Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement

Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement

Heather C. Hill
Brian Rowan
Deborah Loewenberg Ball

In press at American Educational Research Journal.

Heather C. Hill is an Assistant Research Scientist at the School of Education, 610 E. University #1600, University of Michigan, Ann Arbor, MI, 48109; email hhill@umich.edu. Her areas of specialization include educational policy, instruction, and teachers' content knowledge.

Brian Rowan holds the Burke A. Hinsdale Collegiate Professorship in Education at the School of Education, 610 E. University areas of specialization lie at the intersection of organization theory and school effectiveness research.

Deborah Loewenberg Ball holds the William H. Payne Collegiate Professor in Education at the School of Education, 610 E. University #1600, University of Michigan, Ann Arbor, MI, 48109; email dball@umich.edu. Her areas of specialization include the study of efforts to improve teaching through policy, reform initiatives, and teacher education.

Abstract

This study explored whether and how teachers' mathematical knowledge for teaching contributes to gains in students' mathematics achievement. We used linear mixed model methodology in which first (n=1190) and third (n=1773) graders' mathematical achievement gains over a year were nested within teachers (n=334 and n=365), who in turn were nested within schools (n=115). We found teachers' mathematical knowledge was significantly related to student achievement gains in both first and third grades, controlling for key student and teacher-level covariates. While this result is consonant with findings from the educational production function literature, our result was obtained using a measure of the specialized mathematical knowledge and skills used in teaching mathematics. This result provides support for policy initiatives designed to improve students' mathematics achievement by improving teachers' mathematical knowledge.

KEYWORDS: educational policy; mathematics; student achievement; teacher knowledge

In recent years, teachers' knowledge of the subject matter they teach has attracted increasing attention from policymakers. To provide students with "highly qualified teachers," No Child Left Behind requires teachers to demonstrate subject-matter competency through subject matter majors, certification, or other means. Programs such as California's Professional Development Institutes and the National Science Foundation's Math-Science Partnerships are aimed at providing content-focused professional development intended to improve teachers' content knowledge. This focus on subject matter knowledge has arisen, at least in part, because of evidence suggesting that U.S. teachers lack essential knowledge for teaching mathematics (e.g., Ball 1990; Ma 1999), and because evidence from the educational production function literature suggests that teachers' intellectual resources significantly affect student learning.

Despite this widespread interest and concern, what counts as "subject matter knowledge for teaching" and how it relates to student achievement has remained inadequately specified in past research. A closer look at the educational production function literature, for example, reveals that researchers working in this tradition have typically measured teachers' knowledge using proxy variables, such as courses taken, degrees attained, or results from basic skills tests. This stands in sharp contrast to another group of education scholars who have begun to conceptualize teachers' knowledge for teaching differently, arguing that teacher effects on student achievement are driven by teachers' ability to understand and use subject matter knowledge to carry out the tasks of teaching (Ball 1990; Shulman, 1986; Wilson, Shulman, Richert & 1987). In this view, mathematical knowledge for teaching goes beyond that captured in measures of mathematics courses taken or basic mathematical skills. For example, teachers of mathematics not only need to calculate correctly, but also know how to use pictures or diagrams to represent mathematics concepts and procedures to students, provide students with explanations for common rules and mathematical procedures, and analyze students' solutions and explanations. By inadequately measuring teachers' knowledge, existing educational production function research could be limited in its conclusions, not only about the magnitude of effects that teachers' knowledge has on student learning, but also about the kinds of teacher knowledge that matter most in producing student learning..

As we discuss below, only a few educational production function studies have measured teachers' mathematical knowledge directly and used this as a predictor of student achievement (Harbison & Hanushek, 1992; Mullens, Murnane & Willett, 1996; Rowan, Chiang & Miller, 1997). Most other production function studies used tests of teacher verbal ability to predict achievement outcomes. As a result, despite conventional wisdom that elementary U.S. teachers' subject matter knowledge influences student achievement, no large-scale studies have demonstrated this empirically (Wayne & Youngs, 2003). Nor is the situation ameliorated by examining process-product research on teaching, in which both the measurement of subject-specific teaching behaviors and the direct measurement of teachers' subject matter knowledge were notably absent.

