Developing Measures of Teachers' Mathematics Knowledge for Teaching

Heather C. Hill
Stephen G. Schilling
Deborah Loewenberg Ball
University of Michigan

The Elementary School Journal
Volume 105, Number 1

0013-5984/2004/10501-0002$05.00

Abstract

In this article we discuss efforts to design and empirically test measures of teachers' content knowledge for teaching elementary mathematics. We begin by reviewing the literature on teacher knowledge, noting how scholars have organized such knowledge. Next we describe survey items we wrote to represent knowledge for teaching mathematics and results from factor analysis and scaling work with these items. We found that teachers' knowledge for teaching elementary mathematics was multidimensional and included knowledge of various mathematical topics (e.g., number and operations, algebra) and domains (e.g., knowledge of content, knowledge of students and content). The constructs indicated by factor analysis formed psychometrically acceptable scales.

In the past 2 decades teachers' knowledge of mathematics has become an object of concern. New theoretical and empirical insights into the work of teaching (e.g., Shulman, 1986, 1987; Wilson, Shulman, & Richert, 1987) have spurred greater attention to the role such knowledge plays in teacher education and in the quality of teaching itself (e.g., National Commission on Teaching and America's Future, 1996). Other studies have documented the mean and variation in teachers' knowledge of mathematics for teaching (e.g., Ball, 1990; Ma, 1999). Results of these efforts have been reflected in teaching standards published by Interstate New Teacher Assessment and Support Consortium (INTASC), the National Board for Professional Teaching Standards, as well as by many other states, localities, and professional teaching organizations (e.g., the National Council of Teachers of Mathematics). Concerns that teachers possess necessary knowledge and skills for teaching mathematics have also led to the development and use of teacher licensing exams, such as PRAXIS, an assessment developed by the Educational Testing Service and now administered in 38 states. Other states and testing firms have developed and administer similar assessments.

Given the development of such standards and assessments, one might conjecture that there is substantial agreement about the knowledge needed for teaching children mathematics. However, a closer look at released items from the elementary mathematics portion of these teacher licensure exams suggests lack of agreement over what teachers need to know. Some exams assess individuals' ability to solve middle-school-level mathematics problems (e.g., California Basic Educational Skills Test), others the ability to construct mathematical questions and tasks for students (e.g., Exam for the Certification of Educators in Texas), and still others the ability to understand and apply mathematics content to teaching (e.g., Massachusetts Tests for Educator Licensure). This implicit disagreement over the knowledge of mathematics that teachers need can be traced through the theoretical and empirical literature on teaching knowledge, where different authors have proposed divergent elements and organizations for such knowledge. The disagreement is reflected in current debates about the mathematics teachers need to know in order to teach. Some argue, for instance, that a teacher's capability in general mathematics is the most important qualification (U.S. Department of Education, 2002). Others believe that general mathematical ability must be complemented by additional professional knowledge, for example, of student thinking about content, or of mathematical tasks specific to the work of teaching. To date, however, few empirical data have been publicly available to help judge the validity of either claim.

We sought to shed light on this debate by analyzing data collected in the service of constructing an assessment of teachers' content knowledge for teaching mathematics. To develop this assessment, we used elements from existing theories about teacher knowledge (e.g., Ball & Bass, 2003; Grossman, 1990; Shulman, 1987; Wilson et al., 1987) to write a set of survey-based teaching problems thought to represent components of the knowledge of mathematics needed for teaching. We then factor analyzed teachers' responses to this item set to determine the structure of the knowledge we tried to represent. The principal question guiding our work was: Is there one construct that can be called "mathematics knowledge for teaching" and that explains patterns of teachers' responses, or do these items represent multiple constructs and thus several distinct mathematical competencies of elementary mathematics teachers? A second question was: Given the structure of teachers' mathematical knowledge for teaching, can we construct scales that measure such knowledge reliably?

In this article we describe this effort and its results. We begin with an overview of the original literature about content knowledge for teaching. Next we discuss our own efforts to write items that represented such knowledge, with an emphasis on the potential constructs that might emerge from the items. Finally, we describe initial results from a field test of these items, including factor analyses and attempts to scale the items for use in statistical work.

Literature Review

In the mid-1980s Lee Shulman and his colleagues introduced the notion of "pedagogical content knowledge" to refer to the special nature of the subject-matter knowledge required for teaching (Shulman, 1986, 1987; Wilson et al., 1987). Conceived as complementary to general pedagogical knowledge and general knowledge of subject matter, the concept of pedagogical content knowledge was thought to include familiarity with topics children find interesting or difficult, the representations most useful for teaching an idea, and learners' typical errors and misconceptions. Labeling this as pedagogical content knowledge not only underscored the importance of understanding subject matter in teaching but also suggested that personal knowledge of the subject—that is, what an educated adult would know of a subject—was insufficient for teaching that subject. This distinction represented an important contribution to solving the puzzle about qualities and resources needed for effective teaching.

Research on pedagogical content knowledge has conceived of such knowledge as particular, rooted in the details of school subject matter and of what is involved in helping others understand it. Working in depth in different subjects, scholars probed the nature of the content knowledge entailed by teaching; comparisons across fields were also generative. Grossman (1990), for example, articulated how teachers' orientations to literature shaped the ways in which they approached particular texts with their students. And Wilson and Wineburg (1988) illuminated how social studies teachers' disciplinary backgrounds affected the ways in which they represented historical knowledge for high school students. In mathematics, scholars showed that what teachers would need to understand about fractions, place value, or slope, for instance, would be substantially different from what would suffice for other adults (Ball, 1988, 1990, 1991; Borko et al., 1992; Leinhardt & Smith, 1985).

Despite this wealth of research, we argue that the mathematical content teachers must know in order to teach has yet to be mapped precisely. Most foundational work in mathematics has relied principally on single-teacher case studies, expert-novice comparisons, cross-national comparisons, and studies of new teachers. Although such methods have been critical in beginning to articulate the content of subject-matter knowledge for teaching mathematics, these methods lack the power to propose and test hypotheses regarding the organization, composition, and characteristics of content knowledge for teaching.

