# The Nature and Predictors of Elementary Teachers' Mathematical Knowledge for Teaching

Heather C. Hill Harvard Graduate School of Education

This article explores elementary school teachers' mathematical knowledge for teaching and the relationship between such knowledge and teacher characteristics. The Learning Mathematics for Teaching project administered a multiple-choice assessment covering topics in number and operation to a nationally representative sample of teachers (n = 625) and at the same time collected information on teacher and student characteristics. Performance did not vary according to mathematical topic (e.g., whole numbers or rational numbers), and items categorized as requiring specialized knowledge of mathematics proved more difficult for this sample of teachers. There were few substantively significant relationships between mathematical knowledge for teaching and teacher characteristics, including leadership activities and self-reported college-level mathematics preparation. Implications for current policies aimed at improving teacher quality are addressed.

*Key words*: Assessment; Elementary, K-8; Policy issues; Survey data; Teacher knowledge; Teachers (characteristics of)

This article outlines an effort to describe, on a broad scale, elementary teachers' mathematical knowledge for teaching. Few other topics have been the focus of such concern and resource investment over the past dozen years. Rarely can a national mathematics panel or commission meet without pointedly noting that teachers require strong knowledge of content to be effective and making recommendations for how that knowledge should be acquired (e.g., Greenberg & Walsh, 2008; NCTM, 2000; National Commission on Mathematics and Science Teaching for the 21st Century, 2000; National Mathematics Advisory Panel [NMAP], 2008). Financial resources have accompanied this policy concern. For example, between 2002 and 2007, the National Science Foundation and the U.S. Department of Education Mathematics and Science Partnerships spent nearly \$1.2 billion<sup>1</sup> providing content-focused mathematics and science learning experiences for

This research was funded by the National Science Foundation under grant NSF-0335411. Any opinions, findings, conclusions, or recommendations in this article are those of the author and do not necessarily reflect the views of the National Science Foundation. I would like to thank Deborah Loewenberg Ball, Merrie Blunk, Charalambos Y. Charalambous, Seán Delaney, Jennifer Lewis, Lesli Scott, Laurie Sleep, Geoffrey Phelps, Mark Thames, Thomas Tomberlin, and Deborah Zopf.

preservice and in-service teachers. Spending by states, districts, and teachers themselves has no doubt matched or exceeded this amount.

Anecdotal evidence suggests that there is good reason to make such investments. Both observations and interview data suggest that U.S. elementary teachers vary widely in their grasp of the mathematics required to teach this subject. Some scholars approach the problem as one of deficits or constraints: Teachers lacking mathematical knowledge are less likely to present material clearly and provide error-free content (Ball, 1990a; Borko et al., 1992; Cohen 1990; Heaton, 1992; Ma, 1999; Putnam, Heaton, Prawat, & Remillard, 1992; Stein, Baxter, & Leinhardt, 1990). Other scholars have studied this problem from the perspective of what knowledge affords teachers, noting that mathematically stronger teachers can do many tasks beyond simply solving problems in front of students. These tasks include sensibly interpreting and responding to student mathematical productions and producing more conceptually grounded mathematics lessons (Fennema & Franke, 1992; Fennema, Franke, Carpenter, & Carey, 1993; Lloyd & Wilson, 1998; Sowder, Philipp, Armstrong, & Schappelle, 1998; Swafford, Jones, & Thornton, 1997).

Despite this evidence, very little is known about the nature and predictors of elementary teachers' mathematical knowledge for teaching. In fact, most studies on this topic focus on a single teacher or a handful of nonrandomly sampled teachers. Although this permits a depth of analysis not enabled by other methods, it leaves open questions regarding the representativeness of findings. For instance, Ma (1999) conducted a multiple-case comparison of U.S. and Chinese teachers. She reported that mathematics problems evoking conceptual understanding of division of fractions, area and perimeter, and even place value proved difficult for a small sample of beginning and experienced U.S. elementary teachers. Whether this finding holds in the larger population and across more mathematical topics and tasks is not known.

Thus, we argue that there is a need for descriptive information regarding elementary teachers' mathematical knowledge for teaching. Teacher education programs must be focused where they will be most useful, and knowing which topics and tasks teachers find to be challenging provides one source of guidance. Identifying predictors of teacher knowledge could help district and school hiring officials, who often do not have access to mathematics-specific teacher test scores, identify more able candidates. And increasingly, new programs and policies come with implicit, but testable, assertions regarding the content knowledge of teachers: that teacher leaders have content expertise to help their peers and that alternatively certified teachers possess or quickly develop better mathematics knowledge than traditionally certified teachers. Policymakers also have focused recently on putting the best teachers in front of the most at-risk children (U.S. Department of Education, 2009).

In light of professional education needs and the assertions embedded in policy, we use data from a nationally representative sample of elementary mathematics

<sup>&</sup>lt;sup>1</sup> Retrieved December 21, 2007, from www.nsf.gov/awardsearch/ and www.ed.gov/programs/ mathsci/awards and totaled by my research assistant.

teachers to investigate teachers' performance on a set of mathematical knowledge for teaching (MKT) measures focused on number and operations. We chose MKT because there exist a set of measures in this domain that have been validated visà-vis instruction and linked to student outcomes (Hill, Blunk, et al., 2008; Hill, Rowan, & Ball, 2005). The MKT measures also have the advantage of focusing on job-embedded tasks, such as responding to students' mathematical productions and selecting accurate representations and explanations, rather than more abstract or higher level mathematics. Our expectation is that this job-embedded knowledge matters more in teachers' daily performance. We chose number and operations because it constitutes the majority of mathematics instruction in the elementary school grades (Rowan, Harrison, & Hayes, 2004). Our specific questions include the following:

- 1. What does elementary teachers' performance on a paper-and-pencil assessment suggest about the nature of their:
  - a. Knowledge within the strand of number and operations?
  - b. Knowledge of specific tasks of teaching?
- 2. Are there predictors of teachers' mathematical knowledge for teaching (MKT)?
  - a. How are teachers' reports of their own educational background related to their MKT?
  - b. Are those who have taken leadership positions in mathematics especially qualified in the area of mathematical content?
  - c. Do alternatively certified teachers possess greater amounts of MKT?
- 3. Are students of different socioeconomic status assigned to teachers who have, on average, lower MKT?

Finally, by comparing test performance of this sample to that of previous nonrepresentative samples (e.g., teachers attending mathematics professional development programs) and to teachers' own reports of their mathematical knowledge, we can also address an important methodological question: how best to measure teachers' mathematical knowledge. We report on the study's grounding, methods, and results in subsequent sections.

# LITERATURE REVIEW

In-service teacher knowledge of mathematics—and other subjects, for that matter—is seldom directly measured, particularly on a large scale. Although prospective elementary teachers in the majority of states take the Praxis II (ETS, 2010a), which contains a subscale composed of mathematics items (ETS, 2010b), this exam is calibrated to the knowledge requirements of beginning teachers and the results are not public. Few other efforts exist. One reason for the lack of large-scale research into elementary teacher mathematical knowledge involves issues

that surround assessments of teacher knowledge, including potential resistance from teachers and the need for a large number of items to achieve a reliable measure. Including a measure of teacher knowledge on the Schools and Staffing Survey conducted by the U.S. Department of Education (NCES, 2010), for instance, would not be feasible, given the comprehensive nature of that instrument. Another involves the expense of conducting a large-scale assessment. Although convenience samples are often free, it is unclear whether results from these samples generalize to the broader population of U.S. teachers. And obtaining broader samples requires a sampling strategy, efforts to strengthen response rate, and compensation for teachers' time investment—putting the acquisition of a representative sample beyond the budgets of many research projects investigating teacher knowledge.

Another complication is a lack of agreement on the knowledge teachers need for teaching and, thus, the types of items that should compose such an assessment. Some policymakers (Greenberg & Walsh, 2008; Massachusetts Department of Elementary and Secondary Education & Pearson Education, 2008) have designed tests composed mainly of pure content items representing more advanced mathematical topics than what teachers actually teach. In this view, knowing content "up the curriculum" is the most crucial component to producing high-quality mathematics in classrooms. Other researchers focus on the unique mathematical knowledge teachers may have in their specific grade range. This knowledge may be purely mathematical, yet specific to the work of teaching—what some call *specialized content knowledge* (Ball, Thames, & Phelps, 2008). It may also consist of *pedagogical content knowledge*—knowledge of how students learn content or of ways to teach specific topics (Shulman, 1986; Wilson, Shulman, & Richert, 1987). Together with basic grade-level content knowledge, Ball et al. (2008) term these teaching-specific forms of knowledge *mathematical knowledge for teaching*.

