschang, Gregory K.W.K. Chung, Girlie C. Delacruz, and Eva L. Baker
CRESST/University of California, Los Angeles

September 2012

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305C080015.

The findings and opinions expressed here do not necessarily reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.

To cite from this report, please use the following as your APA reference: Buschang, R.E., Chung, G.K.W.K., Delacruz, G.C., & Baker, E.L. (2012). *Validating measures of algebra teacher subject matter knowledge and pedagogical content knowledge* (CRESST Report 820). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

VALIDATING MEASURES OF ALGEBRA TEACHER SUBJECT MATTER KNOWLEDGE AND PEDAGOGICAL CONTENT KNOWLEDGE

Rebecca E. Buschang, Gregory K.W.K. Chung, Girlie C. Delacruz, and Eva L. Baker
CRESST/ University of California, Los Angeles

Abstract

The purpose of this study was to validate inferences about scores of one task designed to measure subject matter knowledge and three tasks designed to measure aspects of pedagogical content knowledge. Evidence for the validity of inferences was based on two expectations. First, if tasks were sensitive to expertise, we would find group differences. Second, tasks that measured similar types of knowledge would correlate strongly, and tasks that measured different types of knowledge would correlate weakly. We recruited and assessed four groups of participants including 46 experienced algebra teachers (2+ years experience), 17 novice algebra teachers (0-2 years experience), 10 teaching experts, and 13 subject matter experts. Results indicate one task differentiated among levels of expertise and measured several aspects of knowledge needed to teach algebra. Results also highlight that future studies should use a combination of tasks to accurately measure different aspects of teacher knowledge.

Introduction

Math teacher knowledge has been conceptualized and measured in several ways over the years. Many studies of teacher knowledge have focused on teachers' subject matter knowledge and have attempted to relate this type of knowledge to student achievement. Positive effects have been found by some (Hill, Rowan, & Ball, 2005; Mandeville & Liu, 1997; Monk, 1994); however, historically, the majority of studies and reviews have reported inconsistent effects (Ashton & Crocker, 1987; Begle, 1972; Eisenberg, 1977). No consistent evidence of a positive effect of teacher knowledge on student achievement has been empirically found. Possible reasons for the inconsistency in results are variation in the definition of teacher knowledge, the different measures being used, including different "depth" and "breadth" of knowledge being assessed, a lack of alignment of student measures to curriculum, variation in study design, and differences in grade level of participants (Choi, Ahn, & Kennedy, 2008). This lack of consistent results may also be the result of measuring the construct inaccurately.

Studies that conceptualize teacher knowledge solely as subject matter knowledge typically have used measures which most college-educated people could have presumably scored high on (e.g., multiple-choice algebra tests or number of college subject matter courses taken). However, measuring teacher knowledge in this way, as only a function of subject matter knowledge, fails to take into account the knowledge only teachers have, and therefore fails to measure teacher knowledge accurately. Instead, teacher knowledge should be conceptualized as a function of both subject matter and occupational knowledge, as more recent studies have done.

Conceptualizing math teacher knowledge as a combination of subject matter knowledge and occupational knowledge has resulted in a more detailed way to think about math teacher knowledge; however, empirical evidence and valid measures using this reconceptualization are still minimal (Hill, Ball, & Schilling, 2008). Our study examined four measures of teacher knowledge and gathered evidence to determine if these measures could, first, be used to differentiate among varying levels of expertise of knowledge for teaching algebra, and second, determine if there was sufficient evidence to confidently describe the type of knowledge measured by each task (AERA, APA, & NCME, 1999).

Background Information

A Shift in Understanding of Teacher Knowledge

In 1986 Lee S. Shulman began a shift in the conceptualization and measurement of teacher knowledge by introducing the term pedagogical content knowledge to the field (Shulman, 1986). Instead of thinking of teacher knowledge as either subject matter knowledge (e.g., procedural or conceptual knowledge of the subject) or knowledge of teaching (e.g., knowledge of lesson planning, classroom management), he introduced a new type of knowledge that he called pedagogical content knowledge. Shulman defined this knowledge as the knowledge a teacher has about teaching their specific subject area, in particular the "useful forms of representations…, [such as] the most powerful analogies, illustrations, examples, explanations, and demonstrations… that make it comprehensible to others" (Shulman, 1986, p. 9). What made this term important and unique to the field of teacher knowledge was the idea that instead of thinking about teacher knowledge as either subject matter knowledge or knowledge of pedagogy, this term offered a new way to investigate teacher knowledge that linked pedagogical knowledge to a specific subject area.

In mathematics, Shulman's conceptualization of pedagogical content knowledge has been refined further into what has been termed mathematical knowledge for teaching (Ball, Thames, & Phelps, 2008; Hill et al., 2008). This view of teacher knowledge breaks the kinds of knowledge mathematics teachers have into two main categories: subject matter knowledge and pedagogical content knowledge. Subject matter knowledge is further broken down into common content knowledge, specialized content knowledge, and knowledge at the mathematical horizon.

The concepts of common content knowledge and specialized content knowledge are utilized in this paper. Common content knowledge includes mathematical knowledge not specific to teaching such as being able to recognize wrong answers, knowing correct mathematical notation, and being able to compute answers. Specialized content knowledge, which is the deeper subject matter knowledge, is more specific to teaching such as understanding different models for math concepts or understanding how topics interrelate.

The second category of mathematical knowledge for teaching, pedagogical content knowledge, is further broken down into knowledge of content and students, knowledge of content and teaching, and knowledge of curriculum. Knowledge of content and students is the knowledge a teacher possesses about how students learn particular topics and also includes knowledge of likely misconceptions they will have or what topics students will have trouble understanding and why. Knowledge of content and teaching is the knowledge related to teaching mathematics such as knowing which examples to use at specific times or which instructional decisions to make. Together, this view of mathematical knowledge for teaching has made it possible to look at math teacher knowledge in a new and more detailed way. Most importantly, it suggests that teacher knowledge should be examined using a finer grain size than previously used.

Measuring Mathematics Teacher Knowledge

Shulman's conceptualization of teacher knowledge changed how people thought about teacher knowledge. Before Shulman, the way people measured teacher knowledge was typically with multiple-choice tests or using a proxy such as the number of subject matter courses in college. What has proven interesting are the different ways researchers have conceptualized and measured pedagogical content knowledge since Shulman.

One way to examine teacher knowledge is to measure the amounts of different kinds of knowledge and explore the interrelationships between various aspects of knowledge needed for teaching. Hill, Schilling, and Ball (2004) use this viewpoint to explore the structure of elementary mathematics teacher knowledge. A multiple-choice test specifically designed to measure teachers' mathematics knowledge for teaching at the elementary level was created by researchers at the University of Michigan as part of the Learning Mathematics for Teaching (MKT) study (Hill et al., 2004). Initial use of this measure with over 1000 elementary mathematics teachers indicated that elementary teacher knowledge was multidimensional on aspects of both subject matter and pedagogical content knowledge—it is possible for a teacher to have subject matter knowledge but lack knowledge for teaching mathematics.