To remedy this situation, this study analyzes teachers' scores on a measure of <u>mathematical knowledge for teaching</u>. By "mathematical knowledge for teaching," we mean the mathematical knowledge used to carry out the <u>work of teaching mathematics</u>. Examples of this "work of teaching" include explaining terms and concepts to students, interpreting students' statements and solutions, judging and correcting textbook treatments of particular topics, using representations accurately in the classroom, and providing students with examples of mathematical concepts, algorithms, or proofs. Our previous work has shown that a measure composed of several multiple choice items representing these teaching-specific mathematical skills can both reliably discriminate among teachers and meet basic validity requirements for measuring teachers' mathematical knowledge for teaching (Hill, Schilling, and Ball, 2004). Here, we use teachers' scores on such a measure as a predictor of students' gains in mathematics achievement. An important purpose of the study is to demonstrate the independent contribution of teachers' mathematical knowledge for teaching to student achievement, net of other possible measures of teacher quality, such as teacher certification, educational coursework, and experience.

Framing the Problem

Since the 1960s, scholars and policymakers have explored the relationship between teacher characteristics, behaviors, and student achievement. Yet measures of teacher characteristics have varied widely, as have results from these investigations. Below, we outline how different research programs have measured characteristics of teachers and teaching and briefly summarize results from investigations using these measures.

Teachers in the Process-Product Literature

In classroom-level education research, attempts to predict student achievement from teacher characteristics have their origins in what has been called the process-product literature on teaching, that is, the large set of studies describing the relationship between teacher behaviors and student achievement. Moving beyond using affective factors such as teacher appearance and enthusiasm to predict student achievement, scholars in this tradition took the view that what teachers did in their classrooms might affect student achievement. By the late 1970s, these scholars had accumulated substantial evidence that certain teaching behaviors did affect students' achievement gains. For example, focusing class time on active academic instruction rather than classroom management, student choice/game time, personal adjustment, or non-academic subjects was found to be one consistent correlate of student achievement gains; so was presenting materials in a structured format via advance organizers, making salient linkages explicit, and calling attention to main ideas. Brophy & Good (1986), Gage (1978), Doyle (1977) and others provide excellent reviews of these findings. As this research progressed, scholars also designed experiments, training teachers in the behaviors indicated by previous research and comparing the academic performance of students in trained teachers' classrooms to that of students in untrained teachers' classrooms. Notably, Good, Grouws, & Ebmeier (1983) conducted such an experiment in mathematics and found that teachers who employed active teaching practices had students who performed better in basic skills but not problem-solving.

Critiques of process-product studies studies ranged from methodological – e.g., an excessive reliance on correlational data – to conceptual. Chief among the conceptual critiques was the lack of attention given in these studies to subject matter, and to how the subject being taught conditioned the findings described above (Shulman, 1986). What worked well to increase student achievement in mathematics, for instance, often did not work well to produce achievement gains in reading. Critics also pointed to the lack of attention to teachers' subject matter knowledge as a predictor of effective teaching and learning in this work.

Teachers in the Educational Production Function Literature

At the same time process-product studies were examining the relationship of classroom teaching behaviors and student achievement, other social scientists were focusing on the relationship between educational resources and outcomes. These studies, originating with the Coleman Report, collectively have been called "educational production function" studies. The main goal of this research program was to predict student achievement on standardized tests from the resources held by students, teachers, schools and others. Key resources were seen to include students' family background and socioeconomic status, district financial commitments to teacher salaries, teacher-pupil ratios, other material resources, and teacher and classroom characteristics (Hanushek, 1981; Greenwald, Hedges & Laine, 1996). Studies focusing specifically on teacher characteristics and student achievement employed two approaches, sometimes in combination, to measure the resources teachers bring to classrooms. In the first approach, information about teacher preparation and experience was collected and used as a predictor of student achievement. Key measures here included teacher education level, certification status, number of post-secondary subject matter courses taken, number of teaching methods courses taken, and years of experience in classrooms. By using such measures, researchers implicitly assumed a connection between formal schooling and employment experiences and the more proximate aspects of teachers' knowledge and performance that produce student outcomes. Reviews of this work have disputed the extent to which variables like teacher preparation and experience in fact contribute to student achievement (Begle 1972, 1979; Greenwald, Hedges, & Laine, 1996; Hanushek, 1981; 1996), with conflicting interpretations resting on the samples of studies and methods used for conducting meta-analyses. Beyond these methodological issues, however, another potential reason for the inherent uncertainties in research findings might be that teacher preparation and job experience are poor proxies for the kinds of teacher knowledge and skill that in fact matter most in helping students learn academic content.