Notwithstanding debates over what to measure, the literature provides several warrants for a large-scale inquiry into elementary teachers' mathematical knowledge. One such warrant is that teachers' knowledge is arguably related to important educational processes and outcomes. Although the National Mathematics Advisory Panel argued that "research that has used teacher test scores and other ad hoc measures [to predict student achievement] has produced mixed results" (NMAP, 2008, p. 5-16), the panel also observed that the closer the measure of teacher knowledge to the work done in the classroom, the more likely a positive result will occur. For instance, Harris and Sass (2007) found no effect of teacher SAT scores on student achievement. By contrast, Hill et al. (2005) found that a measure of MKT predicted student achievement. And a recent study of first-year teachers in New York City (Rockoff, Jacob, Kane, & Staiger, 2008) showed specialized mathematics knowledge to be a better predictor of student mathematics outcomes than a series of other indicators, including general cognitive ability. Using a representative sample of German Grade 10 classes and their mathematics teachers, Baumert and colleagues (2010) found that teachers' pedagogical content knowledge was more predictive of student learning gains than content knowledge. Overall, this developing evidence suggests that teachers' mathematical knowledge-particularly

when conceptualized as more than a grasp of basic facts and procedures or even advanced knowledge—acts as a resource for student learning.

The literature also supports claims that teachers' mathematical knowledge relates to the quality of their classroom work. In one study, strong correlations characterized the relationship between 10 teachers' mathematical knowledge for teaching and mathematical elements of classroom work, such as the presence of teacher mathematical errors, the richness of mathematical work, and the depth of teacher interpretations of student mathematical productions (Hill, Blunk, et al., 2008). Studies using smaller samples and more in-depth analyses of specific topics (e.g., Ball, 1990b; Charalambous, 2010; Cohen, 1990; Heaton, 1992; Lloyd & Wilson, 1998; Wilson, 1990; see Fennema & Franke, 1992 for a review) found similar correspondences.

A second warrant for this study concerns the research questions left open by scholarly inquiry into teachers' mathematical knowledge. For instance, to our knowledge there has been no comprehensive comparison of content difficulty; even within relatively narrow topics, such as fractions, it is common for studies to focus on only a subset of problem types (e.g., Newton, 2008). Working with small teacher samples has also meant that inferences regarding what teachers find difficult are tenuous. Nevertheless, there is a general sense from reviewing this literature that rational number is particularly challenging for teachers of the elementary grades (see An, Kulm, & Wu, 2004; Ball, 1990a; Borko et al., 1992; Leinhardt & Smith, 1985; Ma, 1999; Sowder et al., 1998; Tirosh, 2000), but that problems may arise in any content area for at least some teachers. Additional and more comprehensive information about content difficulty would enable a firmer judgment about how to design teacher education curricula.

There are also questions regarding which dimensions of mathematical knowledge for teaching teachers find easier and which they find more difficult. Again, one might surmise that mathematical elements such as interpreting and using representations, providing mathematical explanations, and interpreting and responding to student productions are more difficult for teachers than simply solving mathematics problems (Borko et al., 1992; Ma, 1999; Thompson & Thompson, 1994; Hill, Blunk, et al., 2008). However, there is no comprehensive evidence for this point either.

The literature also provides few investigations of the predictors of teachers' mathematical knowledge. For instance, although there is considerable—and conflicting—evidence regarding the effect of mathematics coursework, teaching methods, and professional development on student outcomes (NMAP, 2008), little is known about whether such coursework is associated with teacher mathematical knowledge. Likewise, there is little information regarding whether teachers' own mathematical self-concept is related to their knowledge as objectively measured. This is a key methodological point, as many professional development evaluations and research programs rely principally upon teacher self-reports of knowledge or learning (e.g., Garet, Porter, & Desimone, 2001). Principals and district officials may also benefit from knowing whether any of these background characteristics and/or self-reports can help identify mathematically knowledgeable teachers.

Similarly, many current policies are based on seldom-investigated assumptions about the characteristics of individuals entering and leading the profession. For instance, many argue that relaxing teacher education and certification requirements will encourage better-qualified candidates to enter teaching (e.g., Hess, 2002; Paige, 2002), and research suggests that many prospective teachers have taken advantage of alternative programs (Peterson & Nadler, 2009). However, existing research (Cohen-Vogel & Smith, 2007; Tournaki, Lyublinskaya, & Carolan, 2009) suggests few differences among traditionally and alternatively certified teachers in general qualifications and teaching styles. Likewise, the qualifications of those assuming mathematics leadership positions deserve attention. Although the current literature is rife with talk of career ladders, math coaches, and peer assistance and mentoring (e.g., Knight, 2009; Showers & Joyce, 1996), little is known about the intellectual capabilities of the individuals holding those positions.

The last warrant for this study involves equity. Several studies (Hill, 2007; Hill & Lubienski, 2007; Hill et al., 2005; Loeb & Reininger, 2004) suggest that knowledgeable teachers are inequitably distributed across student populations, with mathematically stronger teachers serving more affluent and less racially mixed populations. There is some indication that the mechanisms that support such inequities result from teachers' preferences regarding employment. Recent research (Boyd, Lankford, Loeb, & Wyckoff, 2003) shows that teacher labor markets are largely local; for example, nearly "83% of teachers entering the New York State public school workforce took jobs within 40 miles of their home" (p. 71). If the results from this study generalize to other locations, graduates from weaker school systems may tend to return to teach in those school systems, perpetuating the cycle of unequal access to educational resources.

To extend existing knowledge and address the need for information regarding the effects and potential effects of current policies, we conducted a large-scale study of elementary teachers' mathematical knowledge for teaching. We describe this study in the following sections.

#### **METHOD**

Although this study was quantitative, this article is fundamentally descriptive. We wish to know: What can be said about elementary school teachers' knowledge of specific topics within number and operations? Which types of mathematics teaching tasks prove easier and which prove more difficult for teachers? Which teacher characteristics can be associated with possession of more or less of this knowledge? How are teachers distributed across schools and students? Because we employ a cross-sectional design, we cannot make causal statements regarding relationships between these variables. However, we think of descriptive inference as more than the poor cousin of causal inference: For sensible instructional policies and future research to proceed, an accurate portrayal of the state of elementary teacher mathematical knowledge is necessary.

# Sample

The goal of sample selection was to represent the population of elementary school teachers with mathematics teaching responsibilities in the 48 contiguous states. To do so, our data contractor, the Institute for Social Research at the University of Michigan (ISR), determined the appropriate final target sample size to be approximately 1,090, stratified the elementary schools in the 2005–2006 Common Core Database by geographic region and urbanicity, then drew an initial random sample of schools from each stratum to total 1,200 schools. Next, ISR consulted a database maintained by Quality Educational Data, a firm that develops and maintains a list of teachers in each school. Using this list, ISR staff telephoned each school to confirm and/or update the roster of grades 1 through 6 teachers with instructional responsibilities in mathematics. These calls resulted in 1,110 schools confirming or updating their rosters. From this set, 1,090 schools were sampled to take part in the study and one teacher was sampled at random from within each school. Twenty teachers left their school prior to the first mailing, leading to a final sample size of 1,070 schools.

A first wave of surveys reached teachers in late February 2008; teachers were reminded up to three times to return their surveys, and each was paid \$50 for completing the survey. This resulted in a dataset of 625 teachers, a 59% response rate. Nonresponders were concentrated in the south (56%) and west (54%), as well as in large cities (44%). Although this response rate may be considered low by industry standards, it is routine for studies of this kind, which require that teachers spend between 60 and 90 minutes solving mathematics problems. In addition, descriptive data suggest that our respondents did not differ-in at least demographic characteristics-from those in a 2000 survey done by Horizon Research (Weiss, Banilower, McMahon, & Smith, 2001). In fact, our respondents included more nonwhite teachers than the earlier study and were roughly similar in experience and gender (see Table 1). Weights calculated to take both sampling and nonresponse into account are applied to the analysis reported subsequently, and we comment on the potential effect of the low response rate in the Results section. For more information on the weighting procedure used, Hubbard (2008) is available upon request from the author.

The result is a sample that is fairly typical of the U.S. elementary teacher population. Table 1 shows the distribution of respondent characteristics, including experience, race, gender, and grade. The sample is overwhelmingly white, female, and relatively inexperienced; over half the sample has been teaching 10 or fewer years. Very few teachers—less than 5%—reported a first language other than English.