A second way teacher knowledge has been measured is to focus on very specific aspects of knowledge for teaching mathematics and measure how those aspects of knowledge interrelate. For example, in one study, three different types of knowledge for teaching mathematics were elicited, including (a) knowledge of alternative solutions to the same problem, (b) awareness of student misconceptions, and (c) knowledge of instructional strategies, and then the relationships between these types of knowledge were explored. Two hundred and eighteen secondary math teachers were administered a measure with 35 open-ended questions. Results suggested a strong relationship between what the authors called content knowledge and pedagogical content knowledge (Krauss, Neubrand, Blum, & Baumert, 2008). Similarly, in a second study attempting to elicit both algebra knowledge (e.g., teachers' understanding of algebraic concepts such as rational number equivalence, principles for solving equations, and properties of arithmetic) and knowledge for teaching algebra (e.g., types of knowledge needed to solve certain problems), and then examine the relationship among these constructs, a knowledge mapping task was designed specifically for this purpose and tested with 25 secondary math teachers and one subject matter expert. Results showed moderate correlations among different subject matter knowledge tasks and tasks designed to measure teachers' knowledge of teaching algebra (Delacruz et al., 2007).

A final way teacher knowledge has been examined in the literature is related to specific teaching practices. One example is a study that describes teachers' ability to analyze student work by having teachers analyze samples of student responses to math problems and answer several questions such as, "If this student was in your class, what would you do next in your instruction?" Findings indicated that while teachers were able to identify concepts that specific math problems attempted to measure, teachers had a difficult time deciding their next teaching steps based on student work (Heritage, Kim, Vendlinski, & Herman, 2008).

These results highlight that Shulman's conceptualization of pedagogical content knowledge has led to a wide variety of measures being used, different conceptualizations of math teacher knowledge, and a range of study results. This also highlights that many different aspects of teacher knowledge exist and are able to be measured. However, this also raises the question of which is the appropriate method. Because of the multiple aspects of teacher knowledge, multiple measures may be needed to more fully measure teacher knowledge. The topic of pre-algebra has been used in this study to examine multiple measures of teacher knowledge due to the lack of valid and reliable measures for mathematical knowledge of teaching.

Expertise and Domain Specificity of Teacher Knowledge

Research on expertise suggests that expertise is not a general characteristic of the individual but is rather related to a specific domain of expertise (Chase & Simon, 1973). This implies that individuals will not develop general expertise in all subject areas, but instead will need to develop expertise in a particular subject area.

As expertise in a particular subject area develops, that knowledge becomes deeper and more complex, and more relations can be seen in the mental models of experts (Chi, Glaser, & Farr, 1988). Novices on the other hand tend to focus on surface-level features and have fewer connections among concepts. Different levels of expertise in a specific domain can be differentiated with tasks designed to be sensitive to different levels of expertise.

These findings suggest that when studying a domain, it is important to be clear about the specific domain of knowledge being assessed. In this study, we limited our examination to knowledge for teaching algebra. Findings from the expert/novice literature also suggest that expertise develops over time into a more complex and deeper understanding, and that experts and novices can be distinguished based on the complexity of their mental model representations.

Context of Current Study

There is empirical evidence that teacher knowledge is multidimensional. There is also evidence that this multidimensionality cannot be measured by one item type or with one method, but instead needs to be measured with a battery of tests that covers the intended type of knowledge to be measured. If these tests are intended to be used as measures of the amount of particular types of knowledge, they must be specifically designed and tested to distinguish among levels of expertise.

Prior studies have attempted to capture a different range of mathematics teacher knowledge than this study. For example, Hill et al. (2008) examined a broad range of mathematical knowledge that included teaching practices by examining the relationship between mathematical knowledge for teaching and the mathematical quality of instruction of second through sixth grade teachers. In contrast, this study attempts to examine multiple aspects of a narrow subject area, specifically knowledge for teaching algebra, and uses a wide variety of measures. This study also attempts to test measures that are specifically designed to differentiate between experts and novices in knowledge for teaching algebra.

Our design choices are very specific because of our purposes. First, because we wanted to elicit a variety of knowledge for teaching algebra including aspects of subject matter knowledge and pedagogical content knowledge, four different tasks were used. Second, because our purpose was to test if the selected tasks could differentiate among levels of expertise, our participants were specifically selected as having one of four levels of expertise.

Research Questions

Validity is the "degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests" (AERA, APA, & NCME, 1999, p.9). This definition requires that evidence be gathered to validate score interpretations for certain purposes but the tests themselves are not validated. The purpose of this study was to determine if sufficient evidence could be accumulated to first determine if these tasks could be used to differentiate among levels of expertise in algebra and teaching of algebra, and second, determine if the types of knowledge assessed in the different tasks could be confidently outlined. To accumulate this evidence, we focused on two main research questions:

1. To what extent do these tasks differentiate among levels of expertise?
2. What is the relation among tasks, both those designed to measure the same type of knowledge and different types?

The purpose of the first research question was to determine if there was sufficient evidence that the tasks in this study could differentiate among individuals with different levels of algebra subject matter knowledge and knowledge for teaching algebra. If measures were able to differentiate among different groups, this would provide validity evidence that the measures could be used to categorize individuals into certain groups of algebra subject matter knowledge and knowledge for teaching algebra.

The purpose of the second research question was to explore the types of knowledge or constructs being measured by each task. Measures which correlate highly with each other would provide evidence that similar types of knowledge or constructs are being measured. Measures that correlate poorly with each other would indicate different types of knowledge or constructs being measured. Exploring these two questions will help us gather evidence in support of the types of inferences that can be made from the results of these tasks (AERA, APA, & NCME, 1999).

On a more general level, the investigation of these research questions is important because it will provide insight into the measurement of different aspects of teacher knowledge. Results will provide evidence either in support of or against attempting to measure multiple aspects of teacher knowledge in one study. It may also provide insight into the types and amounts of knowledge individuals with different levels of teaching expertise have and the relationship among the types of knowledge.

Method

Participants

Four groups of participants were recruited for this study: (a) subject matter experts (*n* = 13), (b) pedagogical content knowledge experts (*n* = 10), (c) novice teachers (*n* = 17), and (d) experienced teachers (*n* = 46) for a total of 86 participants. See Tables 1 and 2 for a description of participants. We expected these groups of participants to have different levels of algebra subject knowledge and knowledge for teaching algebra.