Cognizant of this problem, a smaller number of production function studies have sought to measure teachers' knowledge more directly by looking at teachers' performance on certification exams or other tests of subject matter competence. By using such measures, these studies implicitly assume a relationship between teacher content knowledge as measured by such assessments and the kinds of teaching performances that produce improved student achievement. Studies using this approach typically find a positive effect of teacher knowledge, as measured by certification exams or tests of subject matter competence, on student achievement (e.g., Boardman, Davis & Sanday, 1977; Ferguson 1991; Hanushek 1972; Harbison & Hanushek, 1992; Mullens, Murnane & Willett, 1996; Rowan, Chiang & Miller, 1997; Strauss & Sawyer, 1986; Tatto, Neilsen, Cummings, Kularatna & Dharmadasa, 1993; for an exception, see Summers & Wolfe, 1977; for reviews, see Greenwald, Hedges & Laine, 1996; Hanushek, 1986; Wayne & Youngs, 2003). However, although this is an important research finding, it cannot fully describe how teacher knowledge relates to student achievement. One reason is that the studies just described have been conducted only in a limited number of academic subjects. For example, many studies have shown a relationship of teachers' verbal ability to gains in student achievement, but only three have focused explicitly on both teachers' and students' <u>mathematical</u> knowledge and student gains in mathematics achievement

(Harbison & Hanushek, 1992; Mullens, Murnane & Willett, 1996; Rowan, Chiang & Miller, 1997). Unfortunately, the design of these studies limited the degree to which their findings could be generalized. Two of the mathematics studies cited above, for example, took advantage of an assumed greater variation in teacher preparation and ability in other countries to estimate the effects of mathematics content knowledge on students' mathematics achievement (Harbison & Hanushek, 1992; Mullens, Murnane & Willett, 1996). Although these analyses have been fundamental to building the theoretical case for the importance of teachers' mathematical knowledge in producing student achievement gains in mathematics, the findings might not generalize to U.S. contexts, where teacher preparation and knowledge might be both higher and more uniform than in less-developed nations. Other production function studies also have been flawed by additional problems, including problems of aggregation bias, the use of cross-sectional rather than longitudinal data, and the use of composite measures of both teachers' knowledge and students' achievement.

From our perspective, however, the most pressing problem in production function studies remains the imprecise definition and indirect measurement of teachers' intellectual resources, and by extension, the mis-specification of the causal processes linking teachers' knowledge to student learning. Measuring quality teachers through performance on tests of basic verbal or mathematics ability may overlook key elements in what produces quality teaching. Effectiveness in teaching resides not simply in the knowledge a teacher holds personally but how this knowledge is used in classrooms. Teachers highly proficient in mathematics or writing will only help others learn mathematics or writing if they are able to use their own knowledge to perform the tasks they must enact as teachers —– for example, to hear students, to select and make use of good assignments, and to manage discussions of important ideas and useful work on skills. Yet these additional content-related abilities specific to the work of teaching have not been measured or included in the educational production function models. Harbison and Hanushek (1992), for instance, administered the same 4th grade math assessment to teachers and students, using scores from the first group to predict performance among the second. Mullens, Murnane, and Willett (1996) used teachers' scores recorded on the Belize National Selection Exam, a primary-school leaving exam[1] administered to all students seeking access to secondary school. Rowan, Chiang and Miller (1997) used a one-item assessment of teacher knowledge; however, because no scaling or validation work was done on that item, little can be said about what and how well it measures. While the results of each of these studies suggested the importance of teachers' knowledge in producing student learning, we argue that recent theoretical work on how teachers' content knowledge matters for the quality of teaching leads to a need for measures more closely attuned to the mathematical knowledge used in teaching. We turn next to this literature in order to elaborate our argument.