Researchers have, however, conjectured about the organization of such knowledge, and these conjectures proved useful starting points for this investigation. Shulman (1986) originally proposed three categories of subject-matter knowledge for teaching. His first category, content knowledge, "refers to the amount and organization of knowledge per se in the mind of teachers" (p. 9). Content knowledge, according to Shulman, included both facts and concepts in a domain but also why facts and concepts are true and how knowledge is generated and structured in the discipline (Bruner, 1960; Schwab, 1961/1978). The second category Shulman and his colleagues (Shulman, 1986; Wilson et al., 1987) advanced, pedagogical content knowledge, "goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching" (p. 9). This category has become of central interest to researchers and teacher educators alike. Included here are representations of content ideas as well as an understanding of what makes learning a topic difficult or easy for students. Shulman's third category of subject-matter knowledge for teaching, curriculum knowledge, involves awareness of how topics are arranged both within a school year and over longer periods of time and ways of using curriculum resources, such as textbooks, to organize a program of study for students.

Shulman's theory of teacher knowledge listed also general pedagogical knowledge (classroom management techniques and strategies), knowledge of learners and their characteristics, knowledge of educational contexts (e.g., school board politics, communities), and knowledge of educational ends, purposes, and values.

Leinhardt and Smith (1985) proposed a different organization of teacher knowledge in their study of expert-novice differences in mathematics teaching. Working from a psychological/cognitive perspective, they identified two aspects of knowledge for teaching: lesson structure knowledge—which includes planning and running a lesson smoothly and providing clear explanations—and subject-matter knowledge. They included in the latter "concepts, algorithmic operations, the connections among different algorithmic procedures, the subset of the number systems being drawn upon, the understanding of classes of student errors, and curriculum presentation" (p. 247).

Other ways of dividing the terrain have been advanced as well. Grossman (1990) reorganized Shulman and colleagues' categories into four and extended them slightly: subject-matter knowledge, general pedagogical knowledge, pedagogical content knowledge (knowledge of students' understanding, curriculum, and instructional strategies), and knowledge of context. Ball (1990) described differences between teachers' ability to execute an operation (division by a fraction) and their ability to represent that operation accurately for students, clearly demarcating two dimensions in teachers' content knowledge—the ability to calculate a division involving fractions, and the kind of understanding of that operation needed for teaching. And, based on analyses of classroom lessons, Ball proposed a distinction between knowledge of mathematics and knowledge about mathematics, corresponding roughly to knowledge of concepts, ideas, and procedures and how they work, on one hand, and knowledge about "doing mathematics"—for example, how one decides that a claim is true, a solution complete, or a representation accurate—on the other hand. In more recent work, Ma (1999) used comparisons of U.S. and Chinese elementary teachers to describe "profound understanding of fundamental mathematics" as instantiated in the connectedness, multiple perspectives, basic (fundamental) ideas, and longitudinal coherence that occurred during their teaching.

By posing these potential categories of content knowledge for teaching, researchers have contributed to the development of theory about this knowledge. One contribution has been to refocus researchers' attention on the centrality of subject matter and subject-matter knowledge in teaching. A second was to draw attention back to disciplines and their structures as a basis for theorizing about what teachers should know. A third has been to focus attention on what expert teachers know about content and how they use or report using this knowledge of subject matter in their teaching.

Although the theoretical work on teachers' knowledge for teaching has contributed to the field, much remains to be done. For example, there is still much to be understood about the organization and structure of subject-matter knowledge in different disciplines and what these structures suggest for teaching. Little is known yet about whether and how content knowledge for teaching relates to the content knowledge of other professionals or of ordinary educated adults. And to date, scholars have not attempted to measure teachers' knowledge for teaching in a rigorous manner and thus cannot track its development or contribution to student achievement.

Method

To learn more about these issues, we began in 2001 to write, and later pilot test, numerous multiple-choice items intended to represent the mathematical knowledge used in teaching elementary mathematics. We wrote items in categories that were modified versions of those proposed by Shulman, Wilson, Ball, and others. The uneven research base across domains was noticeable, leading us to mark off the territory in slightly different ways. For instance, although we found substantial research on student errors and strategies in mathematics, we uncovered less about what representations work best for particular mathematical topics. Such insight would be needed to write items that would tap knowledge of "representations most useful for teaching specific content" (Shulman, 1986). Without these items and this category, faithfully representing pedagogical content knowledge as conceived in the theoretical literature was not possible. The more extensive research base about student learning of mathematics led us to craft a separate category on this topic. We explain our categories in more detail below.

Item writing served several purposes: at the most practical level, we hoped to develop measures by which we could gauge growth in teachers' content knowledge for teaching and learn more about how such knowledge contributes to student achievement. Item writing was also another way to explore the nature and composition of subject-matter knowledge for teaching. During the process of examining curriculum materials and student work, writing and refining items, and thinking about what items represented, we sharpened and defined our ideas about the mathematical knowledge and skill needed for teaching mathematics. Finally, pilot testing these items allowed us to use factor analyses and scaling techniques to learn about the organization and characteristics of mathematical knowledge for teaching. Before describing the results of our analyses and efforts to build scales, we recount the process by which we developed survey items and describe the possible ways these items might be categorized.

Developing Survey Items

Our approach to studying content knowledge for teaching was grounded in a theory of instruction, taking as a starting point the work of enacting high-quality instruction (Ball & Bass, 2000; Ball & Cohen, 1999; Cohen & Ball, 1999). From that perspective, we asked, What mathematical knowledge is needed to help students learn mathematics? Our interest was in identifying what and how subject-matter knowledge is required for teaching. Using this theoretical perspective as well as the research base on teaching and learning mathematics, analyses of curriculum materials, examples of student work, and personal experience, researchers at the Study of Instructional Improvement developed 138 mathematics items in the spring of 2001 in the categories shown in Figure 1. Two content areas—number concepts and operations—were selected because they comprise a significant portion of the K–6 curriculum and because important and useful work existed on the teaching and learning of these topics. Patterns, functions, and algebra was chosen because it represents a newer strand of the K–6 curriculum and thus allows insight into what and how teachers know about this topic now and perhaps how knowledge increases over time as better curriculum and professional development become available and as teachers gain experience teaching this topic. Initial item-writing efforts also focused on two kinds of teacher knowledge: knowledge of content itself and combined knowledge of students and content. By demarcating the domains in this way, writers intended to reflect elements contained in Shulman and others' typologies of content knowledge for teaching. Crossing the three content areas with the two domains of teacher knowledge yielded six cells. However, the lack of research on students' learning of patterns, functions, and algebra in 2001 led us to conclude that we could not develop items in this cell during these initial item-writing efforts (see Fig. 1). Since then, more research in this area has become available (e.g., Carpenter, Franke, & Levi, 2003) and item writing is planned.