Table 1

Description of Continuous Variables for All Participant Groups

	Subject matter experts (*n* = 13)		Pedagogical content knowledge experts (*n* = 10)		Novice teachers (*n* = 17)		Experienced teachers (*n* = 46)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	25.08	2.78	54.70	6.90	28.06	5.51	41.52	11.11
Years teaching	--	--	24.31	10.12	1.29	0.66	11.16	8.83

Table 2

Description of Categorical Variables for All Participant Groups

Variable	Subject matter experts (*n* = 13)	Pedagogical content knowledge experts (*n* = 10)	Novice teachers (*n* = 17)	Experienced teachers (*n* = 46)
Gender				
Female	7.7%	60.0%	82.4%	60.9%
Male	92.3%	40.0%	17.6%	39.1%
Ethnicity				
Asian	27.3%	0%	58.8%	11.4%
African American	0%	0%	0%	0%
Latino/a	0%	0%	0%	2.3%
White	72.7%	100%	29.4%	79.5%
Other	0%	0%	11.8%	6.8%

Variable	Subject matter experts (*n* = 13)	Pedagogical content knowledge experts (*n* = 10)	Novice teachers (*n* = 17)	Experienced teachers (*n* = 46)
Bachelor's Degree				
Math	100%	77.8%	23.5%	45.5%
Education	0%	0%	17.6%	13.6%
Science/Engineering	0%	11.1%	35.3%	4.5%
Other	0%	11.1%	23.5%	36.4%
Grade Level Taught				
Middle School	N/A	90%	46%	87%
High School	N/A	10%	40%	13%
Other/Student Teaching	N/A	0%	13%	0%

Participants were recruited via graduate student and teacher list-serves and email lists. All participants who qualified based on criteria for each category listed below were accepted as participants.

Requirements to be classified as a *subject matter (SM) expert* (*n* = 13) included being a current student in a mathematics Ph.D. program with no K-12 algebra classroom teaching experience. These requirements were established to create a group with high subject matter knowledge but low pedagogical content knowledge for algebra. All participants had undergraduate degrees in mathematics, and teaching assistant experience in advanced math, physics, and engineering university-level courses but not algebra-level courses. Nine had tutoring experience mostly at the college level in non-algebra courses. Because of their limited experience designing lesson plans and teaching in algebra classroom situations, we were confident that this group had low knowledge related to teaching algebra. Using Ball et al.'s (2008) conceptualization of mathematical knowledge for teaching, these individuals would most likely have high common content knowledge of algebra, and low specialized content knowledge of algebra, low knowledge of content and students in algebra, and low knowledge of content and teaching in algebra because of their training and experience.

Requirements to be classified as a *pedagogical content knowledge (PCK) expert* included being a former or current math teacher with either National Board Certification (*n* = 3) or holding a position working with and training other math teachers (*n* = 7). National Board Certified teachers were chosen because it has been shown that National Board Certification is a good indicator of high teacher quality (Cavalluzzo, 2004; Goldhaber & Anthony, 2007). Teacher trainers were chosen because of their specialized training and high knowledge of mathematics teaching methods. This sample comprised individuals with high math knowledge of algebra, substantial training in the teaching of mathematics, and an average of almost 25 years of algebra teaching experience. Using Ball et al.'s (2008) conceptualization of mathematical knowledge for teaching, these individuals were most likely to have high algebra subject matter knowledge and the different types of pedagogical content knowledge related specifically to algebra.

Requirements to be classified as an *experienced teacher* included being a teacher with more than two years of experience teaching math without National Board Certification or positions training other teachers. Participants averaged 11 years of teaching experience in algebra. Requirements to be classified as a *novice teacher* included being a teacher with two years or less of algebra teaching experience. Teachers in this sample, on average, were in their first year of teaching. The experienced and novice teacher samples represented a typical group of teachers with a range of undergraduate degrees and years of teaching experience. Using Ball et al.'s (2008) conceptualization of mathematical knowledge for teaching, these individuals were most likely to have high common knowledge of algebra like all other groups, but a range of pedagogical content knowledge of algebra and specialized content knowledge.

Measures

We administered four tasks and a background survey to all participants. Descriptions of the tasks and scoring procedures are explained in the two sections below. In the first section, the task designed to measure subject matter knowledge is explained. In the second section, the three tasks designed to measure aspects of pedagogical content knowledge are explained.

Measure of subject matter knowledge. A knowledge mapping task was used to measure subject matter knowledge of algebra. A knowledge map is a network of nodes that represent the concepts in a domain, and links that visually represent how an individual perceives concepts to relate to each other. Knowledge maps have been used to assess knowledge of students and adults (Herl, Baker, & Niemi, 1996; Herl, O'Neil, Chung, & Schacter, 1999; Klein, Chung, Osmundson, Herl, & O'Neil, 2002). Results of these studies suggest that knowledge maps are a valid and reliable way to distinguish among different levels of expertise in a particular area of knowledge.

The specific task used in this study to assess subject matter knowledge is the Concept-Only Knowledge Map. The Concept-Only Knowledge Map asked participants to represent how a given set of algebra concepts interrelate. This task was designed to evaluate subject matter knowledge. It was designed to evaluate subject matter knowledge at a deeper level than can be measured with a test of algebra problems though it is still common content knowledge in that it is knowledge of algebra not specific to teaching that people in professions other than teaching could have (Ball et al., 2008; Hill et al., 2008). This task allows us to evaluate a participant's mental model of pre-algebra concepts and then compare it to an expert's model (Herl et al., 1996; Delacruz et al., 2007).

To complete the task, participants created a knowledge map on paper using 24 predetermined concept stickers (e.g., “additive identity” or “fractions”) and nine predetermined link stickers (e.g., “shows” or “can represent”). Participants placed concept stickers on a 2 foot by 2 foot sheet of paper and connected concepts with the provided link label stickers as they determined appropriate. Participants were instructed that the arrow direction on the link stickers specified the direction of the relationship between the two concepts (see Figure 1 for a sample of part of a Concept-Only knowledge map).

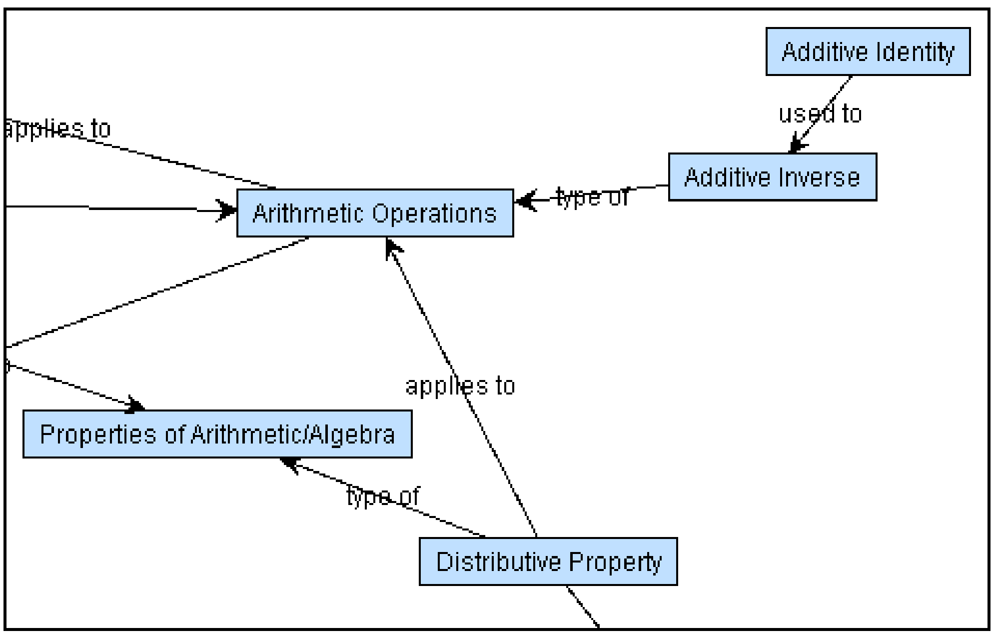


Figure 1. Portion of a participant Concept-Only Knowledge Map (Part 1). In the example, concepts are shown in boxes. Links are written on the arrows.