Teachers in the Teacher Knowledge Literature

Existing alongside production function research, an alternative literature focused directly on teacher knowledge has begun to ask what teachers need to know about subject matter content in order to teach it to students. In this research program, researchers propose to distinguish between the ways in which academic content must be known to teach effectively and the ways in which ordinary adults know such content. Shulman (1986; 1987) and colleagues (e.g., Wilson, Shulman, & Richert, 1987) launched this line of inquiry with their groundbreaking work on what accomplished teachers know. In his 1986 presidential address to the American Educational Research Association, Shulman originally proposed three categories of teacher subject matter knowledge. His first category, content knowledge, was intended to denote "the amount and organization of knowledge . . . 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, 1961/1974). The second category advanced by Shulman and his colleagues (Shulman, 1986; Wilson, Shulman, & Richert, 1987) was pedagogical content knowledge. With this category, he went "beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching" (p. 9, emphasis added). The concept of pedagogical content knowledge attracted the attention and interest of researchers and teacher educators alike. Components of pedagogical content knowledge, according to Shulman (1986), are representations of specific content ideas, as well as an understanding of what makes the learning of a specific topic difficult or easy for students.[2] Shulman's third category, curriculum knowledge, involves awareness of how topics are arranged both within a school year and over time and ways of using curriculum resources, such as textbooks, to organize a program of study for students.

Shulman and colleagues' work expanded ideas about how knowledge might matter to teaching, suggesting that it is not only knowledge of content but also knowledge of how to teach such content that conditions teachers' effectiveness. Working in depth within different subject areas — history, science, English, mathematics — scholars probed the nature of the content knowledge needed by teachers. In this program of work, comparisons across fields were also generative. Grossman (1990), for example, articulated how teachers' orientations to literature shaped the ways in which they approached texts with their students. Wilson and Wineburg (1988) showed how social studies teachers' disciplinary backgrounds — political science, anthropology, sociology –– shaped 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, Eisenhart, et al., 1992; Leinhardt & Smith, 1985).

Until now, however, it has not been possible to link teachers' professionally usable knowledge of their subjects to student achievement. Most of the foundational work on teacher knowledge has been qualitative in orientation and has relied principally on teacher case studies (e.g., Grossman, 1990), expert-novice comparisons (Leinhardt & Smith, 1985), international comparisons (Ma, 1999), and studies of new teachers (Ball, 1990; Borko et al., 1992). Although such studies have been essential in beginning to specify the mathematical content knowledge needed by teachers, they have not been designed to test hypotheses regarding how elements of such knowledge contribute to helping students learn. The result has meant that although many assume, based on the educational production function literature, that teachers' knowledge does matter in producing student achievement, whether and how it affects student learning has not yet been empirically established.

To address these issues, the Study of Instructional Improvement (SII) began in 1999 to design measures of elementary teachers' knowledge for teaching mathematics. In response to the above-reviewed literature, these efforts focused on producing a survey instrument that could measure the knowledge used in teaching elementary school mathematics (Ball & Bass 2000, 2003). By "used in teaching," the developers of this instrument meant to capture not only the actual content teachers taught –e.g., decimals, area measurement, or long division – but also the <u>specialized knowledge of mathematics</u> needed for the work of teaching. "Specialized" content knowledge might include knowing how to represent quantities such as 1/4 or .65 using diagrams, how to provide a mathematically careful explanation of divisibility rules, or how to appraise multiple solution methods for a problem such as 35 x 25. The desire to design survey measures of teacher knowledge also led developers to construct items centered directly on the content of the K-6 curriculum, rather than items that might appear on a middle- or high-school exam for students. Details on measure design, construction, and scaling are presented below.

Method

In this section we provide an overview of this project, describing the sample of students and teachers participating in the study, and providing information about data collection instruments and response rates. We also explain data analysis methods and model specifications used to estimate the relationship of teachers' content knowledge for teaching and student gains in mathematics achievement.

<u>Sample</u>

The data presented here come from a study of schools engaged in instructional improvement initiatives. As part of this study, researchers collected survey and student achievement data from students and teachers in 115 elementary schools during the 2000-

01 through 2003-04 school years. Eighty-nine of the schools in this study were participating in one of three leading Comprehensive School Reform programs—America's Choice, Success for All, and Accelerated Schools Project—with roughly thirty schools in each program. Additionally, 26 schools not participating in one of these programs were included as comparison schools. Program schools were selected for the study via probability sampling from lists supplied by the parent programs,[3] with some geographical clustering to concentrate field staff resources. Comparison schools were selected to match program schools in terms of community disadvantage and district setting. Once schools agreed to participate in the study, project staff approached all classroom teachers in each school to encourage their involvement.

The achieved sample of schools in this study differs from a nationally representative sample of schools in two ways. First, the sampling procedure deliberately selected schools engaged in instructional improvement; in addition, the achieved sample was deliberately constructed to over-represent high-poverty elementary schools in urban, urban fringe, and suburban areas. In particular, whereas 1999 statistics show that the average U.S. school served neighborhoods where 13% of the households were in poverty, the average school in the study sample served neighborhoods where 19% of the households were in poverty (Benson, 2002). Moreover, 68% of the schools in the study sample were located large and small cities, and no schools were located in rural areas.