The constructs, or underlying organizing principles, indicated by factor analyses with these items might reflect the five existing domains exactly. Yet a post-hoc analysis of the items revealed other potential hypotheses about the organizational structure. To start, one construct may explain the patterns in teachers' responses to items in all five cells. If this were so, we might conjecture that this single construct could be described as "general mathematical ability" and conclude that there is little need to identify specialized knowledge for teach-

	Domain	
Content area	Knowledge of content	Knowledge of students and content
Number concepts		
Operations		
Patterns, functions, algebra		

FIG. 1.—Mathematics content areas and domains. Shaded area represents construct for which no items were developed.

ing, or that this specialized knowledge is so strongly related to the knowledge other educated adults hold as to be functionally equivalent, at least for measurement purposes. At the other end of the spectrum, however, we might find that items are differentiated in more detail than in Figure 1. For instance, teachers' knowledge might be differentiated at the level of particular topics in the elementary curriculum, for example, whole numbers, fractions, decimals, operations (e.g., addition) with whole numbers, and so on. If this were the case, we might conjecture that teachers have highly particularized knowledge of the material they teach, and we would study these knowledge clusters in more depth.

Mathematical content areas are not the only potential means of organizing items. These items were situated in yet another possible categorization system, what we would call tasks of teaching. This way of categorizing items is based on the idea that teachers' mathematical knowledge is used in the course of different sorts of tasks—choosing representations, explaining, interpreting student responses, assessing student understanding, analyzing student difficulties, evaluating the correctness and adequacy of curriculum materials. These are tasks teachers might face in teaching any subject, and they provide another potential way to organize teachers' content knowledge for teaching mathematics.

Finally, items may differentiate themselves within the cells shown in Figure 1. For instance, some items appear to require respondents to draw on common knowledge of content (CKC), for instance, items that ask teachers to find the decimal halfway between 1.1 and 1.11 or to find the eighty-third shape in a sequence. As Shulman and others have pointed out, such mathematics knowledge is used in the course of teaching because teachers must compute, make correct mathematical statements, and solve problems. Other items, however, appear to be based on the ways mathematics arises in elementary classrooms, or on what we call specialized knowledge of content (SKC), including building or examining alternative representations, providing explanations, and evaluating unconventional student methods. One way to illustrate this distinction is by imagining how someone who has not taught children but who is otherwise knowledgeable in mathematics might interpret and respond to these items. This test population would not find the items that tap ordinary subject-matter knowledge difficult. By contrast, however, these mathematics experts might be surprised, slowed, or even halted by the mathematics-as-used-in-teaching items; they would not have had access to or experience with opportunities to see, learn about, and understand mathematics as it is used at the elementary level.

Two of the four mathematics items in the Appendix illustrate this distinction. In the first item, about powers of ten, teachers must draw on their knowledge of properties of numbers—in this case, place value as represented within exponential notation—to answer the problem. This content knowledge is used in teaching; students learn about exponential notation in the middle to late elementary grades, and thus teachers must have adequate knowledge to develop this topic. However, many adults, and certainly all mathematicians, would know enough to answer this item correctly. The next item illustrates a special kind of content knowledge, one that arises through teaching content to young children. In the second item, teachers must inspect three solutions to the same two-digit multiplication problem (35 × 25) and assess whether approaches used for each solution would generalize to all whole-number multiplication. To respond in such situations, teachers must draw on their mathematics knowledge—inspecting the solution to understand what was done at each step, then gauging whether the method makes sense and would work in all cases. Analyzing procedures and justifying their validity is a mathematical process. However, doing it in this way and in this context (i.e., appraising student solutions to a computation problem) is a task that arises regularly in teaching and not necessarily in other arenas. Hence, it is a type of and context for mathematical reasoning in which teachers must engage, and it appears to draw on a specialized type of mathematical knowledge, one that makes it possible to analyze and make sense of a range of methods and approaches to a computation.

The KSC category also contained subtle distinctions. As we (and mathematicians associated with this project) reviewed items, we saw that some such items required knowledge of students and their ways of thinking about mathematics—typical errors, reasons for those errors, developmental sequences, strategies for solving problems. Teachers may need to know, for instance, what errors students make as they learn about the place value system, or strategies students might use to remember the answer to 8 × 9. In other cases, knowledge of students and content items might draw on student thinking and/or mathematics content knowledge. For instance, teachers may use both types of knowledge to interpret student statements about the commutative property, analyzing what students have said about this topic to assess understanding and depth of knowledge.

The third and fourth problems in the Appendix illustrate distinctions within our KSC category. The third item asks teachers to consider which of three lists of decimal numbers would be best to assign to discriminate students' understanding of and skill in ordering decimals, or whether the three lists would be equally useful for this purpose. Two of the three lists would allow students to respond correctly without paying any attention to the decimal point. In interviews, many people who have never taught this topic, including mathematicians, report seeing no difference among the three lists; teachers with knowledge of decimals for teaching are more likely to see the differences immediately. Thus, knowledge of students and of the typical mistakes they make in ordering decimals is necessary to answer this problem correctly. The next item, on buggy algorithms, requires either knowledge of typical student mistakes or the ability to perform a detailed mathematical analysis to arrive at a correct answer.

Data Collection and Analysis

Items in each cell of Figure 1 were assigned roughly equally to one of three forms (A, B, or C), thus balancing forms across both content and domains. Forms were also balanced in terms of projected item difficulties. These forms were pilot tested in California's Mathematics Professional Development Institutes (MPDIs). These institutes were publicly funded, large-scale efforts to boost California teachers' knowledge of subject matter in mathematics. The institutes had over 40 sites, cost about $65 million, and served 23,000 K–12 teachers in the first 3 years of the program. Pilot testing took place only with elementary teachers enrolled in number and operations institutes. At a typical number and operations institute, teachers were paid up to $1,500 to attend summer sessions ranging from 1 to 3 weeks. Academic mathematicians and mathematics educators were the instructors; the content was mathematics—number and operations. Although MPDI sites were selected on the basis of their willingness to work with low-performing/high-poverty districts and schools, teachers were not recruited on the basis of their mathematical knowledge or other characteristics (Madfes, Montell, & Rosen, 2002). As a condition of funding, each institute was required to measure growth in teacher content knowledge. By supplying these measures, the Study of Instructional Improvement and officials at California's MPDIs formed a mutually beneficial partnership, allowing both the pilot testing of items and, potentially, an evaluation of the institutes' effectiveness. Items and forms, however, were not written or constructed to align with any particular MPDI, because content varied across the 21 institutes included in this analysis, and because we wanted to design measures that could be used beyond the MPDI setting.