This task was scored in two ways. First, the number of concepts used and the number of links made to other concepts were counted. Expertise research indicates that complexity of knowledge maps is an indication of a high level of expertise (Chi, et al., 1988). Therefore, the number of concepts used and links made by teachers was used as a proxy for level of expertise of the different groups.

Second, after the check on expertise was completed, similarity scores were created. Similarity scores were based on the PCK expert maps. PCK experts were chosen as the reference group because results of the independent scoring of the Concept-Only map indicated that this group had the most complex maps. They were also the participants who had the most training in the teaching of mathematics at the pre-algebra and algebra level.

To create the similarity scores, each proposition was evaluated based on the concept-link-concept proposition and the direction of the link. Scoring was based on how similar each participant's map was to the PCK expert maps. One point was awarded for an exact proposition match. Zero points were awarded for partial or non-matching propositions. To create a total score for each participant, the Concept-Only Knowledge Map was scored against all PCK expert maps. Then, the mean of these scores was taken for each participant.

What is measured with this task is similar to the common content knowledge because it is knowledge a teacher might have, but a type of knowledge that is not exclusive to teaching. It was expected that all groups would have relatively similar performance on this task.

Measures of pedagogical content knowledge. The following section outlines three tasks designed to measure aspects of pedagogical content knowledge. As outlined earlier, pedagogical content knowledge is the deeper knowledge about a specific subject area that teachers have. The first task described is an extension of the Concept-Only Knowledge Map, the second is a multiple-choice measure of knowledge for mathematics teaching, and the third task required participants to evaluate student work.

Problems-added knowledge map. The Concept-Only Knowledge Map described above was intended to measure content knowledge. The extension of this knowledge map, the Problems-Added Knowledge Map, was designed to assess knowledge of teaching algebra. Specifically, it was designed to measure teachers' ability to decide what types of knowledge a student would need to be able to solve an algebra problem (Delacruz et al., 2007). This task asked participants to "Relate the concepts in your [Concept-Only] knowledge map to the mathematics problems provided by considering *what concepts are relevant to solve the problem correctly*."

This task was chosen because we were interested in assessing a teacher's knowledge of how students learn a certain topic and teachers' ability to identify the instructional sequence of algebra knowledge, which is similar to knowledge of content and students (Ball et al., 2008; Hill et al., 2008). Completion of this task required participants to know what background knowledge students would need to complete certain algebra problems.

To complete the task, participants added up to 21 predetermined algebra problems to the map they created in the Concept-Only Knowledge Map. First, participants placed the problem stickers on the same 2 foot by 2 foot piece of paper and were then instructed to link concept(s) they perceived necessary to solve that algebra problem (Figure 2).

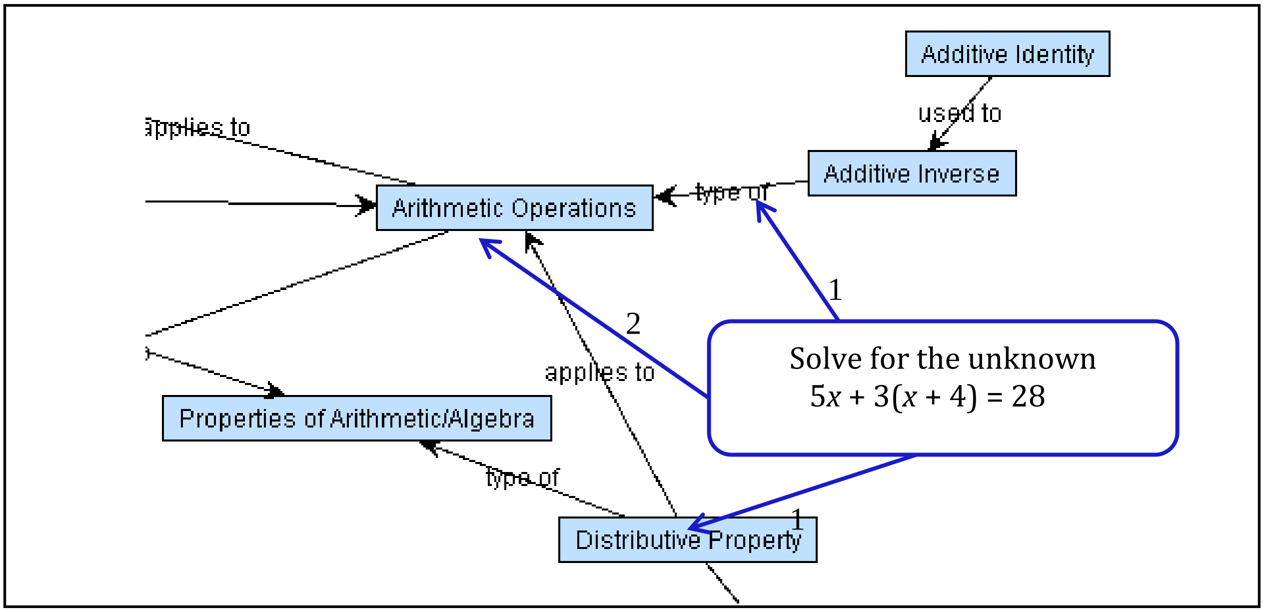


Figure 2. Portion of a participant concept map plus problem knowledge map (Part 2).

Here the expert has linked an algebra problem, shown in the box with rounded corners, to the concepts needed to solve the problem and rated the links as either 1 or 2.

Participants also rated each problem-concept pair using the following guidelines:

- 1 = necessary, **but not** sufficient (you need to know this concept, but it is not enough to solve the problem correctly)
- 2 = necessary **and** sufficient (if you know this concept, you can solve the problem correctly)

A similar process to the Concept-Only Knowledge Map was used to score the Problems-Added Knowledge Maps. As with the Concept-Only Knowledge Map, one point was awarded for an exact match to the reference map, and zero points were awarded for a partial or non-matching proposition. However, in the Problems-Added Knowledge Map, the direction of the link was disregarded since it did not affect interpretation of the relationship. Scoring was based on how similar each participant's map was to the PCK expert maps.

Mathematical knowledge for teaching instrument. The Mathematical Knowledge for Teaching (MKT) instrument is a multiple-choice test designed to measure mathematical knowledge for teaching designed by researchers at the University of Michigan (Hill et al., 2004) as part of the Learning Mathematics for Teaching project. This measure covers concepts such as algebra, number concepts and operations, functions, and geometry.

Item A.

Mrs. Smith is looking through her textbook for problems and solution methods that draw on the distributive property as their primary justification. **Which of these familiar situations could she use to demonstrate the distributive property of multiplication over addition [i.e., a (b + c) = ab + ac]?** (Mark APPLIES, DOES NOT APPLY, or I'M NOT SURE for each.)