Table 1 examines the characteristics of students who entered the present study in kindergarten and compares this sample to the nationally-representative sample of kindergarten students participating in Early Childhood Longitudinal Study (ECLS).[4] Table 1 suggests kindergarten sample in our study differed only slightly from a nationally-representative sample of kindergarten students. This finding strongly suggests that our study sample includes a sufficient range of children, schools, and educational contexts to make reasonable statements about the contribution of teachers' knowledge to student achievement. In particular, there is little indication that student variables are truncated in our sample in ways that would limit statistical inferences or the ability to generalize our findings to a larger population of schools and students. Our final sample of students includes 1190 first graders and 1773 third graders.

Just as students in the study were from varied social backgrounds, schools in the study were located in many different policy and social environments. For example, schools in the study sample were located in 42 districts in 15 states. States varied in size, in state average NAEP scores, and in approaches to improving low-performing schools. While 3 states were scored as being among the least interventionist on the accountability index designed by Carnoy and Loeb (2002), another 4 states scored at the top of this scale, indicating they were pursuing strong state-level rewards and sanctions to improve schools and student performance. The remaining 8 states clustered near the less interventionist end of the Carnoy and Loeb scale. In one state and several districts, participation in a comprehensive school reform was mandatory for schools performing below a certain level; in other states and districts, comprehensive school reforms were entirely optional.

The teacher sample for this study included 334 first grade and 365 third grade teachers. These teachers were fairly typical of the elementary teaching force, particularly in urban schools. Within the study sample 86% of teachers were female; 55% were White, 23% were Black, and 9% were Hispanic. About 90% of teachers in the sample were fully certified, and the average teacher in the sample had just over 12 years teaching experience.

Data Collection Instruments

Data collection for the study centered on two cohorts of students, one that entered the study as kindergarteners and was followed through second grade, and another that entered the study as third graders and was followed to the end of fifth grade. For each cohort, data were collected in two waves; in 2000-2001, the study collected information on first and third graders in 53 schools; in 2001-2002, the study collected information on an additional 62 schools. These two waves of data collection have been collapsed in the data analyses. In the remainder of this paper, we discuss the data collected on first and third grade students in this design, reporting response rates for instruments separately for the two waves of data collection (see below).

Data for the study were collected using a variety of survey instruments. Data on students, for example, came from two major sources: student assessments and parent interviews. Student assessments were administered in the fall and spring of every academic year, for a maximum of six administrations over the course of the study. The test forms and content of these assessments is discussed in more detail below. For now, we simply note that the assessments were given to eight randomly selected students per classroom and administered outside students' usual classroom by trained project staff. Project staff also contacted the parents or guardians of sampled students once by telephone in order to gather information about students' academic history, parent/guardian employment status, and other relevant home background variables. The completion rate for the student assessment averaged 96% across the 2000-2001 and 2001-

2002 school years.[5] The completion rate for the parent interview was 85% and 76% in 2000-2001 and 2001-2002, respectively.

Teacher data were gathered from two main instruments, a teacher log that teachers completed up to 60 times during one academic year and an annual questionnaire filled out each year of the study. The log was a highly structured self-report instrument asking teachers to record the amount of time devoted to mathematics instruction on a given reporting day, the mathematics content covered on that day, and the instructional practices used to teach that content. Teachers filled out logs for six-week periods in the fall, winter, and spring. Each log recorded one day of learning opportunities provided to one of eight randomly selected target students for whom achievement data also were collected. The response rates for log data were quite high. Overall, 97% (2000-2001) and 91% (2001-2002) of eligible teachers agreed to log, and of the roughly 60 logs assigned to each logging teacher, 91% were completed and returned in useable form to project staff.

The mathematics log used here was subjected to extensive development, piloting, and validation work. An observational study of a pilot version of the log found that agreement rates between teachers and trained observers was 79% for large content descriptors (e.g., number, operations, geometry), 73% for finer descriptors of instructional practice (e.g., instruction on why a standard procedure works), and that observer and teacher reports of time in mathematics instruction differed by under 10 minutes of instruction for 79% of lessons (Ball, Camburn, Correnti, Phelps, & Wallace, 1999).