#### Measures

Measures development took place in the context of a larger project intended to provide instruments to evaluators, scholars, and others studying the impact of preservice education and professional development on teacher knowledge. This

Table 1	
Teacher	<b>Characteristics</b>

	Percent of teachers with this characteristic	Percent of K–4 math teachers in Weiss et al. (2001) with this characteristic
Experience		
0–4 years	26.6%	32% (0-5 years)
5–10 years	28.4%	18% (6-10 years)
11-20 years	23.4%	26% (11-20 years)
20+ years	21.7%	29% (21+)
Race		
Hispanic	6.5%	3%
Black, not of Hispanic origin	4.5%	4%
White, not of Hispanic origin	79.7%	90%
Asian or Pacific Islander	1.4%	0%
American Indian or Alaska		
Native	0.4%	2%
Multiracial/biracial	1.1%	
Other	0.9%	
No response	7.6%	
Sex		
Female	89.5%	96%
Male	8.6%	4%
Grade		
1	25.5%	
2	22%	
3	20.2%	
4	17.5%	
5	18.3%	
6	5.7%	
7-8	1%	
9 or higher	0.3%	
Not currently teaching mathematics	4.7%	

*Note.* Percentage totals do not sum to 100 due to rounding. In the case of grade level, some teachers taught more than one grade level.

intended use of these measures conditioned their design in two important ways. First, we wanted to design a set of measures with strong reliability-that is, measures that can accurately differentiate between two individuals only slightly apart on the underlying trait, mathematical knowledge for teaching. Thus, we targeted the assessment such that, by design, the average teacher would score only 50% correct. Assessments on which nearly all teachers get nearly all items correct or incorrect are neither reliable nor usable in the context of research programs. To reach our 50% goal, we aimed to produce items with a wide range of difficulties, from ones that nearly all teachers answered correctly to ones that nearly all could not answer correctly. This approach differs from criterion referencing, which includes the creation of a detailed construct map and the establishment of benchmarks for performance. Although one might draw conclusions about teacher proficiency levels from a criterion-referenced test, one cannot do so from the design we adopted. Thus, this study cannot make claims about the overall strength and appropriateness of elementary teachers' mathematical knowledge for teaching; instead, as represented in the first research question, the article examines the nature of that knowledge (e.g., which topics the sampled teachers found easier or more difficult).

Second, we wanted to write items that would measure teachers' mathematical knowledge for teaching (Ball et al., 2008) rather than their knowledge of high school or college mathematics (e.g., calculus, trigonometry, differential equations) or their pure mathematical aptitude or skill. We chose to use the MKT framework over other possibilities (e.g., one based on the NCTM standards, high school or college coursework, or various commission reports' findings) for several reasons. To start, it was based in the real work teachers do in classrooms, with children. In fact, it was developed from a grounded study of mathematics teaching that involved observing teachers, students, and their interactions with mathematical content in real classrooms. The MKT framework also specifies a way of thinking about the various mathematics-related tasks teachers are asked to complete in classrooms, as opposed to a list of topics that teachers should master and upon which they should be assessed.

Finally, MKT incorporates multiple forms of teacher knowledge that may affect instruction, in line with Shulman and colleagues' observations about the nature of teacher knowledge (Shulman, 1986; Wilson et al., 1987). One element of MKT is *common content knowledge* (CCK)—an ability to correctly recall and execute grade-level appropriate ideas and procedures (Ball et al., 2008). CCK represents the knowledge that we expect mathematically literate nonteaching adults to hold and also represents the content traditionally taught to elementary school students. However, we also sought to assess teachers' *specialized content knowledge* (SCK), the mathematical knowledge that lies beyond that held by a well-educated adult. Examples include knowing mathematical explanations for common rules or procedures; constructing and/or linking nonsymbolic representations of mathematical subject matter; interpreting, understanding, and responding to nonstandard mathematical methods and solutions; deploying mathematical definitions or proofs in accurate yet also grade-level-appropriate ways; and diagnosing errors in student

work. This knowledge, like CCK, is wholly mathematical; to get SCK items correct, one need not know about students, instructional methods, or materials.

The first two sample items in Appendix A provide examples of these two aspects of teachers' content knowledge. These items, as with others discussed below, were released from older forms because they do not perform well in psychometric analyses; nevertheless, they are instructive for the types of knowledge we wish to assess. In the first item, a teacher is thinking about the number zero and asks her sister about specific statements in a new textbook. This item then asks respondents to determine whether 0 is a number (yes), whether it is even (yes), and whether the number 8 can be written as 008 (yes). Although this is not knowledge that every U.S. adult may possess, it is purely mathematical knowledge, common across professions such as accounting, computing, and engineering.

By contrast, the next item, which asks teachers to choose a diagram that does not represent

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

calls on knowledge and skills unlikely to be held by nonteachers. Whereas most adults would know a conventional algorithm for multiplying these fractions, we expect that few would have had experiences identifying the unit (parts [a], [b], and [d]) and "seeing" the multiplicands and products. In part (b), for instance, one might see each large rectangle as a unit, note that  $1^{1/2}$  of those units are shaded gray, and then identify the hash marks as 2/3 of the originally stated quantity,  $1^{1/2}$ . More important, many adults might fail to notice that the  $1^{1/2}$  in part (c) is not represented with the same unit, and thus we cannot define 1 and  $1^{1/2}$ . This knowledge is mathematical in nature—the respondent does not need to possess knowledge about how students learn content or about the best way to represent content to learners (see Shulman, 1986). However, it is different from the common knowledge evoked previously, and thus, Ball et al. (2008) categorize it as SCK.

In this survey, we also sought to measure two dimensions typically associated with *pedagogical content knowledge*. The first—*knowledge of content and students* (KCS)—is exemplified by the third item in Appendix A. In it, a teacher is considering problems that can be used to introduce students to proportional reasoning. In considering these problems, the teacher might note that options (a) and (c) contain scale factors that lend themselves to simpler solutions. In (a), for instance, students might notice that they can increase the number of buttons by half to arrive at Mr. Tall's height, as the scale factor is 1.5. Option (c) is similar. In (b) students might notice that the scale factor between Mr. Short and Mr. Tall is 1.2; however, there is no easy way to "scale up" 7 by this factor. The teacher examining these problems may decide that because students often discover this method for solving proportional reasoning problems before more formal cross-multiplication, she will not assign (b) in her introductory problem set.

The second category of pedagogical content knowledge covered by this survey was *knowledge of content and teaching* (KCT). Items categorized here ask teachers to design instruction based on considerations of both content and likely student responses to content. This includes choosing and sequencing representations and examples, constructing problems with similar interpretations (e.g., partitive model of division), and deciding how to select student responses to highlight and move mathematical discussions forward. Item 4 in Appendix A illustrates this type. Ms. Miller has a very specific instructional purpose: she wants her students to develop an initial definition for triangle and then improve that definition by testing it on different shapes. Each poster affords general information about shapes and thus could be instructionally useful. However, only one is useful for the purpose she has chosen. Poster (a) would enable students to use their initial definitions to decide that circles, squares, and rectangles are not triangles. Most elementary students would easily recognize these as not-triangles, and the definitions would not be challenged. Poster (d) includes only triangles and an inaccurate set of descriptors. Poster (c) contains only triangles, and although it might promote students' understandings of triangles that are not oriented parallel to the sides of the page, it would do little to help other aspects of the definition-building process. Poster (b) contains triangles and other polygons, some strategically designed to provoke confusion or uncertainty in students (e.g., the lightning bolt). Although other triangles meet the definition, many students would not recognize them as triangles because of their orientation or dimensions (e.g., the wedge). Only the poster with this variety should allow students to fulfill the goals of instruction and improve their definition of triangle (see Clements, Swaminathan, Zeitler-Hannibal, & Sarama, 1999).

Finally, this assessment focused only on number and operations (arithmetic) content. We did so because evidence from large-scale studies indicates that this topic comprises roughly 50% of the instruction delivered in elementary schools (e.g., Rowan et al., 2004) and because including all topics across the curriculum would mean insufficient power to carefully compare the difficulty of item content and knowledge associated with particular tasks of teaching. This form included 20 items involving whole numbers, 6 involving integers, and 39 involving rational numbers.

Based on the categories in Ball et al. (2008), this form contained 6 CCK items, 23 SCK items, 1 KCS item, and 7 KCT items. The imbalance was driven by our preference for specialized over common content items and a notable lack of success in writing KCS and KCT items (see Hill, Ball, & Schilling, 2008). A discussion of the implications of this imbalance for the analyses follows.