	Applies	Does not apply	I'm not sure
a) Adding	1	2	3
b) Solving $2x - 5 = 8$ for x	1	2	3
c) Combining like terms in the expression $3x^2 + 4y + 2x^2 - 6y$	1	2	3
d) Adding 34 + 25 using this method:	1	2	3

Item B.

Students in Mr. Carson's class were learning to verify the equivalence of expressions. He asked his class to explain why the expressions a – (b + c) and a – b – c are equivalent. Some of the answers given by students are listed below.

Which of the following statements comes <u>closest</u> to explaining why a – (b + c) and a – b – c are equivalent? (Mark ONE answer.)

a) They're the same because we know that a – (b + c) doesn't equal a – b + c, so it must equal a – b – c.

b) They're equivalent because if you substitute in numbers, like a=10, b=2, and c=5, then you get 3 for both expressions.

c) They're equal because of the associative property. We know that a – (b + c) equals (a – b) – c which equals a – b – c.

d) They're equivalent because what you do to one side you must always do to the other.

e) They're the same because of the distributive property. Multiplying (b + c) by –1 produces –b – c.

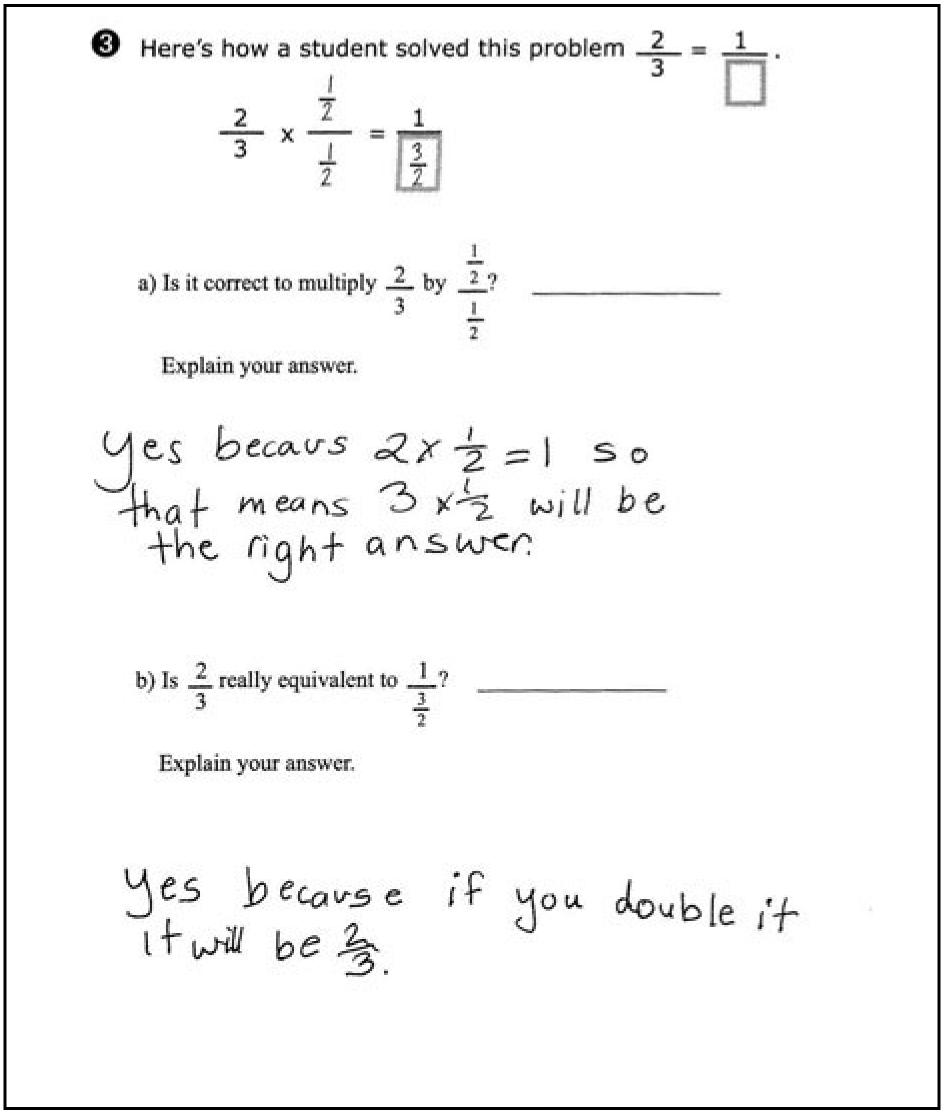
Figure 3. Released items from the Mathematical Knowledge for Teaching (MKT) measure.

The MKT instrument was used in this study because (a) it is a measure based on previously validated measures of pedagogical content knowledge (Hill et al., 2004), and (b) it is a measure of teachers' knowledge of instructional choices, common student errors, and appropriateness of student responses. The questions on the MKT instrument address specialized content knowledge, knowledge of content and students, and knowledge of content and teaching (Ball et al., 2008; Hill et al., 2008).

A total of 20 out of the original 28 items were selected and administered to participants. The eight problems not selected were geometry items. Those items were dropped so that only knowledge related to algebra was being assessed in this task.

Each answer was scored as correct or incorrect. The total correct answers for the 20 items were counted for a total score. Because several items had multiple correct answers (e.g., “select all that apply”), the total number of correct answers was counted to equal the total score. The maximum possible was 45. Totaling the number of correct items was chosen over analysis methods such as item response theory because we were interested in looking at total scores of individuals rather than the performance of individual items. Cronbach’s alpha was calculated for all items and established at alpha = 0.91 indicating high internal consistency among items.

Student response analysis task. The Student Response Analysis (SRA) task was designed to assess pedagogical content knowledge, in particular the knowledge needed to evaluate student work and determine the next teaching steps (Heritage, Kim, & Vendlinski, 2008; Heritage, Kim, Vendlinski, & Herman, 2008). Participants analyzed several student responses posted online, then responded to the prompt, “If these students were in your class what would you want to teach next? Why?”



3 Here's how a student solved this problem $\frac{2}{3} = \frac{1}{\square}$.

$$\frac{2}{3} \times \frac{\frac{1}{2}}{\frac{1}{2}} = \frac{1}{\frac{3}{2}}$$

a) Is it correct to multiply $\frac{2}{3}$ by $\frac{\frac{1}{2}}{\frac{1}{2}}$? ______________

Explain your answer.

Yes becaus $2 \times \frac{1}{2} = 1$ so that means $3 \times \frac{1}{2}$ will be the right answer.

b) Is $\frac{2}{3}$ really equivalent to $\frac{1}{\frac{3}{2}}$? ______________

Explain your answer.

Yes because if you double it it will be $\frac{2}{3}$.

Figure 4. Sample of student work given to teachers for the SRA task.

This task was chosen because it mirrors daily tasks of teachers including interpreting student work and creating next teaching steps which is similar to what is called knowledge of content and teaching (Ball et al., 2008; Hill et al., 2008). To complete this task, an individual would need to have specialized content knowledge as well as knowledge of common student errors, and then have the capability to take that knowledge and apply it to next teaching steps. This task is also different from the other two measures of pedagogical content knowledge because it is a constructed response task.