Each year of the study, teachers were also asked to complete a questionnaire containing items about their educational background, involvement in and perceptions of school improvement efforts, professional development, and language arts and mathematics teaching. Notably, this survey is the source of items included in the content knowledge for teaching mathematics measure used here and described below. Table 2 shows that roughly three-quarters of eligible teachers returned completed teacher questionnaires each year; because most of the non-content knowledge questions (e.g., on certification) remained the same on each teacher questionnaire, we were able to construct many of the variables described below even when teachers did not complete a questionnaire during the time period when their students were under study.

Measures

Having described major instruments and response rates, we next turn to the specific measures used in the study. We begin by describing student achievement measures and then work outward to measures of family, teacher, classroom, and school characteristics. Table 3 shows means and standard deviations for the measures discussed below.

Student achievement. The measures of student achievement used here were drawn from CTB/McGraw Hill's Terra Nova Complete Battery (for spring of kindergarten), the Basic Battery (in spring of 1st grade), and the Survey (in third and fourth grades). Students were assessed in the fall and spring of each grade by project staff, and student scores were computed by CTBS using item response theory (IRT) scaling procedures. These scaling procedures yielded interval-level scores from student's raw responses. For the analyses conducted here, we computed gain scores from these IRT scale scores. For the first grade sample, we simply subtracted each student's spring of kindergarten IRT mathematics scale score from their spring of first grade mathematics score. For the third grade sample (for which spring 2nd grade data were not available), we subtracted the fall of third grade mathematics scale score from the fall of fourth grade score. The result in both cases was a number representing how many IRT scale score points students gained over one year of instruction.

The Terra Nova is widely used in state and local accountability and information systems. Its use here, therefore, adds to the generalizability of study results in the current policy environment. However, the construction of the Terra Nova adds several complexities to our analyses. To start, data from the mathematics logs indicate that the average student has a 70% chance of working on number concepts, operations, or pre-algebra and algebra in any given lesson (Rowan, Harrison & Hayes 2004). For this and other reasons, our mathematics knowledge for teaching measure was constructed solely from items on these three "focal" topics. However, the Terra Nova contains items from many additional content domains that are spread more broadly across the elementary school mathematics curriculum. At level 10 (spring of kindergarten), only 43% of Terra Nova items covered the focal topics included on our teacher knowledge measure. At level 12 (fall of third grade), 54% of Terra Nova items aligned with the focal topics. As this implies, there is an imperfect alignment between our measures of mathematical knowledge for teaching and measures of students' mathematical knowledge. It is well known that imperfect alignment of independent and criterion measures in research on teaching can lead to underestimates of effect sizes, suggesting that our empirical analyses probably underestimate the effects of teachers' content knowledge on student gains in mathematics achievement.

Student mobility also affects the data analyses reported here. By design, the larger study from which our data were drawn collected student achievement data on eight randomly selected students per classroom. In this design, students who left the classroom in one year were replaced through random selection by students who entered the study in the next year. As a consequence of this design, neither the leavers nor new students would have complete data across the time points included in the analyses reported here, and this produced sample attrition. In particular, student mobility results in complete data for only 3.9 students per classroom in the first grade sample, largely because mobility is typically high in grades K and 1, and 6.6 students per classroom in the third grade sample. Available data show that first graders who left the study scored 7 points lower on the spring kindergarten Terra Nova compared to those with complete data across both time points; for third graders, the corresponding difference was 6 points. Comparisons also showed that African-American and Asian students left the study at higher rates than other students. However, available data suggest that student attrition was unrelated to teacher scores on the main independent variable of interest here—teachers' content knowledge scores as measured by the survey. For example, in the third grade, the difference in teacher knowledge scores for students who left and those who stayed was not significantly different ($t=.282$, $p > .5$). The lack of relationship between student mobility and teacher knowledge scores suggests that our estimates of teacher knowledge effects on student achievement gains are not subject to much selection bias as a result of student mobility.

Another problem caused by missing data can occur when the standard deviations of key variables are affected by the loss of portions of the student or teacher population. When this is the case, standardized regression coefficients can be biased (although unstandardized coefficients will not be much affected). A comparison of key student-level variables (SES, minority status, gender, initial test score) using pre- and post-attrition samples shows standard deviations vary less than 5% in the case of initial test scores, and only by 1% or less for the other variables. Moreover, only .5% of first grade teachers and 4% of third grade teachers had no complete student data, suggesting that the standard deviations of teacher-level variables will not be much affected by missing data.