Prior to conducting the study, these items were reviewed by both internal project members (including one mathematician) and four external mathematicians for mathematical accuracy. To understand better what we were measuring, we undertook a variety of validation work. A set of cognitive tracing interviews indicated that teachers' answers do, in general, reflect their underlying mathematical thinking (Hill, Dean, & Goffney, 2007). Only a small fraction of teacher responses (roughly 8%) demonstrated inconsistencies between their thought process and the multiple-choice answer ultimately selected. A study of 10 elementary teachers showed the mathematical quality of their instruction, estimated by analyzing nine videotaped lessons per teacher, to correlate highly with their MKT scores (r = 0.74, p < 0.05;

Hill, Blunk, et al., 2008). A similar study of 26 middle school teachers (Hill, Umland, & Litke, 2010) demonstrated moderate to strong correlations between specific dimensions of mathematics teaching (teacher errors [-0.65, p < 0.01], richness of the mathematics [0.32, p > 0.05], responding to students productively [0.51, p < 0.05]) and MKT score. We and others have also linked teachers' performance on the elementary MKT measures, and more specifically two different sets of common and specialized knowledge items, to gains in student achievement on standardized assessments: the students of teachers who answered more items correctly gained more over the course of a year of instruction, controlling for student background and classroom composition (Hill et al., 2005; Rockoff et al., 2008). Finally, we have conducted content validity checks, ensuring that our item pools provide fair coverage of the topics (e.g., number and operations) they are intended to represent. Although by no means definitive, these validity checks suggest that this survey-based measure of teachers' mathematical knowledge for teaching is a strong predictor of the quality of their classroom practice and, to a lesser degree, student outcomes.

In addition to measuring mathematical knowledge for teaching, our survey carried a series of questions intended to gauge teachers' backgrounds and activities. Some descriptors, such as grade level; years of experience; specific leadership activities; and content, methods, and professional development experiences, were measured with single items. An open-ended question asked teachers who did not attend traditional teacher education programs to report their mode of entry. Responses included district internship programs overseen by local universities, state internship programs, Teach for America, and a variety of specific programs (e.g., Preparing Responsive Educators Program). This question was then handcoded to either traditional mode of entry (teacher education program) or alternative program. Two other constructs were measured via Likert scale. The first measured teachers' self-reported instructional practices in the area of student explanation, analysis, and proof. Three separate items required teachers' estimates of how often students engaged in each activity. These items were averaged to form a scale in which more positive values indicate more of these mathematical practices  $(\alpha = 0.82)$ . The second construct measured teachers' mathematical self-concept of ability (Newton, 2009), including three items that gauged their estimates of their own mathematical knowledge, one item that tapped whether they believe that their mathematical knowledge is sufficient for teaching the subject to students, and one item that asked whether they consider themselves a master (expert) mathematics teachers. A factor analysis found these items scaled well together, and they were combined into a common scale in which positive scores indicate a better mathematical self-concept ( $\alpha = 0.81$ ).

# Data Analysis

We began by conducting a factor analysis to determine the structure of the data and to ascertain whether items loaded as we expected. MKT items were categorized a priori as CCK or SCK, KCS, or KCT using Ball and colleagues' (2008) definitions of knowledge types; this categorization was then compared to exploratory factor analysis results. We used ORDFAC (Schilling, 2005), a program written specifically for our MKT data that enabled the inclusion of *testlets*, or stems with multiple related items beneath. Results of the factor analysis were not clear-cut; items did not, in fact, load cleanly onto hypothesized factors or even come close to doing that.

Because of this, we elected to combine all items into one indicator, which we named mathematical knowledge for teaching (MKT). We chose to do this for several reasons. First, we did not have a sufficient number of items to return adequate person-level reliabilities for most subscales. Second, although we might have omitted the one KCS item or constructed a measure of only SCK items, this would have had the effect of decreasing the measure's accuracy and reducing the amount of information provided about various aspects of teachers' knowledge. In general, omitting items is not warranted unless there is a strong theoretical and/or empirical rationale for doing so. Similarly, we could have confined the data to a subset, those for which predictive validity has been established (e.g., items contained in Hill et al., 2005), yet we would have faced the same problem as above-a marked reduction in reliability. Third, the ideal composition of an MKT measure is, in fact, unknown; until we have more information regarding which dimensions contribute with which weight to student outcomes, we can only guess what such a measure should look like. Even with imperfect balance among the dimensions, however, we believe that the instrument provides a first approximation to a broader measure of MKT.

In total, the MKT measure used in this study included 37 stems and 65 items. Items outnumbered stems because some stems (problem situations) have more than one question for teachers to answer, as in the case of item 1 in Appendix A. All items were used in every analysis unless otherwise noted.

Teachers' answers to the survey were entered into an Item Response Theory (IRT) model using Bilog 3.0 (Zimowski, Muraki, Mislevy, & Bock, 2003). IRT uses teachers' correct and incorrect answers to return person parameters, or scores, expressed in standard deviations, with mean 0 and standard deviation 1. Scores were normally distributed between -2 and +2 with only a few outliers. Teachers' scores are thus interpreted as the distance for a specific teacher from the average teacher in the sample, with higher scores indicating higher relative ability and lower scores indicating lower relative ability. We chose to express teacher scores in standard deviations rather than the more readily interpretable percent correct because percentages do not map onto the underlying MKT dimension linearly (e.g., in terms of the underlying MKT dimension, the difference between two individuals in the 5th and 10th percentiles is not the same size as the difference between two individuals in the 50th and 55th percentiles). We used a two-parameter IRT model to score teacher responses; these models construct individual scores by overweighting items with strong person-discrimination indices and underweighting items with low person-discrimination—in essence, underweighting items that mainly consist

of either noise or different constructs. Finally, although IRT models typically describe the accuracy of test scores in terms of test "information" provided at different points along the distribution of respondents, Bilog also translates this measure into the more interpretable internal reliability coefficient, which we report here. The person-reliability estimate for the MKT measure was 0.91.

In addition to person-level scores, IRT models also yield information about items. Items were described by two parameters, their slope (a discrimination index) and difficulty (distance from average of the scale, expressed in standard deviation units). Higher slopes indicate items that yield more information about examinees; higher difficulties indicate more difficult items. To answer the first research question, about the nature of teachers' MKT, we used a subset of the data consisting only of items with an adequate discrimination index (slope above 0.4)<sup>2</sup> resulting in a subset containing 48 items. We took this approach because the difficulties of items with poor discrimination indices are often estimated inaccurately. Descriptive analyses of the relationship between teacher characteristics and MKT score used all 65 items scaled as described above. We made this decision because the reliability of the overall measure was not affected by their inclusion. Analyses consisted mainly of calculating frequencies, correlations, and a series of simple regressions.

In reporting relationships between teacher characteristics, we needed to make a decision about the need for a correction in significance level due to the potentially large number of estimated correlations (roughly 45, if all teacher characteristics are correlated with one another as well as with MKT). One option was to calculate significance levels with a Bonferroni correction, which is conservative regarding significance levels. Another option was to present only the main correlations of teacher characteristics to each other. On the theory that these correlations may be of interest to some readers, we display them in Appendix B, although we do not discuss their significance levels.

## RESULTS

#### Item Difficulties

Our goal for item writing had been to design an assessment such that (a) there was wide dispersion of item difficulties and (b) the average item was answered correctly only 50% of the time. We succeeded in attaining that goal. Using a subset of the data that consisted of only those items with a reasonable discrimination index (IRT slope > 0.4), roughly half were above and half below the 50% mark. The distribution of teacher scores was roughly normal. Because this matches the aforementioned goal in test development, we cannot make conclusions regarding the relative knowledge of U.S. elementary teachers in mathematics.

<sup>&</sup>lt;sup>2</sup> The decision to use 0.4 was done on the advice of a psychometrician.

Continuing to examine items with adequate discrimination indices, we looked for patterns in item difficulty. This analysis showed that there were no differences in difficulty by content (whole numbers, rational numbers, integers; p = 0.17), or by certain subcategories of specialized knowledge: interpreting nonstandard solution methods, using representations, and choosing examples. However, two item descriptors were highly predictive of difficulty. First, CCK items were, on average, much easier than items designed to represent SCK. There was a difference of 0.68 standard deviations in item difficulty (p = 0.05). Table 2 shows the distribution of item difficulties by content and by knowledge type assessed. Second, the 10 specialized knowledge items that focused specifically on explanations for mathematical ideas and procedures were more difficult than the average item in the set as a whole. (There was a difference of 0.65 standard deviations, p = 0.13.)

Although this project does not release most items to ensure their secure use in ongoing program evaluations, a description of specific easier and more difficult items can help illuminate this trend. The least difficult dozen items included several that asked teachers to work with place value concepts, composing and recomposing numbers, and representing decimals. Another two easy items tapped the conceptual underpinnings of two whole number procedures, multidigit subtraction and long division. For subtraction, respondents were asked to differentiate between student responses, one of which provided a conceptual explanation for the standard procedure and two of which did not. For division, respondents were asked to compare two division word problems and select a response that identifies how they differ. In this case, the use of the remainder was different in each problem.