Responses were coded on three dimensions: (a) if the participant would re-teach the material, (b) what method the participant would use to re-teach the material, and (c) if and how the participant would differentiate instruction for students. Total scores were created by adding the scores for each of the three dimensions. Accuracy of scoring was checked in several ways. Interrater reliability was checked on this set of data and was established at 84.3% agreement. Cronbach's alpha was established at alpha = 0.56 indicating low internal consistency among items.

For the Problems-Added Knowledge Map, the MKT, and the SRA tasks, our expectation was that each task required a high level of knowledge of teaching algebra, specifically on the daily tasks teachers carry out. Therefore, we expected that the PCK experts would perform at higher levels than individuals with low pedagogical knowledge in algebra such as our SM experts and possibly our novice teachers. We also expected high correlations with the other tasks intended to measure pedagogical content knowledge since these tasks are measuring similar types of knowledge and moderate correlations with the Concept-Only Knowledge Map.

Procedure

Via U.S. mail, participants received paper-based tasks (the Concept-Only and Problems-Added Knowledge Map and the MKT instrument), a 2-minute CD with video instructions for the knowledge map tasks, and instructions for accessing the online materials (the background survey and the Student Response Analysis task). Participants returned completed tasks via U.S. mail.

Tasks received via mail were entered manually into the computer. Knowledge maps were also scored automatically using the CRESST Human Performance Knowledge Mapping Tool (Chung et al., 2006). Out of the 119 participants who received the complete set of tasks, there was a 73% return rate. A $100 honorarium was given for full participation.

Results

Two separate analyses were conducted; however before the main analyses could be carried out, an analysis of the expert groups was conducted to verify appropriate labeling of expert groups for the knowledge mapping tasks. Using the Concept-Only Knowledge Map, the number of concepts used and the number of connections created was counted. Results indicate that PCK experts have the most complex representations related to algebra. They used significantly more concepts on the concept map ($F(3, 76) = 6.42$, $p = .001$), and they used more links per concept than other groups (see Table 3). These results suggest that the PCK experts are an appropriate reference group for creating similarity scores for the knowledge mapping tasks.

Table 3

Number of Concepts and Links on Concept-Only Knowledge Map by Group

	Subject matter experts (*n* = 13)		Pedagogical content knowledge experts (*n* = 10)		Novice teachers (*n* = 17)		Experienced teachers (*n* = 46)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Number of concepts	7.58	2.31	14.37	9.26	7.27	3.56	7.31	3.59
Number of links	12.75	5.17	18.50	8.31	11.93	4.41	13.55	6.66

Group Differences

To answer the first research question that focused on whether individual tasks differentiated among levels of expertise, group differences were conducted on each task using one-way analysis of variance (ANOVA) with expertise level as the between-subjects factor. Results showing significant differences among groups with different levels of subject matter knowledge of knowledge specific to teaching algebra would be evidence for the validity of using these tasks to differentiate among level of expertise.

Several patterns emerged from analysis of group differences (see Table 4). Significant group differences only existed for the MKT task, $F(3, 81) = 10.41$, $p < .001$. A post hoc test using the Tukey HSD revealed that PCK experts (*M* = 37.30, *SD* = 5.42) scored significantly higher than novice teachers (*M* = 28.35, *SD* = 6.86). Results also revealed that SM experts (*M* = 43.67, *SD* = 1.88) scored significantly higher than novice teachers (*M* = 28.35, *SD* = 6.86) and experienced teachers (*M* = 31.67, *SD* = 9.55). No other significant differences were found.

While other tasks did not show significant differences, trends in the results are worth mentioning. For the SRA, the task aimed to measure pedagogical content knowledge of student errors and knowledge of appropriate instructional decisions, PCK experts scored highest, experienced teachers and novice teachers scored next highest, and SM experts scored lowest. For the Concept-Only Knowledge Map, which was the task aimed to measure content knowledge, and the Problems-Added Knowledge Map, which was the task aimed to measure teacher knowledge of how students learn certain topics, SM experts scored highest, then expert teachers, followed by experienced teachers and then novice teachers. These trends merit further exploration since they follow expected patterns of results.

Table 4

Group Differences on Measures

	Subject matter experts (*n* = 13)		Pedagogical content knowledge experts (*n* = 10)		Experienced teachers (*n* = 17)		Novice teachers (*n* = 46)	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
MKT	43.67[a]	1.88	37.30[b]	5.42	31.67[a]	9.55	28.35[ab]	6.86
Concept-Only Knowledge Map	2.37	.82	1.87	1.00	1.89	1.18	1.48	1.08
Problems-Added Knowledge Map	5.70	2.60	5.11	2.13	4.24	2.70	4.33	3.32
SRA	.58	.67	1.20	1.03	.93	.98	1.06	1.14

Note. MKT = Mathematical Knowledge for Teaching instrument; SRA = Student Response Analysis Task. [a]Group differences between SM experts and non-experts (experienced teachers and novice teachers) exist at $p < .05$ (two-tailed). [b]Group differences between PCK experts and non-experts (experienced teachers and novice teachers) exist at $p < .05$ (two-tailed).

Relation of Tasks

Correlation data allowed us to explore the relation among tasks (i.e., which showed evidence of measuring similar constructs and which showed evidence of measuring different constructs). The main purpose of this analysis was to explore how tasks related to each other to gather evidence about the likely types of knowledge each task is measuring. Results showing high correlations between two tasks would provide evidence that there is a relationship between the two tasks, and that they likely measure similar constructs or types of knowledge. Low correlations between two tasks likely would provide evidence that the two tasks measured different constructs or types of knowledge.

Whole group correlations. First, we examined the correlation among tasks for all participants using Pearson's correlations. Moderate significant correlations were found among the MKT, the Concepts-only Knowledge Map, and the Problems-Added Knowledge Map (see Table 5). These results provide evidence that similar constructs were being measured by these tasks. Low correlations were found among the SRA and the other three tasks. This result indicates that the SRA could be measuring a different construct or type of knowledge than the other tasks.

Table 5

Relation of Tasks Designed to Measure Pedagogical Content Knowledge

Measure	Concept-Only Knowledge Map	Problems-Added Knowledge Map	SRA
MKT	.41**	.50	.06
Concept-Only Knowledge Map		.55**	.12
Problems-Added Knowledge Map			.15
SRA			

Note. MKT = Mathematical Knowledge for Teaching instrument; SRA = Student Response Analysis Task.

Correlations among tasks by group. After examining correlation among tasks for the whole group, we explored the correlations among tasks for different participant groups. Spearman's rho was used because of the small sample size in some groups. This analysis allowed us to examine the relations of performance among tasks for different groups, and determine if the relation among tasks differed by group. Differences among groups could provide additional evidence about the types of knowledge being measured by each task and provide information about whether tasks could be used to differentiate among levels of expertise.