By combining the summer pre/postassessments given to teachers, we obtained enough responses to each of three pilot forms—640 cases for form A, 535 for form B, and 377 cases for form C—to conduct statistical analyses. Each form was constructed such that roughly seven stems and 11–15 items represented each cell in Figure 1. "Testlet" items, or items linked by a common stem or scenario like item 2 in the Appendix, were combined into one item for the factor analysis. Within each cell, three "linking items" were constant across all three forms; these items allowed form equating in the evaluation portion of this project and also allowed us to test and confirm hypotheses about particular items across the three forms. The remaining items within a cell differed between forms but still followed the general themes and topics for items outlined above. Thus, factor analyses on each of the three forms could return consistent results broadly (e.g., finding the same number of factors, interpreting factors in the same way) and for a small number of linking items. To perform factor analyses, we used a program written to accommodate testlets (ORDFAC; Schilling, 2002b). To learn more about other item characteristics, we used BILOG (Mislevy & Bock, 1997), a program that enables item response theory analyses (Hambleton, Swaminathan, & Rogers, 1991).

Results and Discussion

In this analysis, we answered two questions: how teachers' mathematical knowledge for teaching is organized, and whether we can, with these items, reliably measure teachers' mathematical knowledge for teaching. The results we present here draw on the descriptions of data analysis presented in Schilling (2002a); readers who wish a more thorough and technical version of the factor analysis should refer to this manuscript.

Organizing Content Knowledge for Teaching

As we described above, the items used in the MPDIs can be organized in several ways. By putting all items on each form into ORDFAC, we could determine which items related to the same underlying constructs, how many such constructs existed, and with what certainty we could identify this structure.

This task was complicated by the limitations of the analytic method of item factor analysis, which identifies patterns of association between items for a particular sample answering a single survey instrument. A pattern of association is a necessary but not sufficient condition for identification of a unidimensional construct. Items measuring conceptually different constructs can also show a pattern of association in item factor analysis because of a strong correlation between the underlying constructs in that sample. Another sample, in which the same constructs do not exhibit as strong a correlation, will often show a different pattern of association, differentiating the two conceptually distinct constructs. As we describe below, this phenomenon was exhibited in two of our forms, where the patterns, functions, and algebra content items loaded on the same factor as the knowledge of students and content items in number concepts and operations. A related question was, assuming a number of conceptually distinct but related constructs for a set of items, the extent to which a single "general" factor could account for the covariation between items compared to the amount of covariation accounted for by specific factors. Addressing this issue provided insight into the meaning that might be attached to the use of a simple total score for our instruments. To address these concerns, we examined the data using three types of analyses: (*a*) exploratory factor analyses of the three forms; (*b*) factor analyses with patterns, functions, and algebra items removed, to seek additional clarity of results; (*c*) bi-factor analyses, to further assess the issue of multidimensionality and to resolve questions regarding knowledge of students and content items.

Exploratory factor analysis of all items on form A suggested that there were three underlying dimensions: (*a*) knowledge of content in number concepts and operations; (*b*) knowledge of content in patterns, functions, and algebra; and (*c*) knowledge of students and content in number concepts and operations. This is illustrated in Table 1, which presents Promax rotated factor loadings for all items. Note that all of the knowledge of number concepts and operations items, with one exception, loaded strongly on the first factor, and all the knowledge of patterns, functions, and algebra items loaded on the third factor. The situation for the knowledge of students and content items is more complicated. Most of these items (9 out of 14) loaded primarily on the second factor, but a significant minority loaded primarily on the first factor. This suggested that either knowledge of content or knowledge of students and content might alternatively be critical for answering these types of items correctly. However, inspection of each item failed to reveal any noticeable difference between the items loading on the first and second factors.

We also ran factor analyses on all items for forms B and C. The results of these analyses were consistent with form A in that the knowledge of number concepts and operations items loaded almost exclusively on the first factor and the knowledge of patterns, functions, and algebra items loaded exclusively on the third factor. However, the results for knowledge of students and content items differed across the three forms. In forms B and C these items loaded most often on the first (content) and third (algebra) factors; the second factor had only a few items with strong loadings on both forms.

Conceptually, there was little reason to believe that the student thinking and patterns, functions, and algebra items should be combined to form a scale. Instead, it was likely that these two constructs were correlated in the form B and C samples, as described above. This, combined with the results of the analysis of form A, suggested that the presence of the patterns, functions, and algebra items might be obscuring relations among the student thinking items. Therefore, we omitted these items from subsequent analyses and focused on whether knowledge of content and knowledge of students and content were distinguishable factors.

Results suggested that we could make such a differentiation. To start, we fit exploratory factor models of increasing complexity (number of factors) to these items. The results of these successive fits for each form are presented in Table 2. Schilling and Bock (in press) recommended that a model of increasing complexity only be accepted if the chi-square statistic for a model is two times the difference in the degrees of freedom between the two models. This heuristic is also employed in the Akaike Information Criterion (AIC) (Agresti, 1990), where a low value indicates better fit. By both criteria, a two-factor model provided the best fit for form A, whereas three-factor models provided the best fit for forms B and C. Table 3 presents the loadings for the two-factor models for all three forms, and Table

TABLE 1. Promax Rotated Factor Loadings, Form A

Item	Factor 1	Factor 2	Factor 3
Knowledge of content:			
Number concepts:			
1	**.512**	.138	.063
2	**.473**	.113	−.059
3	**.260**	−.132	.165
4	**.219**	.062	−.180
5	**.444**	.158	−.133
6	**.228**	.097	.086
7	.139	−.039	**.295**
Operations:			
1	**.732**	.068	−.203
2	**.246**	.084	.136
3	**.637**	−.210	.101
4	**.643**	.042	.036
5	**.704**	−.292	−.023
6	**.511**	−.143	−.052
Patterns, functions, and algebra:			
1	−.290	−.111	**.773**
2	.259	−.038	**.403**
3	.015	−.016	**.675**
4	.156	.175	**.314**
5	.337	−.047	**.419**
6	.039	−.113	**.639**
Knowledge of students and content:			
Number concepts:			
1	.061	**.275**	.121
2	**.263**	**.327**	.158
3	**.311**	.149	−.021
4	.002	**.365**	.010
5	−.109	**.248**	.212
6	.180	**.386**	.047
7	**.352**	−.041	.019
Operations:			
1	−.055	**.946**	−.098
2	**.466**	.181	−.058
3	.018	**.249**	.130
4	**.493**	−.022	.017
5	−.125	**.699**	−.124
6	−.012	**.417**	−.022
7	−.015	.149	.151

NOTE.—Boldface indicates highest loadings for each item.