Finally, although attrition was more common among students who performed more poorly on the initial pre-test, students with test scores similar to those who left the study nevertheless remain in the sample. As a result, the growth of this "class" of lower-performing students can be accurately estimated, particularly given the independence of the probability of attrition and teachers' content knowledge, the main concern of this paper.

<u>Student background</u>. Several measures of student background were included in this study. The rate of student absence from mathematics instruction was generated by aggregating log reports of daily student absence to the student level. Just over 9 logs were recorded for the average first grader and 8 logs for the average third grader, and the reliability of this aggregated estimate in discriminating among students' rate of absence is .41. Using these data, we created a dummy variable indicating students whose absence rate exceeded 20% (the reference category being students with less than a 20% absence rate). Information on students' gender and minority status was collected from teachers

and other school personnel at the time of student sampling. Information on family socio-economic status was collected via the telephone interview with the parent/legal guardian of the students in the study. The composite variable labeled "SES" represents an average of father's and mother's education level, father's and mother's occupation, and family income.

Teacher background and classroom characteristics. Teacher background variables came primarily from the teacher questionnaire, from which data were used to construct measures of teacher experience, certification, and undergraduate/graduate coursework. These teacher background characteristics were straightforwardly represented in our statistical models. For instance, teachers' experience was reported as the years in service at Year 2 of the study. Although we had information on non-certified teachers' credentials (e.g., provisional or emergency certification), too few teachers existed in each category to include them independently in statistical analyses; thus our credential variable simply reports the presence (1) or absence (0) of certification. Finally, teachers reported the total number of a) mathematics methods and b) mathematics content courses taken as part of their pre-service and post-graduate higher education. Since reports of methods and content courses were highly correlated ($r = .80$) they produced multicolinearity in regression models estimated at both the first and third grades. As a result, we formed a single measure combining reports of mathematics methods and content coursework. Unfortunately, this strategy does not allow for an examination of the independent effects of methods and content courses, as is standard practice in the educational production function literature (e.g., Monk, 1994).

We included three classroom variables in our analyses. First, information on classroom percent minority students was obtained by aggregating student characteristics for each classroom. Second, to capture variation in the absolute amount of mathematics instruction students were exposed to, we developed a measure of the average time spent on mathematics in a classroom using data from teachers' mathematics logs. The time measure excluded days on which the student or teacher was absent. Finally, the rate of teacher absence from mathematics lessons was calculated by aggregating logs to the teacher level.

Content knowledge for teaching. Between five and twelve items designed to measure teachers' content knowledge for teaching mathematics (CKT-M) were included on each of the teacher questionnaires administered over the course of the study. Because this procedure resulted in only a small number of CKT-M items being administered each year, we constructed one overall measure of teachers' content knowledge for mathematics teaching using data from teachers' responses over multiple questionnaire administrations. This strategy increased both the number of CKT-M items on our measure, and the content domains sampled by the measure.

A key feature of our measure is that it represents the knowledge teachers use in classrooms, rather than general mathematical knowledge. To assure that this is the case, we designed measurement tasks that gauged proficiency at providing students mathematical explanations, representations, and working with unusual solution methods. A more detailed description of the work of designing, building, and piloting these measures can be found in Hill, Schilling & Ball, (2004). Aspects of the measures that are critical to interpreting the results of the current study are discussed below.

The overall measurement project began with a specification of the domains of teachers' content knowledge for teaching that we sought to measure. As noted earlier, we limited item-writing to only the three most-often taught mathematical content areas: number, operations, and patterns, functions, and algebra. Next, we decided which aspects of teachers' knowledge to measure within these three topics. On the basis of a review of the research literature, we originally chose to include items in only two major domains within our original framework: content knowledge for teaching and knowledge of students and mathematics. Because piloting revealed items written in this second category did not meet criteria for inclusion in a large and costly study[6], we selected items from only the content knowledge domain to construct the measure described here.

Once the domain map was specified, we invited mathematics educators, mathematicians, professional developers, project staff and former teachers to write items. Writers cast items in multiple-choice format to facilitate the scoring and scaling of large numbers of teacher responses, and produced items that were not ideologically biased – rejecting, for example, items where a "right" answer indicated an orientation to "reform teaching." Finally, writers strove to capture two key elements of content knowledge for teaching – teachers' "common" knowledge of content, or simply the knowledge of the subject a proficient student, banker, or mathematicians would have; and knowledge that is "specialized" to teaching students mathematics.