First, correlations among tasks designed to measure pedagogical content knowledge, including the MKT, the Problems-Added Knowledge Map, and the SRA, were examined for each participant group (see Table 6). Results of this analysis indicate that the correlation among tasks differed among groups. Notably, there seemed to be differences between groups with high teaching experience (PCK experts and experienced teachers) and those with less teaching experience (novice teachers and SM knowledge). For the MKT and the Problems-Added Knowledge Map, groups with high teaching experience showed strong correlations among these tasks, and those groups with the least amount of teaching experience (novice teachers and SM experts) showed weak correlations. These results suggest that these tasks are measuring a type of knowledge that experienced teachers would have, but less experienced teachers would not. For the SRA, moderate to high correlations existed for PCK experts only among this task and other tasks designed to measure aspects of pedagogical content knowledge. Low to moderate negative correlations among the SRA and the other tasks designed to measure aspects of pedagogical content knowledge existed for the group with no K-12 algebra teaching experience. These results could be an indication that the SRA is measuring a different type of knowledge than the other tasks, or possibly a different skill that only those with very specialized math teaching knowledge, such as PCK experts, would be able to express. Because inter-item reliability is low on the SRA, results must be considered with caution.

Table 6

Relation of Tasks Designed to Measure Pedagogical Content Knowledge for Separate Groups

Task	Problems-Added Knowledge Map	SRA
MKT		
SM experts	.01	-.15
PCK experts	.74*	.47
ET	.56**	.22
NT	.22	-.22
Problems-Added Knowledge Map		
SM experts	---	-.42
PCK experts	---	.82**
ET	---	.22
NT		.12

Note. MKT = Mathematical Knowledge for Teaching instrument; SRA = Student Response Analysis Task; SM = subject matter; PCK = pedagogical content expert; ET = experienced teacher; NT = novice teacher.
**p* < .05 (two-tailed). ***p* < .01 (two-tailed).

Next, correlations among tasks designed to measure subject matter knowledge and those designed to measure pedagogical content knowledge were also examined. Results show differences in associations for different groups (see Table 7). In particular, results indicate moderate correlations for most groups between the Concept-Only Knowledge Map and the Problems-Added Knowledge Map and between the Concept-Only Knowledge Map and the MKT. The only exception to this is a low correlation for the SM expert groups between the MKT and the Concept-Only Knowledge Map. These results indicate that, in general, the MKT and Problems-Added Knowledge Map are measuring constructs related to content knowledge, but not the identical construct. Results for the SRA show strong correlations to the Concept-Only Knowledge Map for PCK experts only, weak correlations for both experienced and novice teachers, and moderate negative correlations for SM experts. These results could be additional evidence that the SRA is measuring a different type of pedagogical content knowledge than the other two tasks designed to measure pedagogical content knowledge.

Table 7

Relation of Tasks Designed to Measure Subject Matter Knowledge and Pedagogical Content Knowledge for the Separate Groups

Task	MKT	Problems-Added Knowledge Map	SRA
Concepts-only Knowledge Map			
SM experts	.14	.64*	-.51
PCK experts	.38	.43	.50
ET	.40**	.56**	.17
NT	.41	.51	.20

Note. MKT = Mathematical Knowledge for Teaching instrument; SRA = Student Response Analysis Task; ET = experienced teacher; NT = novice teacher.
**p < .01 (two-tailed)

Conclusion

Reflecting on the Study's Research Questions

Research Question 1: Do these measures differentiate level of expertise? The first purpose of this study was to gather evidence to determine if scores from the selected tasks could differentiate among levels of expertise. The MKT was sensitive to both subject matter knowledge and pedagogical content knowledge, but only between expert and non-expert levels. This indicates that while the MKT task was sensitive to some levels of expertise, it was not able to differentiate between novice and experienced teachers. One interpretation of this result is that individuals with more teaching experience do not necessarily have more knowledge about teaching mathematics. The MKT was also sensitive to both subject matter expertise as well as pedagogical content expertise, suggesting sensitivity to both subject matter knowledge and pedagogical content knowledge. This result is consistent with factor analysis showing teachers' knowledge as measured by the MKT instrument to be multidimensional with a factor for subject matter knowledge and one for pedagogical content knowledge, and that it was hard to differentiate these from each other on many items (Hill et al., 2004). This result also provides evidence that the MKT measures a combination of subject matter knowledge and pedagogical content knowledge.

The SRA, Concept-Only Knowledge Map, and the Problems-Added Knowledge Map were not sensitive to levels of expertise as measured by group differences. One explanation for these results is that the tasks, while designed to measure different aspects of pedagogical content knowledge, were not sensitive enough to distinguish among levels of expertise possibly due to design or scoring methods. An alternative explanation is that the tasks were tapping into knowledge or skills that were not exclusively either subject matter knowledge or pedagogical content knowledge, for example, test-taking ability or reasoning ability. Therefore, the knowledge that these tasks were designed to measure might be masked by these other types of knowledge or skill.

Based on these results, there is evidence that scores from the MKT can distinguish between expert levels of knowledge and non-expert (e.g., experienced and novice teachers) levels of knowledge. Trends found on other tasks warrant further exploration especially in light of the small sample size.

Research Question 2: What is the relation among tasks, both those designed to measure the same type of knowledge and different types? The relation among the different tasks was explored to answer the second research question. This research question was examined to gather evidence about the types of inferences that could be made about certain measures. Specifically we investigated if the tasks showed evidence of measuring the constructs they were designed to measure. Results suggest that three aspects of knowledge for teaching algebra were measured by the set of tasks used in this study. First, the Problems-Added Knowledge Map and the MKT seemed to measure a similar type of knowledge for teaching algebra that was highly dependent upon subject matter knowledge. Evidence that these tasks measure a similar type of knowledge come from moderate to strong correlations between these two tasks for the whole group. Evidence that these tasks were measuring an aspect of knowledge needed to teach algebra that depends on subject matter knowledge comes from moderate whole group correlations to the Concept-Only Knowledge Map, and from differential correlations of certain groups, specifically moderate to high correlations for groups with high teaching knowledge. Since pedagogical content knowledge is typically defined as a combination of subject matter knowledge and knowledge for teaching, it is not surprising that there are correlations among the task designed to measure only subject matter knowledge (i.e., Concept-Only Knowledge Map) and the tasks that were designed to measure pedagogical content knowledge (i.e., the MKT, and the Problems-Added Knowledge Map). These results also support previous studies that have found links among tasks or items designed to measure subject matter knowledge and those designed to measure knowledge for teaching (Hill et al., 2004; Krauss et al., 2008; Delacruz et al., 2007).

The second aspect of knowledge for teaching algebra being measured by this set of tasks is subject matter knowledge by the Concept-Only Knowledge Map. This inference is based mostly on the nature of the Concept-Only Knowledge Map that asks individuals to create a knowledge map of algebra using terms related to the domain of algebra. It is also based on the moderate correlations of the Concept-Only Knowledge Map to tasks where teachers carried out activities or answered questions similar to what is done while teaching algebra (i.e., Problems-Added Knowledge Map, MKT). These moderate correlations are an indication that the knowledge measured with the Concept-Only Knowledge Map is related to but not the same as the knowledge measured by the other tasks. No other tasks in this study were designed to exclusively measure subject matter knowledge, therefore no analyses between tasks designed to only measure subject matter knowledge could be carried out.