The most difficult dozen items included, notably, an item that asks teachers to identify the reason the standard U.S. long division algorithm works, an item that asks teachers to identify a correct representation for integer subtraction with

Average Item Difficulty by MKT Doi	main and Content	
Item type	Average item difficulty	Number of items
MKT Domain		
Common content knowledge	-0.45	14
Specialized content knowledge	0.24	29
Knowledge of content and teaching	1.05	5
Average/Total	0.12	48
Content		
Whole numbers	0.46	16
Rational numbers	-0.07	26
Integers	0.08	6
Average/Total	0.12	48
Rational numbers Integers Average/Total	-0.07 0.08 0.12	26 6 48

Table 2

Average Item	Difficulty	by MKT	'Domain	and Conten
--------------	------------	--------	---------	------------

black-and-white chips, an item that asks how to interpret remainders in a division of fractions problem, and one that asks teachers to identify the reason one cannot divide by zero. These items are a mix of common and specialized content knowl-edge. Some are purely mathematical, although admittedly complex. For instance, the problem "Michelle needs 4/5 cups of flour to make a batch of play-dough. How many batches can she make if she has 6 cups of flour?" does not require knowledge special to teaching, but interpreting the remainder (1/2) as either half of a cup of flour or half of a batch of play-dough requires careful thought.

Other difficult items, such as the integer subtraction representation and explanation for long division, were categorized according to the map laid out in Ball et al. (2008) as tapping specialized content knowledge. Whereas many professionals use integers in their daily work, seldom do nonteachers need to represent integers using concrete manipulatives. The same is true of the explanation for long division. Although most adults are familiar with the long division algorithm, few may understand why it works.

Several other difficult items stand out because they contain content that is potentially less demanding but ask teachers to make more subtle judgments. Two SCK items ask teachers to identify, based on student statements, the student who has the most advanced understanding of particular topics (fractions and whole number subtraction). Two KCT items ask teachers to identify the best numeric example for a particular purpose—demonstrating that the associative property can sometimes make expressions easier to evaluate and teaching an initial lesson on primes and composites. Although we expect that few teachers would have trouble stating the associative property or identifying prime and composite numbers, the additional knowledge involved in teaching these topics is, we suspect, nontrivial. For instance, the teaching task of showing that the associative property can be helpful in evaluating certain expressions makes some examples better than others. Consider the following four possibilities:

a. 
$$(27 + 54) + 6 =$$
  
b.  $(833 \times 5) \times 20 =$   
c.  $(45 + 29) + 17 =$   
d.  $55 \times (6 \times 37) =$ 

The associative property can be used to regroup terms for any of these expressions. However, for only two of the expressions (a and b) does using the associative property make the expression easier to evaluate. For instance, in (a) students might first add 27 and 54, then 6. However, by combining 54 and 6 first, the result (60) becomes easier to add to 27. The same principle applies to (b). In the other cases, regrouping results in equivalently complex computations. Thus, for teachers who wish to demonstrate why the associative property can make certain expressions easier to evaluate—an important component of strategic competence (NRC, 2001) and a precursor to algebraic thinking—choosing examples strategically (and making use of those choices) is a critical skill. In fact, we argue that knowing that the associative property (and the commutative and distributive properties as well) has a mathematical purpose beyond simply allowing us to rearrange numbers is a key component of teacher knowledge.

This description of content that was difficult for teachers is not meant to suggest that elementary teachers of mathematics are not prepared to teach this subject. Because the design of the instrument does not allow inferences regarding teachers' performance relative to any benchmark or criterion, we cannot say whether teachers require such content knowledge to teach well. Instead, we summarized these results to suggest directions for teacher education and professional development, and we comment on this issue further in the conclusion.

#### Describing Teachers' MKT

Shifting now to use all MKT items rather than only a subset of items, as we did in the last analysis, we focus on the second and third research questions, which ask whether any characteristics and experiences correlate with teachers' MKT. Descriptive statistics indicate substantial variance in each of these teacher characteristic indicators. For instance, Table 3 shows this sample of teachers as actively engaged in professional learning activities. Most teachers report having enrolled in both mathematics content coursework and mathematics methods coursework. Most also engaged in mathematics professional development within the past year, although the modal level of engagement, at less than 6 hours, is quite low, particularly in comparison to those same teachers' reports of professional development in other subjects. And a sizeable fraction of the sample-28% across the three types of leadership experiences listed in Table 4-engaged in a leadership activity of some sort within 3 years prior to the survey. Although we enrolled a negligible number of mathematics coaches, per se, in the study,<sup>3</sup> both peer mentors and in-service providers have responsibility for conveying mathematics to other teachers. Individuals in these roles are, in our opinion, also likely to become mathematics coaches in subsequent years.

The set of mathematics self-concept of ability items shows teachers appear to regard themselves as having adequate, if not exceptional, mathematics knowledge. Most agree or strongly agree with statements worded to indicate having sufficient mathematical knowledge for their job but are more lukewarm with regard to statements indicating content and teaching mastery. The mean score is 3.54 on this scale, with scores ranging from 1 to 5 and a standard deviation of 0.72. Table 5 shows that teachers did report engaging their students in mathematically challenging practices (explaining, analyzing, and proving) on a semiregular basis. The mean on this scale is 4.08, with standard deviation 1.12 and values ranging from 1 (never for any of these activities) to 6 (daily for all activities).

Turning now to correlations, Table 6 shows that the relationships between teacher background and MKT are relatively weak. This is particularly true for indicators of teachers' educational background. Teachers who reported taking more

 $<sup>^{3}</sup>$  This occurred partly by design; we sampled only practicing classroom teachers. The four coaches who participated in our study also had teaching responsibilities.

Mathematics coursework/ Mathematics methods coursework	Mathematics: Graduate or undergraduate	Mathematics methods: Graduate or undergraduate
No classes	5%	13%
1-2 classes	37%	65%
3-5 classes	45%	18%
6 or more classes	13%	4%

 Table 3

 Teacher Participation in Professional Learning Activities

Mathematics professional development/Other professional development within the past year	Mathematics	Other subjects
None	27%	4%
Less than 6 hours	37%	15%
6–15 hours	21%	35%
16-35 hours	9%	26%
35+ hours	6%	20%

*Note.* Numbers in the table are percentages of total sample engaged in each category of activity.

# Table 4 Relationship Between MKT Score and Recent Leadership Experiences

	Percent of teachers with this leadership experience	Mean M (in standard	KT score l deviations)
Within past 3 years		Leadership	No leadership
Served on a district or school mathematics committee	19%	0.18	-0.05*
Acted as a peer math coach or mentor while continuing my regular teaching duties	14%	0.13	-0.03
Taught an in-service workshop or course related to mathematics or mathematics teaching	6%	0.17	-0.02

\* Difference between leadership and no leadership means is significant (p < .01).

mathematics methods and mathematics coursework scored negligibly higher on the measure of MKT (r = 0.06 and 0.09, respectively). Although the correlation with mathematics content courses is significant, it is not substantively large. There were no significant relationships between MKT and mathematics-related professional development experiences, suggesting that extensive professional development participation is not an indicator of mathematically knowledgeable teachers. Finally, a separate analysis shows that scores from the roughly 30 teachers who entered teaching without graduating from a formal teacher education program suggest there is no relationship between taking one of these alternative pathways and stronger content knowledge (difference of means = 0.09, p > .50).<sup>4</sup> Although the sample size of nontraditionally certified teachers is small, these results accord with other studies in the field (Cohen-Vogel & Smith, 2007; Tournaki et al., 2009).

There is no significant correlation between teachers' MKT and their reports of engaging students in analysis, explanation, and proof. There are many potential reasons for this finding, and the relationship between MKT and the cognitive demand of student work merits further study.

Several positive and significant relationships stand out in Table 6. Teachers' reported current grade-level assignment is among the stronger associations with MKT, with a correlation of 0.30 (p < .001). K–1<sup>5</sup> teachers score almost one half standard deviation below the sample average; fifth-grade teachers score a third of

	Never	Less than once per month	1–3 times per month	1–2 times per week	3–4 times per week	Every day
Explain an answer or a solution method for a particular problem.	1%	1%	6%	22%	31%	39%
Analyze similarities and differences among repre- sentations, solutions, or methods.	4%	9%	19%	29%	26%	12%
Prove that a solution is valid or that a method works for all similar cases.	15%	18%	18%	25%	15%	9%

# Table 5 Mathematics Instructional Practices

Note. Percentages are rounded to the nearest whole number.

<sup>&</sup>lt;sup>4</sup> Because regular and alternatively trained teachers differ in average years of experience, we also conducted a comparison of alternatively trained teachers with a randomly chosen sample of teachers with comparable experience. Again, there were no differences in MKT between the two groups.

<sup>&</sup>lt;sup>5</sup> Anticipating that there may be blended K-1 classrooms, our survey asked whether teachers taught at K-1 rather than first grade.