4 presents the loadings of the three-factor models for form C. Examination of the factor loadings for the two- and three-factor models for form B revealed the two models to be essentially the same, with the exception that three of the knowledge of students and content in operations items comprised a third factor for the latter model. The two- and three-factor models for form C differed in that none of the knowledge of students and content in number concepts items had substantial loadings on the second factor for the two-factor model, but four of the eight items loaded on a third factor for the three-factor model.

Taken together, these exploratory analyses suggested at least three dimensions across all the items reflecting the following constructs: (*a*) knowledge of content (KC) in elementary number and operations; (*b*) knowledge of students and content (KSC) in elementary number and operations; and (*c*) knowledge of content (KC) in patterns, functions, and algebra. Although results differed across forms in the area of knowledge of students and content, they

TABLE 2. Exploratory Factor Analyses, Number Concepts/Operations

	χ^2	*df*	Akaike's Information Criterion
Form A:			
1 factor			25345
2 factor	96.96	26	25300
3 factor	44.82	25	25305
4 factor	43.34	24	25310
Form B:			
1 factor			22506
2 factor	72.76	28	22489
3 factor	66.20	27	22477
4 factor	47.76	26	22481
Form C:			
1 factor			16219
2 factor	111.98	29	16165
3 factor	69.46	28	16152
4 factor	37.56	27	16168

corresponded to the categorization system shown in Figure 1, assuming the combining of number and operations.

After these analyses suggested this general shape to our data, we ran a five-factor bi-factor model. This model specified the number of factors (four) and allows each item to load in two places: on a general factor that explains teachers' responses to all items, and on a specific factor representing its place in the categorization scheme described in Figure 1. Because exploratory factor analysis showed no difference between number and operations content items, however, we assigned both sets of items to only one factor. We tested this bi-factor model for three reasons. First, results would assist us in determining to what extent a general factor versus specific factors explained patterns in teachers' responses to these items. Second, results would allow us to better assess our hypothesis that there is a difference between common and specialized knowledge of content (SKC). Here we might expect items that tap common knowledge to load on the general factor and items that tap specialized knowledge to load on a specific factor. Finally, results would allow us to better understand the multidimensionality within the knowledge of students and content items.

Results from this bi-factor analysis were informative. First, the general factor explained between 72% and 77% of the overall variation in teachers' responses to items on each of the three forms (see Table 5). This factor explained variation in a substantial number of content knowledge items, suggesting that the factor can be interpreted as common knowledge of content (CKC) and suggesting an influence of general grasp of mathematics on teachers' responses to items. However, multidimensionality was also apparent here, because the factors describing knowledge of students and content (KSC) and knowledge of content (KC) in patterns, functions, and algebra accounted for between 21% and 45% of the communality in items written to represent these areas. Further, the SKC factor explained 12%–23% of the communality of items written to represent knowledge of content in number and operations.

Similar to results from the exploratory analysis, some knowledge of students and content items continued to load on the CKC factor, others loaded on their own factor, and many loaded on both. There were no firm patterns among items in how they loaded; both the CKC and KSC factors included items that referenced student errors, common strategies, similar subject-matter content, and a range of item difficulties. Whatever the cause of these loading patterns, it made sense to think that mathe-

TABLE 3. Number Concepts and Operations Promax Rotated Factor Loadings, Two-Factor Models

Item	Form A		Form B		Form C	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Content knowledge:						
Number concepts:						
1	**.546**	.159	.253	**.491**	**.641**	−.067
2	**.456**	.059	.192	**.342**	**.473**	.026
3	**.373**	−.126	.174	**.238**	**.222**	.027
4	.126	.016	**.408**	−.012	.193	**.267**
5	**.357**	.130	**.353**	.191	**.616**	−.118
6	**.273**	.110	**.591**	.269	**.551**	−.181
7	**.328**	.004	**.327**	**.457**	**.697**	−.090
8			−.126	**.517**		
9			**.594**	.068		
10			**.454**	.247		
Operations:						
1	**.606**	.031	**.844**	−.084	**.691**	−.005
2	**.338**	.090	**.395**	.098	**.411**	.012
3	**.761**	−.250	**.552**	−.140	**.415**	.043
4	**.681**	.027	.291	**.402**	**.234**	**.326**
5	**.699**	−.304	**.321**	.230	**.386**	.101
6	**.503**	−.176			**.659**	.004
7					**.492**	−.071
8					**.295**	**.366**
Knowledge of students and content:						
Number concepts:						
1	.164	**.264**	.052	**.217**	.128	.188
2	**.360**	**.342**	.195	**.474**	**.647**	−.107
3	**.285**	.147	−.040	**.426**	**.580**	−.086
4	.018	**.356**	**.311**	**.271**	−.007	.060
5	.029	**.270**	.187	**.423**	**.579**	−.081
6	.208	**.396**			**.578**	−.008
7	**.359**	−.042			.213	.152
8					.201	.079
Operations:						
1	−.176	**1.006**	−.103	**.431**	−.264	**.919**
2	**.427**	.168	.077	**.473**	.302	**.388**
3	.097	**.265**	−.411	**.719**	−.190	**.659**
4	**.500**	−.012	**.310**	−.041	.210	**.372**
5	−.195	**.669**	.124	.150	−.045	**.700**
6	−.047	**.441**	.062	**.464**	.174	**.552**
7	.057	.182	.137	**.316**	.035	**.518**
8			.071	.224		
9			−.036	**.551**		

NOTE.—Boldface indicates highest loadings for each item.

matical content knowledge and knowledge of students and mathematics should be interrelated, for it is difficult to imagine teachers having strong knowledge of students' learning without some basic knowledge of the mathematics they study.