Finally, the SRA, while also designed to measure pedagogical content knowledge, showed very weak correlations to the other task designed to measure pedagogical content knowledge when analyzed for the whole group. This result suggests that the SRA is possibly measuring a type of knowledge or skill different to what the other tasks are measuring. Furthermore, when correlation data were analyzed separately by group, there were high correlations for the PCK experts, which is the group that likely has the highest pedagogical content knowledge, and low correlations for individuals with low knowledge for teaching. This is additional evidence that the SRA is likely measuring a different type of pedagogical content knowledge or skill than the other tasks. Since participants were asked to answer open-ended questions on this task as opposed to selecting responses on the MKT or Problem-Added Knowledge Map, this different method of response could also be responsible for the differential correlation. It could indicate that only certain groups have the ability to express this type of knowledge or that solely being asked to answer in a different way could in itself cause the weak association among the SRA and the other tasks.

Overall, there is evidence that this set of tasks measure three combinations of knowledge needed to teach algebra: a mixture of subject matter knowledge and pedagogical content knowledge by the Problems-Added Knowledge Map and the MKT, subject matter knowledge by the Concept-Only Knowledge Map, and a special type of pedagogical content knowledge by the SRA. Using Ball et al.'s (2008) conceptualization of mathematical knowledge for teaching, the Problems-Added Knowledge Map and the MKT would most likely be associated with a combination of specialized content knowledge, knowledge of content and students, and knowledge of content and teaching. The Concept-Only Knowledge Map would most likely be associated with common content knowledge. The SRA would most likely be associated with knowledge of content and students and knowledge of content and teaching.

General Conclusions

The results of this study suggest that teacher knowledge is a complex construct to measure and that no one assessment can accurately measure it in its entirety. Instead a battery of well developed tests is needed to measure pedagogical content knowledge of teachers. This is an important finding in light of the renewed interest in measurement of teacher knowledge. Future studies attempting to measure teacher knowledge broadly must use several measures to accurately assess teacher knowledge. Those studies attempting to measure only a particular aspect of teacher knowledge need to be cautious in designing or picking an assessment that accurately measures the intended aspect of teacher knowledge.

Limitations of This Study

Interpretation of these results is limited by the small sample size of several of the groups. The small sample size could have effects on the statistical tests and may have resulted in Type II errors. Our interpretation is also limited by selection criteria and sampling procedures.

The reliability of the SRA task was also low. One possible explanation for the low internal consistency stems from the correlational analysis broken down by level of expertise that shows that groups have differential expertise. Results should be interpreted with caution.

A final limitation of this study is a lack of knowledge about how tasks were completed and if participants consulted resources to complete the tasks due to distribution through the mail. Participants were asked to complete the tasks individually. However, it is possible that participants consulted with other individuals or resources while completing the tasks.

Future Directions

This study has helped to gather evidence for how scores from these tasks could be interpreted. It has also helped in our understanding of teacher knowledge as related to expertise.

One possible future direction is attempting to develop measures that are sensitive to novice and experienced teachers' levels of knowledge. One of our tasks, the MKT, was sensitive among levels of experts and non-experts, but was not able to differentiate among levels of non-expertise. To assess a full range of knowledge, measures need to be able to differentiate among these levels as well. Measures that differentiate among ranges of knowledge levels could be developed and be especially useful to those interested in professional development of teachers to help them gauge knowledge levels of teachers who are not yet experts in the field. These measures could also be used instructionally with teacher training programs.

Another direction for future studies is to attempt to create and test more tasks designed to measure only one aspect of pedagogical content knowledge. It would be unusual to attempt to design a task that measures pedagogical content knowledge without measuring some aspects of subject matter knowledge since the two are unquestionably linked. However, it might be possible to measure only knowledge of students and content and not knowledge of teaching and content, for example. If measures of this type could be made, it would help teacher education programs and professional development providers individualize instruction for teachers.

In general, the results are encouraging and leave us with a large amount of evidence about the appropriate use of the tasks used in this study. Needless to say, results also raise many more questions to explore and avenues for future research.

References

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

Ashton, P., & Crocker, L. (1987). Systematic study of planned variations: The essential focus of teacher education reform. *Journal of Teacher Education, 38*(3), 2-8.

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.

Begle, E. G. (1972). *Teacher knowledge and student achievement in algebra* (Report No. 9). Stanford, CA: Stanford University.

Cavalluzzo, L C. (2004). *Is National Board Certification an effective signal of teacher quality*? Alexandria, VA: CNA Corporation. Available online at http://www.cna.org/documents/CavaluzzoStudy.pdf

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing*. New York, NY: Academic.

Chi, M. T., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Choi, J., Ahn, S., & Kennedy, M. (2008). *The role of teachers' subject-matter knowledge*. Unpublished manuscript. Michigan State University.

Chung, K. W. K., Baker, E. L., Brill, D. G., Sinha, R., Saadat, F. & Bewley, W. L. (2006). *Automated assessment of domain knowledge with online knowledge mapping.* (CSE Technical Report No. 692). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Delacruz, G. C., Chung, G. K. W. K., Heritage, M., Vendlinski, T., Bailey, A., & Kim, J.-O. (2007, April). *Validating knowledge elicitation techniques: Examining the relation between measures of content knowledge and knowledge of teaching algebra.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Eisenberg, T. A. (1977). Begle revisited: Teacher knowledge and student achievement in algebra. *Journal for Research in Mathematics Education, 8*(3), 216-222.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *Review of Economics and Statistics*, *89*(1), 134-150.

Heritage, M., Kim, J. O., & Vendlinski, T. (2008, March). *From evidence to action: A seamless process in formative assessment?* Paper presented at the annual meeting of the American Educational Research Association, New York.

Heritage, M., Kim, J. O., Vendlinski, T., & Herman, J. (2008). *From evidence to action: A seamless process informative assessment* (CSE Technical Report No. 741). Los Angeles, University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive science structure of U.S. history knowledge. *Journal of Educational Research, 89*, 206-218.

Herl, H. E., O'Neil, H. F., Jr., Chung, G. K. W. K., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior, 15,* 315–334.

Hill, H., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teacher's topic-specific knowledge of students. *Journal for Research in Mathematics Education*, *39*(4), 372-400.

Hill, H., 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., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11-30.

Klein, D. C. D., Chung, G. K. W. K., Osmundson, E., Herl, H. E., & O'Neil, H. F. (2002). *Examining the validity of knowledge mapping as a measure of elementary students' scientific understanding* (CSE Technical Report No. 557). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.

Krauss, S., Neubrand, M., Blum, W., & Baumert, J. (2008). *The professional knowledge of German secondary mathematics teachers: Investigations in the context of the COACTIV Project*. Paper presented at the 11th International Congress on Mathematical Education, Monterrey, Mexico.

Mandeville, G. K., & Liu, Q. (1997). The effect of teacher certification and task level on mathematics achievement. *Teaching and Teacher Education*, *13*(4), 397-407.

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

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.