Table 6 Correlations l	3etween MK1	r and Teacher C	haracteristic	S					i	
				Math	Other	Math	Math		Math	School free/
	MKT	Years of	Grade	prof'l	prof'l	content	methods	Instructional	self-	reduced
	score	experience	taugnt	uev.	nev.	course	course	practices	concept	Iuncn
MKT	1.00	•60.0	0.30**	0.03	0.02	*60.0	0.06	0.01	0.25**	+60.0-
score		(582)	(200)	(613)	(909)	(609)	(909)	(298)	(612)	(619)
* Significant	at <i>p</i> < .05. **	Significant at p	< .001							

Nature and Predictors of Elementary MKT

a standard deviation above, with a roughly linear trend in between. This relationship may be due to a recency effect—upper-elementary teachers are more likely to have taught rational number, which comprises a substantial portion of the test—or may reflect teacher selection of and assignment to grade level based, in part, on mathematical ability.

Teacher experience is modestly related to mathematical knowledge for teaching. More experienced teachers—and particularly those with over 20 years of experience—have more MKT, and this overall relationship looks approximately linear (Table 7). This may occur through one of several paths. One is that teachers learn on the job through the various resources available to them: curriculum materials, colleagues, students, and professional development. Another possibility is that the trend illustrated in Table 7 results from a cohort effect; recent work (Bacolod, 2007; Corcoran, 2007) has demonstrated changes in the cognitive skills of entering teachers. As employment opportunities opened for women during the 1970s and 1980s, fewer top-decile women entered the teaching profession. Finally, the pattern shown in Table 7 could result from both learning on the job and a cohort effect.

Teachers' mathematical self-concept in Table 6 correlates to their MKT score at 0.25 (p < .001). Although this is among the stronger associations in the data, from an absolute perspective, it seems weak. These self-concept questions were placed at the end of the survey, after teachers had worked through roughly 1 hour of mathematics problems. The self-concept questions focused mainly on mathematical content knowledge rather than mathematics pedagogy or perceived effects on student learning (self-efficacy). The reliability of both this set of items and the MKT assessment is high, suggesting very little attenuation due to measurement error. Finally, this association remains at 0.25 or less regardless of whether question wording focuses on mathematical adequacy (e.g., "My knowledge of number and operations is adequate to the task of teaching these subjects.") or being a master (e.g., "I consider myself a 'master' mathematics teacher.").

Similar results occur for the relationship between mathematics leadership positions and MKT. Table 4 shows that teachers who have taken on leadership roles within the past 3 years have only slightly better-than-average MKT. The difference is largest for individuals who served on a district mathematics committee (0.23 standard deviations) and smallest for acting as a peer mentor (0.16 standard deviations). The difference between individuals serving and not serving on a mathematics

Tears of Experience and MIKT		
Years of experience	MKT score	
0-4 years	-0.09	
5-10 years	-0.04	
10–20 years	0.04	
20+ years	0.1	

Table	7
Years	of Experience and MKT

committee is significant (p < .05); for the other two variables, the difference is insignificant. This lack of strong correlations is discouraging, in that mathematicsspecific leadership of any kind—policy leadership, instructional leadership, gradelevel leadership—requires content expertise. Additional correlations (not shown) demonstrate that teachers who served in these leadership roles were not more likely to have strong mathematical self-concepts than those who did not (r = 0.02; p > .50); they were also only slightly more likely to use higher-demand instructional techniques (r = 0.08, p < .05) than those who did not report leadership activities.

Clearly, individuals may be chosen or volunteer for leadership positions because of key nonmathematical areas of expertise, including pedagogical expertise, enthusiasm for mathematics, the ability to work with others, or the ability to navigate the politics of mathematics curriculum and instruction. Thus, it is plausible that lowknowledge individuals can serve as excellent leaders or mentors in certain situations. However, individuals lacking content expertise can provide only limited guidance with regard to assisting teachers with representations, explanations, student productions, and other mathematics-related classroom practices.

Finally, there is a slight tendency for lower-knowledge teachers to work in higherpoverty school districts (r = -0.09, p < .05 between student free and reduced-price lunch eligibility and MKT). This is similar to estimates from other samples (Hill, 2007; Hill et al., 2005) and suggests that the students who may be most in need of mathematically knowledgeable teachers are slightly less likely, on average, to get them.

Although this is a purely descriptive study, we can learn about several key points from a regression. Instead of attempting to discern causality, we instead are interested in several factors: the significance of key variables controlling for other (potentially correlated) predictors, the overall amount of explained variation in teacher knowledge, and how explained variance changes with the addition of specific variables. Table 8 shows a set of regressions that provide this information.

In keeping with the preceding intent for the regression, variables were entered in four stages. In Table 8, Model 1 shows a regression of only background characteristics—specifically grade level, experience, and school-level free/reduced lunch eligibility—on MKT. Two of the three were significant and signed in the direction suggested by the preceding discussion; free/reduced lunch eligibility was significant only at the p < 0.10 level. In Model 2, teachers' reported opportunities to learn were included; only teacher participation in additional mathematics coursework was significantly associated with the outcome. In Model 3, teachers' self-concept and reported use of instructional practices that required students to explain, analyze, and prove were entered. Both were significant, although the model shows that teachers with higher MKT asked students to do less—not more—explanation, analysis, and proof. Finally, in Model 4 we see teachers' reports of serving in leadership or mentoring roles were unrelated to MKT once other controls were entered.

In the first model, background characteristics accounted for roughly 10% of the variance in MKT in this sample. The addition of other predictors increased the explained variation, but not by much; in the final model, the full set explained only

Regression of Teacher M.	KI On Teacher C		a Activities	
	Model 1	Model 2	Model 3	Model 4
Intercept	0.59****	-1.12****	-1.76****	-1.76****
	(0.12)	(0.29)	(0.35)	(0.36)
Grade taught	0.18****	0.18****	0.17****	0.18****
	(0.02)	(0.03)	(0.02)	(0.03)
Years of	0.01*	0.01*	0.01*	0.01*
experience	(0.004)	(0.004)	(0.004)	(0.004)
School free and	-0.26	-0.29	-0.29	-0.28
reduced lunch	(0.14)	(0.15)	(0.15)	(0.15)
Mathematics course		0.12*	0.09	0.09
		(0.06)	(0.05)	(0.06)
Mathematics methods		0.01	0.02	0.01
course		(0.07)	(0.06)	(0.07)
Mathematics profes-		0.00	-0.02	-0.02
sional development		(0.03)	(0.04)	(0.04)
Other		0.03	0.04	0.04
Development		(0.17)	(0.04)	(0.04)
Teacher education		0.11	0.03	0.03
Program		(0.17)	(0.17)	(0.17)
Instructional practices			-0.08*	0.07*
			(0.03)	(0.04)
Mathematics			0.31****	0.32****
self-concept			(0.06)	(0.06)
Service on a math				0.05
committee				(0.11)
Peer coach or mentor				-0.01
				(0.13)
Taught in-service	0.10			0.01
				(0.17)
Adjusted r-squared		0.10	0.14	0.14
n	560	541	537	537

 Table 8

 Repression of Teacher MKT on Teacher Characteristics and Activities

 $\it Note.$  Standard errors in parentheses. Teachers without mathematics responsibilities excluded from the analysis.

\* p < .05. \*\* p < .01. \*\*\* p < .001. \*\*\*\* p < .0001.

14% of variation in MKT. This, and the relatively weak regression coefficients for many predictors, suggests that those making decisions about hiring, awarding tenure, and promoting teachers to leadership positions are provided little information by candidates' formal qualifications, self-reported teaching, and expertise in the content area.

## Methodological Issues

The correlation between teachers' mathematical self-concept and their MKT has implications for the measurement of teachers' knowledge and, in particular, the measurement of gains in such knowledge as a result of professional education efforts. We discuss this in more detail in the conclusion to this article.

However, one more methodological finding stands out: that this nationally representative sample performs differently from convenience samples used initially to pilot items for this survey. A sample of teachers who took the assessment as part of teacher professional development programs in 2001, for instance, outscored the sample here by roughly 3.6% over a group of identical items, with wide variation (up to 10%) for a handful of items. A sample of teachers contacted by mail in 2006 using a list supplied by a commercial publisher outscored this nationally representative sample by 5.7%, also with wide variation in item-specific differences. This survey featured no sample verification or follow-ups with teachers and thus had a lower response rate of 30%. This suggests that the more selective the sample, the better overall performance will be on assessments of this type. In the case of this survey, the response rate of 59% may make it vulnerable to similar score inflation—for instance, if more knowledgeable and more confident teachers were more likely to respond by a significant margin. However, the differences are, on average, not large.