Multidimensionality was also apparent in items written to represent knowledge of content in elementary number and operations. To a large extent, items representing common knowledge of content (CKC) tended to appear on the general factor, suggesting again that this factor represented overall mathematical ability. However, variation in teachers' responses to items written to represent specialized knowledge of content (SKC) was much more likely to be explained, at least in part, by the "specific" content knowledge factor. Finding this factor supported the conjecture that some content knowledge used in teaching is specific to key tasks in which teachers must engage.

TABLE 4. Promax Rotated Factor Loadings, Number Concepts/Operations, Three-Factor Solution, Form C

Item	Factor 1	Factor 2	Factor 3
Content knowledge:			
Number concepts:			
1	**.625**	−.066	.018
2	**.355**	−.062	**.284**
3	**.240**	.041	−.038
4	**.242**	**.316**	−.097
5	**.558**	−.137	.102
6	**.614**	−.096	−.201
7	**.783**	.021	−.250
Operations:			
1	**.725**	.046	−.103
2	**.355**	−.010	.112
3	**.393**	.041	.044
4	−.131	.046	**.894**
5	**.344**	.064	.112
6	**.702**	.067	−.124
7	**.571**	.009	−.211
8	**.299**	**.352**	.045
Knowledge of students and content:			
Number concepts:			
1	−.074	−.008	**.543**
2	**.624**	−.098	.016
3	**.462**	−.200	**.304**
4	−.276	−.177	**.661**
5	**.526**	−.097	.091
6	**.499**	−.072	.197
7	.098	.046	**.302**
8	.167	.044	.096
Operations:			
1	−.095	**.899**	−.129
2	**.271**	**.320**	.157
3	−.092	**.655**	−.070
4	.243	**.363**	−.004
5	−.107	**.564**	.303
6	.131	**.462**	.210
7	.003	**.439**	.183

NOTE.—Boldface indicates highest loadings for each item.

TABLE 5. Percentage of Communality Explained by General and Specific Factors

	Knowledge of Content		Knowledge of Students and Content	
	Number and Operations	Patterns, Functions, and Algebra	Number and Operations	Total
Form A:				
General	88.1	54.7	65.6	72.5
Specific	11.9	45.3	34.4	27.5
Form B:				
General	83.3	72.8	77.3	79.1
Specific	16.6	27.2	22.6	20.9
Form C:				
General	76.7	78.9	75.5	76.6
Specific	23.3	21.1	24.5	23.4

Inspecting the items that comprised this factor further supported this hypothesis and suggested some characteristics of this knowledge. These items included those that engage teachers in (*a*) analyzing alternative algorithms or procedures, (*b*) showing or representing numbers (e.g., 10.05) or operations (e.g., $1/2 \times 2/3$) using manipulatives, and (*c*) providing explanations for common mathematical rules (e.g., why any number can be divided by 4 if the last two digits are divisible by 4). These corresponded closely to our initial ideas about the constituent parts of specialized knowledge of content, with one exception: items that asked teachers to match fractions number sentences to stories (e.g., represent $1\frac{1}{4} \div \frac{1}{2}$ with a story) appeared on the general factor. Nevertheless, finding this specific factor supported the idea of specialized knowledge of content.

Evidence that supports the existence of specialized content knowledge for teaching is important. From a measurement standpoint, these results suggested that common and specialized mathematical knowledge are related yet are not completely equivalent; the possibility exists that individuals might have well-developed common knowledge yet lack the specific kinds of knowledge needed to teach. Results also indicated that individuals might develop the specialized knowledge for teaching mathematics—perhaps from teacher preparation, professional development, or working with students or curriculum materials—without having otherwise expert knowledge of mathematical content. This finding has implications for theory, policy, teacher preparation, and measurement. We discuss some of these below.

Apart from the major structure of the data, several things about these findings stand out. First, these items were not organized around generic tasks of teaching (e.g., evaluating curriculum materials, interpreting students' work). Instead, these results suggested that the organization of teachers' knowledge is at least somewhat content specific. Yet these constructs were not highly particular, either: instead of finding specific factors that represented either content (e.g., fractions, whole-number computation) or specific tasks of teaching mathematics (e.g., representing numbers and operations, analyzing student errors), we found broader groupings of items. Finally, the items that appeared on all three forms tended to perform consistently across those forms in our factor analyses, with minor exceptions (see Schilling, 2002a).

Overall, results from these factor analyses revealed that teachers' content knowledge for teaching is at least somewhat domain specific, and that scholars who have hypothesized about the categories around which teacher knowledge might organize are at least partially correct. Subject-matter content does play a role; so do the different ways mathematical knowledge is used in classrooms. Including additional content areas (geometry, data and statistics) and a fuller array of knowledge of students and content items (e.g., in algebra, geometry) would allow further testing of this finding. In the meantime, we consider these findings' implication for constructing measures of teacher knowledge.

Measuring Content Knowledge for Teaching

Given the results from the factor analysis, could we construct reliable measures that accurately represent teachers' ability in these areas? This was a major goal of our work, for these measures are needed to gauge the effectiveness of professional development and other teacher learning opportunities and to estimate the contribution of teacher knowledge to student achievement.

We used BILOG to fit initial item response theory (IRT) models to the data (Hambleton et al., 1991). We present results for scales for (*a*) each cell in Figure 1, (*b*) for combined number concepts/operations knowledge of content and knowledge of students and content scales, and (*c*) for an overall measure of mathematical knowledge for teaching. Table 6 provides descriptive statistics for these scales on each of the three forms—coefficient alphas for a classical test theory measure of reliability, IRT reliabilities computed using BILOG, and points of maximum test information. The reliabilities for patterns, functions, and algebra scales, as well as for scales that combined number and operations items within each domain, were good to excellent, ranging from 0.71 to 0.84. However, the points of maximum information revealed how each scale could be improved. The lowest reliability of 0.71 occurred for the knowledge of students and content scale on form A where the maximum information was 1.9 standard deviations below the population mean. In contrast, the number concepts/operations content knowledge scale for form A had point of maximum information at 0.36 standard deviations below the population mean. This meant that this scale was better targeted to the skills of the population; hence the scale had a higher reliability—0.81.

There are several things to note about these efforts to build measures that reflect individuals' content knowledge for teaching mathematics. First, measures representing teachers' knowledge of content had higher reliabilities than those composed of items meant to measure familiarity with students and content. Second, for most measures, the test provided the most information (test information curve maximum) at abilities below the average teacher, that is, items were, on average, too easy, yielding the best measurement (lowest standard er-

TABLE 6. Reliabilities and Points of Maximum Information

Form/Scale	Items (*N*)	Alpha	IRT Reliability	Max Info
Form A:				
Knowledge of content:				
Number concepts	13	.536	.654	−.51
Operations	13	.617	.709	−.21
Patterns, functions, algebra	12	.740	.771	−.79
Combined number and operations	26	.719	.810	−.36
Knowledge of students and content:				
Number concepts	10	.494	.576	−.67
Operations	10	.450	.534	−1.97
Combined number and operations	20	.622	.709	−1.90
Total	58	.845	.907	−.76
Form B:				
Knowledge of content:				
Number concepts	13	.670	.741	−1.45
Operations	11	.568	.655	−.76
Patterns, functions, algebra	12	.793	.805	−1.21
Combined number and operations	24	.766	.831	−1.27
Knowledge of students and content:				
Number concepts	8	.507	.578	−.50
Operations	11	.544	.610	−1.29
Combined number and operations	19	.657	.727	−1.16
Total	55	.878	.916	−1.33
Form C:				
Knowledge of content:				
Number concepts	11	.653	.742	−.95
Operations	12	.675	.758	.21
Patterns, functions, algebra	10	.824	.801	−.81
Combined number and operations	23	.784	.839	−.17
Knowledge of students and content:				
Number concepts	11	.552	.655	−.43
Operations	10	.649	.689	−1.51
Combined number and operations	21	.698	.781	−1.11
Total	54	.888	.931	−.92

rors) for teachers who scored between .5 and 2.0 standard deviations below average. This trend was most pronounced in the knowledge of students and content measures. Third, there remain some significant problems with multidimensionality with these items, particularly in the areas of knowledge of students and content and, for those who choose to use this construct, the specialized knowledge of content. For more on potential solutions to this problem, see Schilling (2002a).

Finally, any appraisal of the utility of a measure must include an examination of the relation between individuals' performance on the instrument and those individuals' real skill or ability, that is, of validity. For these measures, a best-case investigation of validity would include comparing teachers' measure score with an assessment of their use of mathematics content during classroom teaching. This work is currently under way. Less convincing, although more often done in the field of test construction, are cognitive tracing interviews, in which individuals talk through their thinking and answers about items. If individuals' thinking does not reflect their answers, problems of validity are likely. Although an analysis conducted with items similar to these suggested that, for knowledge of content items, teachers' answers represented their underlying reasoning, results were not so sanguine for student thinking items, where more problems pertained (see Hill, 2002). This suggests that the more varied factor analysis results and lower reliabilities for this second set of items may be related to problems with measurement in this domain.

Conclusion

By developing measures of teacher knowledge for teaching mathematics, we hope to contribute to a number of ongoing efforts in educational research to answer policy-relevant questions: identifying the effect of teacher knowledge on student achievement, explaining how teacher knowledge develops (via experience, professional training, professional development), and answering other key policy questions (e.g., the effects of certification on teacher knowledge). However, we believe developing such measures can also contribute to a renewal of interest in the theoretical aspects of professional knowledge for teaching by allowing insight into how knowledge is held by teachers, how that knowledge relates to common subject-matter knowledge, and perhaps even (through open-ended interviews) how teachers, nonteachers, and subject-matter experts deploy knowledge.

We see the analyses reported in this article as a first step in the measures-development process. The data set was less than ideal, because teachers were sampled nonrandomly and MPDI pre- and posttests were combined for this analysis. Because different subjects answered different forms on the pretest and posttest, combining data did not present significant problems for our use of IRT, other than perhaps producing a nonnormal distribution of ability in the sample for a particular form. Fortunately, IRT models are generally robust to nonnormal distributions of ability (Bock & Aitken, 1981). We also measured typical, rather than expert, teachers, and this may further constrain our results: if typical teachers do not have or have less specialized knowledge for teaching mathematics, we bias our results toward a null finding for this hypothesis. And these findings should also be replicated, both through studies similar to the one reported here and also through the use of multiple methods, including interviews and observations of classroom instruction.

However, our analyses suggest that some tentative results can be reported now. First, repeated analyses across three forms show evidence of multidimensionality in these measures, suggesting that teachers' knowledge of mathematics for teaching is at least partly domain specific rather than simply related to a general factor such as overall intelligence, mathematical ability, or teaching ability. Although results from the bi-factor analysis suggest that such a general factor does operate, additional communality is explained by specific dimensions; this supports Shulman's and others' claims that knowledge for teaching consists of both general knowledge of content and knowledge in more specific domains.

The domains identified in the factor analyses are themselves interesting. Our data indicate that in addition to a general factor, specific factors represent knowledge of content in number and operations, knowledge of students and content in number and operations, and the relatively newer area (for elementary school) of knowledge of content in patterns, functions, and algebra. The data also suggest a specialized knowledge of content measure made up of several types of items: representing numbers and operations, analyzing unusual procedures or algorithms, and providing explanations for rules. Writing items that represent more content areas, more specialized tasks (e.g., using mathematical definitions in teaching), and possibly more domains (e.g., knowledge of teaching and content) will allow us to assess the extent to which content and task continue to play a role in defining domains of teacher thinking.

Our findings imply lessons for theory, policy, and measurement. They provide evidence for the conjecture that content knowledge for teaching mathematics consists of more than the knowledge of mathematics held by any well-educated adult. Although such knowledge of mathematics appears to be an important component of the knowledge needed for teaching, there may be more mathematical depth to teaching elementary school, in other words, than simply the content of a third-, fifth-, or even eighth-grade textbook. We cannot say what specific areas teachers must know to help students learn—such a statement must wait for the results of analyses that compare the effects of different kinds of teacher knowledge on students' growth in classrooms. But our results hint that rather than focusing on how much mathematics an individual knows, as has historically been the case (see Shulman, 1986), researchers must also ask how an individual holds and uses that knowledge—whether a teacher can use mathematical knowledge to generate representations, interpret student work, or analyze student mistakes. Our findings also suggest the utility of continuing to identify the content, so to speak, of our specialized knowledge of content category and thus extending our notions of the knowledge needed to teach.