# CONCLUSION

This study has clear limitations. We assessed teacher knowledge in one area, number and operations at the elementary level, rather than a more extended set of mathematical topics and grade levels. Also, we do not know the extent to which teachers who responded to this survey consulted resources, such as the Internet or curriculum materials, in determining their answers. Despite these limitations, however, this study contains several lessons for teacher educators and policymakers.

Inspecting item parameters, we see that MKT domain, rather than mathematical content, is associated with differences in item difficulties. That is, difficulties were unaffected by whether the item covered whole or rational number; this is a surprise in light of the conventional wisdom that elementary teachers find rational numbers particularly difficult. We also found that items labeled as common content knowledge proved easier for this sample than items categorized as belonging to the specialized and/or pedagogical content knowledge categories. Among the latter group, we found that more subtle mathematics judgments resulted in greater item difficulty. We also found that items requiring mathematical explanation were

markedly more difficult than items involving representation of content and interpreting nonstandard student work (see also Leinhardt, 2001). This analysis excluded items that did not relate well to the underlying construct, rendering it likely that these differences reflect true variability in the types of tasks that are easy and difficult for teachers. With only a small number of items, replication of this study is important. However, we note that the pattern identified here parallels what we found in a nationally representative sample of middle school teachers (Hill, 2007).

These patterns, as well as the description of specific easy and difficult items, lead to implications for the professional education of teachers and the measurement of professional knowledge. To start, we were able to identify areas of mathematics that proved difficult for teachers within the core K–6 curricula. This suggests that it is not necessary to use advanced mathematics topics, such as proportional reasoning or precalculus, to develop an instrument with adequate measurement properties for elementary teachers.

These results suggest that professional education efforts might focus on the specialized and pedagogical content knowledge teachers might use in the course of their work. Although the assessment carried a limited number of common content knowledge items, these proved easier—a pattern we also saw in our middle school data as well (Hill, 2007). This recommendation, however, poses a problem for professional education efforts. Specialized knowledge has yet to be fully mapped (Ball et al., 2008); without such mapping, addressing these topics coherently in professional education efforts will be difficult. Further, the topic of difficult items varied widely, from long division to integer operations. It is likely, to our minds, that there is more such knowledge than could be reasonably taught in a course or several-week professional development setting. Instead, mathematics educators will have to implement strategies that enable teachers to learn this content in their workplace from more experienced colleagues and/or curriculum materials (Ball & Cohen, 1996; Davis & Krajcik, 2005).

These results also hold implications for public policies aimed at exactly this form of professional education. We found that individuals who were active in professional leadership activities such as mathematics committee work, peer mathematics mentoring, and teaching mathematics in-services had scant advantage on the assessment of their MKT. The same held true for individuals who reported attending a generous amount of mathematics-specific professional development. These findings highlight two issues. First, peer math coaches and mentor teachers as well as mathematics in-service instructors already have direct responsibility for conveying content to teachers. Although such mentoring and professional development may have pedagogical benefits (e.g., suggesting good activities, noting how to navigate curriculum materials, dealing with more generic pedagogical issues), we are concerned that these activities may, in general, lack content focus and strength. Second, if full-time mathematics coaches are drawn from this set of individuals already active in mathematics leadership without regard for those prospective coaches' level of content knowledge, this popular policy intervention may lack promise. These findings accord with and help explain a recent study by Murray, Ma, and Mazur (2009). They observed a peer coaching effort among high school teachers and found that most coach-teacher interactions were short, overwhelmingly positive, and lacking in analytic substance.

That teachers cannot provide accurate descriptors of their own level of mathematical knowledge is not surprising but still discouraging. With practice largely privatized, few see others' instruction or have opportunities to engage in in-depth mathematical conversations. Further, there are few opportunities for teachers to receive feedback on their mathematical strengths and weaknesses. Students, who might form one potent source of feedback, do not hold mathematical authority within classrooms. Other adults who view instruction—including administrators and math coaches—do not necessarily comment on the mathematics in any depth (Murray et al., 2009). Finally, mathematical knowledge is relative. A teacher may see herself as the most mathematically knowledgeable second-grade teacher in her school and may be correct; however, she may score well below the national average.

The correlation of 0.25 between teachers' mathematical self-concept and MKT has methodological, as well as policy, implications. One implication is that studies using teacher self-reports of mathematics knowledge or mathematics knowledge gain are estimating this construct with appreciable error. Another is that current policy efforts premised on teachers self-identifying as mathematically weak or strong—for example, mathematics programs that aim to recruit less able teachers or mathematics and science magnet schools that need to recruit mathematically talented teachers—will face significant barriers.

The few and relatively mild significant associations found in this study suggest avenues for further research. Studies that specifically focus on the content knowledge of mathematics leaders and alternatively certified teachers are urgently needed; the information presented here and elsewhere suggests that the premise for these policies—that more mathematically skilled individuals can be drawn into the profession and that teachers will learn mathematics from others—are suspect. A second avenue includes overcoming problems involved in measuring teachers' content knowledge directly. If we are to replace teacher self-report as the arbiter of program or professional development success, evaluators need tools and procedures for objectively measuring different forms of mathematical knowledge. Finally, our descriptions of easy and difficult items conflict with the case study literature, in that this literature noted general deficiencies in knowledge of rational numbers and our analysis did not find the same. Investigating this issue further will help identify topics and mathematical tasks of teaching on which professional education should focus.

#### REFERENCES

An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school, mathematics teachers in China and the U.S. Journal of Mathematics Teacher Education, 7, 145–172.

- Bacolod, M. P. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, 89, 737-751.
- Ball, D. L. (1990a). Prospective elementary and secondary teachers' understanding of division. Journal for Research in Mathematics Education, 21, 132–144.
- Ball, D. L. (1990b). Reflections and deflections of policy: The case of Carol Turner. Educational Evaluation and Policy Analysis, 12, 247–259.
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6–8, 14.
- 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., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47, 133–180.
- Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, 23, 194–222.
- Boyd, D., Lankford, H., Loeb, S., & Wycoff, J. (2003). Understanding teacher labor markets: Implications for educational equity. In M. L. Plecki & D. H. Monk (Eds.), School finance and teacher quality: Exploring the connections. 2003 yearbook of the American Education Finance Association (pp. 55-84). Larchmont, NY: Eye on Education.
- Charalambous, C. Y. (2010). Mathematical knowledge for teaching and task unfolding: An exploratory study. *The Elementary School Journal*, 110, 247–278.
- Clements, D. H., Swaminathan, S., Zeitler Hannibal, M. A., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education*, 30, 192–212.
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. Educational Evaluation and Policy Analysis, 12, 311–329.
- Cohen-Vogel, L., & Smith, T. M. (2007). Qualification and assignments of alternatively certified teachers: Testing core assumptions. *American Educational Research Journal*, 44, 732–753.
- Corcoran, S. P. (2007). Long-run trends in the quality of teachers: Evidence and implications for policy. *Education Finance and Policy*, 2, 395–407.
- Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, 34(3), 3–14.
- Educational Testing Service. (2010a). PRAXIS series passing scores by test and state. Princeton, NJ: Author.
- Educational Testing Service. (2010b). For test takers: Praxis II test content and structure. Retrieved July 8, 2010, from http://www.ets.org/praxis/about/praxisii/content
- Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics (pp. 147–164). New York: Macmillan.
- Fennema, E., Franke, M. L., Carpenter, T. P., & Carey, D. A. (1993). Using children's mathematical knowledge in instruction. *American Educational Research Journal*, 30, 555–583.
- Garet, M. S., Porter, A. C., & Desimone, L. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915–945.
- Greenberg, J., & Walsh, K. (2008). No common denominator: The preparation of elementary teachers in mathematics by America's education schools. Washington, DC: National Council on Teacher Quality.
- Harris, D. N., & Sass, T. R. (2007). Teacher training, teacher quality, and student achievement (Working Paper No. 3). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