If our results hold, they also bear on current policy debates regarding the recruitment and preparation of teachers. Strong knowledge of basic mathematical content does matter; however, policy makers must take seriously the idea that additional capabilities may be layered atop that foundation. Until we can replicate these results, we cannot definitively say that teachers should learn this information in preservice or in-service preparation. Yet finding evidence for these multiple dimensions lends support for a curriculum that goes into depth and that is specific to the work of teaching. Teachers may need to know why mathematical statements are true, how to represent mathematical ideas in multiple ways, what is involved in an appropriate definition of a term or a concept, and methods for appraising and evaluating mathematical methods, representations, or solutions. By helping teachers develop knowledge of mathematics that goes beyond the understanding needed for everyday nonprofessional functioning, faculty and professional developers may assist teachers in preparing for the tasks they will encounter on the job.

From a policy perspective, our research suggests supporting professional development and teacher preparation programs that enable this kind of learning. However, it also carries a lesson for those who construct teacher licensure exams, at least at the elementary level; reviews of several currently used exams indicate that the majority of problems simply ask teachers to compute, rather than to use knowledge in more classroom-authentic ways. If researchers find that the more specific kinds of expertise identified here affect student achievement—or even if researchers simply decide, based on normative arguments, that teachers should possess this knowledge—licensure exams should reflect this emphasis.

From a measurement perspective, these results support constructing separate scales to represent knowledge for teaching mathematics. This is an important point for researchers, who aim to devise measures that are sensitive to differences in individuals' unique combinations of knowledge and skills in order to explore the relationship between such measures and others like student achievement. The presence of multidimensionality also changes the way researchers might model teacher development and its contribution to student achievement. Rather than using one catchall variable, researchers can contrast the effects of teacher growth in various domains on student achievement and predict the effects of growth in different domains resulting from various "treatments," such as the effect of the first years of teaching on knowledge of student strategies, mistakes, and methods.

Appendix

Examples of Items Measuring Content Knowledge for Teaching Mathematics

1. Mr. Allen found himself a bit confused one morning as he prepared to teach. Realizing that ten to the second power equals one hundred ($10^2 = 100$), he puzzled about what power of 10 equals 1. He asked Ms. Berry, next door. What should she tell him? (Mark (X) ONE answer.)

a) 0

b) 1

c) Ten cannot be raised to any power such that ten to that power equals 1.

d) −1

e) I'm not sure.

2. Imagine that you are working with your class on multiplying large numbers. Among your students' papers, you notice that some have displayed their work in the following ways:

Student A	*Student B*	*Student C*
35 ×25 125 +75 875	35 ×25 175 +700 875	35 ×25 25 150 100 +600 875

Which of these students would you judge to be using a method that could be used to multiply any two whole numbers?

	Method would work for all whole numbers	Method would NOT work for all whole numbers	I'm not sure
a) Method A	1	2	3
b) Method B	1	2	3
c) Method C	1	2	3

3. Mr. Fitzgerald has been helping his students learn how to compare decimals. He is trying to devise an assignment that shows him whether his students know how to correctly put a list of decimals in order of size. Which of the following sets of numbers will best suit that purpose?

a)	.5	7	.01	11.4
b)	.60	2.53	3.14	.45
c)	.6	4.25	.565	2.5

d) Any of these would work well for this purpose. They all require the students to read and interpret decimals.

4. Mrs. Jackson is getting ready for the state assessment, and is planning mini-lessons for students focused on particular difficulties that they are having with adding columns of numbers. To target her instruction more effectively, she wants to work with groups of students who are making the same kind of error, so she looks at a recent quiz to see what they tend to do. She sees the following three student mistakes:

	1		1		1
I)	38	II)	45	III)	32
	49		37		14
	+65		+29		+19
	142		101		64

Which have the same kind of error? (Mark ONE answer.)

a) I and II
b) I and III
c) II and III
d) I, II, and III

References

Agresti, A. (1990). *Categorical data analysis.* New York: Wiley.

Ball, D. L. (1988). *Knowledge and reasoning in mathematical pedagogy: Examining what prospective teachers bring to teacher education.* Unpublished doctoral dissertation, Michigan State University, East Lansing.

Ball, D. L. (1990). The mathematical understandings that prospective teachers bring to teacher education. *Elementary School Journal,* **90**(4), 449–466.

Ball, D. L. (1991). Research on teaching mathematics: Making subject matter part of the equation. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. **2**, pp. 1–48). Greenwich, CT: JAI.

Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.

Ball, D. L., & Bass, H. (2003). Making mathematics reasonable in school. In G. Martin (Ed.), *Research compendium for the Principles and Standards for School Mathematics* (pp. 27–44). Reston, VA: National Council of Teachers of Mathematics.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika,* **46,** 443–459.

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**(3), 194–222.

Bruner, J. (1960). *The process of education.* Cambridge, MA: Harvard University Press.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary school.* Portsmouth, NH: Heinemann.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement* (CPRE Research Report RR-043). Philadelphia: Consortium for Policy Research in Education.

Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education.* New York: Teachers College Press.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hill, H. C. (2002) *Cognitive tracing study: Validity of items in representing teachers' thinking.* Unpublished manuscript, University of Michigan, Ann Arbor.

Leinhardt, G., & Smith, D. A. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology,* **77**(3), 247–271.

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

Madfes, T., Montell, F., & Rosen, R. (2002). *Independent evaluator: California Professional Development Institutes, Mathematics.* San Francisco, CA: WestEd.

Mislevy, R. J., & Bock, R. D. (1997). *Bilog: Item analysis and test scoring with binary models.* Lincolnwood, IL: Scientific Software International.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching and America's future.* New York: Author.

Schilling, S. G. (2002a). *Measuring pedagogical content knowledge: An integrated measurement methodology.* Unpublished manuscript, University of Michigan, Ann Arbor.

Schilling, S. G. (2002b). ORDFAC software. Ann Arbor, MI: Author.

Schilling, S. G., & Bock, R. D. (in press). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika.*

Schwab, J. (1961/1978). *Science, curriculum, and liberal education: Selected essays.* Chicago: University of Chicago Press.

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

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review,* **57,** 1–22.

U.S. Department of Education. (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, Office of Policy, Planning, and Innovation.

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). Sussex: Holt, Rinehart & Winston.

Wilson, S. M., & Wineburg, S. (1988). Peering at history through different lenses: The role of disciplinary perspectives in teaching history. *Teachers College Record,* **89,** 525–539.