- Heaton, R. M. (1992). Who is minding the mathematics content? A case study of a fifth-grade teacher. *The Elementary School Journal*, 93, 153–162.
- Hess, F. M. (2002). Tear down this wall: The case for a radical overhaul of teacher certification. Educational Horizons, 80, 169-183.
- 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*, 95–114.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge. Journal for Research in Mathematics Education, 39, 372–400.
- Hill, H. C., Blunk, M. L., Charalambous, Y. C., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, 26, 430–511.
- Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research and Perspectives*, 5, 2-3, 81-92.
- Hill, H. C., & Lubienski, S. T. (2007). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. *Educational Policy*, 21, 747–768.
- 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., Umland, K., & Litke, E. (2010, May). Teacher quality and quality teaching: Examining the relationship of a teacher exam to practice. Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.
- Hubbard, F. (2008). Summary of procedure for calculation of sample weights. Ann Arbor, MI: Institute for Social Research.
- Knight, J. (2009). Coaching. Journal of Staff Development, 30(1), 18-20, 22.
- Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook for research on teaching* (4th ed., pp. 333–357). Washington, DC: American Educational Research Association.
- Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. Journal of Educational Psychology, 77, 247–271.
- Lloyd, G. M., & Wilson, M. R. (1998). Supporting innovation: The impact of a teacher's conceptions of functions on his implementation of a reform curriculum. *Journal for Research in Mathematics Education*, 29, 248–274.
- Loeb, S., & Reininger, M. (2004). Public policy and teacher labor markets: What we know and why it matters. East Lansing, MI: Education Policy Center at Michigan State University.
- Ma, L. (1999). Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States. Mahwah, NJ: Erlbaum.
- Massachusetts Department of Elementary and Secondary Education & Pearson Education, Inc. (2008). Massachusetts tests for educator licensure (MTEL). General curriculum (03) practice test: Booklet 2. Mathematics subtest. Amherst, MA: Pearson.
- Murray, S., Ma, X., & Mazur, J. (2009). Effects of peer coaching on teachers' collaborative interactions and students' mathematics achievement. *The Journal of Educational Research*, 102, 203–212.
- National Center for Education Statistics. (n.d.). Common core database (CCD) 2005-2006, Retrieved from http://nces.ed.gov/ccd
- National Center for Education Statistics (2010). Schools and staffing survey. Retrieved July 12, 2010, from http://nces.ed.gov/surveys/sass
- National Commission on Mathematics and Science Teaching for the 21st Century. (2000). Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century. Washington, DC: U.S. Department of Education.
- National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.

- National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. Retrieved June 1, 2008, from http://www2.ed.gov/about/bdscomm/ list/mathpanel/report/final-report.pdf
- National Research Council. (2001). Adding it up: Helping children learn mathematics (J. Kilpatrick, J. Swafford, & B. Findell, Eds.). Washington, DC: National Academy Press.
- Newton, K. J. (2008). An extensive analysis of preservice elementary teachers' knowledge of fractions. American Educational Research Journal, 45, 1080–1110.
- Newton, K. J. (2009). Instructional practices related to prospective elementary school teachers' motivation for fractions. *Journal of Mathematics Teacher Education*, 12, 89–109.
- Paige, R. (2002). Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality. Washington, DC: U.S. Department of Education, Office of Postsecondary Education.
- Peterson, P. E., & Nadler, D. (2009). What happens when states have genuine alternative certification? We get more minority teachers and test scores rise. *Education Next*, 9(1), 70–74.
- Putnam, R. T., Heaton, R. M., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *The Elementary School Journal*, 93, 213–228.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2008). NBER Working Paper Series: No. 14485. Can you recognize an effective teacher when you recruit one? (JEL No. 121, J45). Washington, DC: National Bureau of Economic Research.
- Rowan, B., Harrison, D., & Hayes, A. (2004). Using instructional logs to study elementary school mathematics: A close look at curriculum and teaching in the early grades. *The Elementary School Journal*, 105, 103–127.
- Schilling, S. G. (2005). ORDFAC: A computer program for ordinal factor analysis [Computer software]. Ann Arbor, MI: University of Michigan.
- Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership*, 53(6), 12–16.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Sowder, J. T., Philipp, R. A., Armstrong, B. E., & Schappelle, B. P. (1998). Middle-grade teachers' mathematical knowledge and its relationship to instruction. Albany, NY: State University of New York Press.
- Stein, M. K., Baxter, J. A., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A case from functions and graphing. *American Educational Research Journal*, 27, 639–663.
- Swafford, J. O., Jones, G. A., & Thornton, C. A. (1997). Increased knowledge in geometry and instructional practice. Journal for Research in Mathematics Education, 28, 467–483.
- Thompson, P. W., & Thompson, A. G. (1994). Talking about rates conceptually, part I: A teacher's struggle. Journal for Research in Mathematics Education, 25, 279–303.
- Tirosh, D. (2000). Enhancing prospective teachers' knowledge of children's conceptions: The case of division of fractions. *Journal for Research in Mathematics Education*, 31, 5–25.
- Tournaki, N., Lyublinskaya, I., & Carolan, B. V. (2009). Pathways to teacher certification: Does it really matter when it comes to efficacy and effectiveness? Action in Teacher Education, 30(4), 96–109.
- U.S. Department of Education. (2009). The American Recovery and Reinvestment Act of 2009: Saving and creating jobs and reforming education. Retrieved September 29, 2009, from http://ed.gov/ policy/gen/leg/recovery/implementation.html
- Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). Report of the 2000 National Survey of Science and Mathematics Education. Chapel Hill, NC: Horizon Research.
- Wilson, S. M. (1990). A conflict of interests: The case of Mark Black. Educational Evaluation and Policy Analysis, 12, 293–310.

- Wilson, S. M., Shulman, L. S., & Richert, A. (1987). '150 different ways' of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 104–124). London: Cassell Educational Limited.
- Zimowski, M., Muraki, E., Mislevy, R., & Bock, D. (2003). BILOG-MG 3 (Version 3) [Computer software]. Lincolnwood, IL: Scientific Software International.

## Author

Heather C. Hill, Harvard Graduate School of Education, 6 Appian Way #445, Cambridge, MA 02138; heather\_hill@harvard.edu

Accepted June 8, 2010

# APPENDIX A

1. Ms. Dominguez was working with a new textbook, and she noticed that it gave more attention to the number 0 than her old book. She came across a page that asked students to determine if a few statements about 0 were true or false. Intrigued, she showed them to her sister, who is also a teacher, and asked her what she thought.

Which statement(s) should the sisters select as being true? (Mark YES, NO, or I'M NOT SURE for each item below.)

	Yes	No	I'm not sure
a) 0 is an even number.	1	2	3
b) 0 is not really a number. It is a placeholder in writ- ing big numbers.	1	2	3
c) The number 8 can be written as 008.	1	2	3

2. At a professional development workshop, teachers were learning about different ways to represent multiplication of fractions problems. The leader also helped them to become aware of examples that do not represent multiplication of fractions appropriately.

Which model below <u>cannot</u> be used to show that  $1\frac{1}{2} \times \frac{2}{3} = 1$ , ? (Mark ONE answer.)



3. Ms. Abdul is preparing a unit to introduce her students to proportional reasoning. She is considering three versions of a problem that are the same except for the numbers used. Which version of the Mr. Short and Mr. Tall problem below is likely to be the most <u>challenging</u> for students? (Circle ONE answer.)

- a) A picture depicts Mr. Short's height as 4 paper clips and as 6 buttons. The height of Mr. Tall (not shown) is given as 6 paper clips. How many buttons in height is Mr. Tall?
- b) A picture depicts Mr. Short's height as 4 paper clips and as 7 buttons. The height of Mr. Tall (not shown) is given as 5 paper clips. How many buttons in height is Mr. Tall?
- c) A picture depicts Mr. Short's height as 2 paper clips and as 9 buttons. The height of Mr. Tall (not shown) is given as 5 paper clips. How many buttons in height is Mr. Tall?
- d) All three of the problems are equally challenging.
- 4. Ms. Miller wants her students to write or find a definition for triangle and then improve their definition by testing it on different shapes. To help them, she wants to give them some shapes they can use to test their definition.

She goes to the store to look for a visual aid to help with this lesson. Which of the following is most likely to help students improve their definitions? (Circle ONE answer.)



				Excluding	g MK I				
							Reform		Free/
					Math	Math	instruc-		reduced
	Years of	Grade	Math	Other	content	methods	tional	Math self-	lunch
	experience	taught	prof'l dev.	prof'l dev.	course	course	methods	concept	eligibility
Years of	1.00	-0.01	02	-0.05	-0.01	0.12	0.03	0.06	-0.11
experience		(260)	579)	(573)	(574)	(574)	(567)	(580)	(582)
Grade taught		1.00	0.05	-0.02	0.03	0.02	0.15	0.09	-0.02
			(586)	(579)	(581)	(579)	(584)	(589)	(200)
Math prof'l			1.00	0.29	0.15	0.13	0.12	0.13	0.11
development				(605)	(605)	(603)	(594)	(809)	(613)
Other prof'l				1.00	0.05	0.11	0.05	-0.02	0.03
development					(665)	(597)	(587)	(601)	(909)
Math content					1.00	0.34	0.10	0.17	0.04
course						(603)	(201)	(603)	(609)
Math						1.00	0.09	0.10	-0.03
methods							(589)	(601)	(909)
Explanation,							1.00	0.15	0.08
analysis, proof								(598)	(598)
Math self-									-0.01
concept									(612)

Appendix B Correlations Between Teacher Characteristics, Excluding MKT