Two sample items included on the teacher questionnaire illustrate this distinction (Figure 1). In the first, respondents are asked to determine the value of x in $10^X=1$. This is mathematics knowledge teachers use; students learn about exponential notation in the late elementary grades, and teachers must have adequate knowledge to provide instruction on this topic. However, many adults, and certainly all mathematicians would know enough to answer this item correctly – it is "common" content knowledge, not specialized for the work of teaching. Consider, however, another type of item. Here teachers inspect three different approaches to solving a multi-digit multiplication problem – 35 x 25 – and assess whether those approaches would work with any two whole numbers. To respond to this situation, teachers must draw on mathematical knowledge – inspecting the steps shown in each example to determine what was done, then gauging whether or not this constitute a "method," and if so, whether it makes sense and whether it works in general. Appraising nonstandard solution methods is not a common task for adults who do not teach. Yet this task is entirely mathematical — not pedagogical; in order to make sound pedagogical decisions, teachers must be able to size up and evaluate the mathematics of these alternatives — often swiftly, on the spot. Other "specialized" items ask teachers to show or represent numbers or operations using pictures or manipulatives, and to provide explanations for common mathematical rules (e.g., why any number can be divided by 4 if the number formed by the last two digits is divisible by 4).

We believe our measure of teachers' content knowledge bridges the literatures described earlier. It includes the common knowledge often measured within the educational production function literature; however, it also uses lessons from the case study literature on teachers' knowledge to identify and measure the unique skills and capabilities teachers might use in their professional context. By employing this more job-specific measure in the context of an educational production function-type study, we might improve upon prior studies and examine untested assumptions about the relevance of elementary teachers' mathematical knowledge to student achievement.

Following a review of draft items by mathematicians and mathematics educators both internal and external to the project, we piloted items in California's Mathematics Professional Development Institutes (MPDIs). Average reliability for piloted forms ranged in the low .80s with very few misfitting items. Further, specialized factor analyses revealed the presence of a strong general factor in the piloted items (Hill, Schilling & Ball, 2004). Because we had a relatively large pool (roughly 90) of piloted items, we could use information from this pilot to select items for inclusion in the current measure that had shown desirable measurement properties, including a strong relationship to the underlying construct, a range of "difficulty"[7] levels, and a mix of content areas.

As part of these pilots, we also conducted validation work on items by a) subjecting a subset of items to cognitive tracing interviews and b) comparing items to National Council of Teachers of Mathematics (NCTM) Standards, to ensure that we covered the domains specified in that document. Results from the cognitive interviews suggest that in the area of content knowledge, teachers produced few (6.5%) "inconsistent" responses to items, where correct mathematical thinking led to an incorrect answer, or incorrect mathematical thinking led to a correct answer (Dean, Goffney, & Hill, 2004). The content validity check of the entire piloted item set indicated adequate coverage across the number, operations, and patterns, functions, and algebra NCTM standards.

The measure of teachers' content knowledge ultimately used in this analysis includes 30 mathematical knowledge for teaching (CKT-M) items on the year 1 through year 3 teacher questionnaires. We balanced items across content domains (13 number items, 13 operations items, 4 pre-algebra items), and specialized (16 items) and common (14 items) content knowledge. In practice, however, teachers typically answered fewer than 30 items. One reason was that by design, only half the sample answered the first teacher questionnaire. Another reason was that missing data ranges between 5-25% on these items.

We used Item Response Theory (IRT) to handle missing data, create equal-interval scale scores, and provide information about the reliability of our measures. Teachers' responses were scored in a two-parameter IRT model[8] using Bayesian scoring methods. When a teacher failed to answer more than 25% of CKT-M items on a given questionnaire, we scored that teacher's missing items as "not presented," which does not penalize teachers for skipping items. Otherwise, missing data were scored as incorrect. To confirm the findings presented below, we rescored the data using different methods (i.e., Maximum Likelihood) and handled missing data in different ways (e.g., scored all missing data as not presented). Results were robust to these different methods of computing teacher scores. The reliability of the resulting measure was .88. Finally, the CKT-M measure was calculated for the entire teacher sample (first through fifth grade) as a standardized variable (i.e., mean = 0, standard deviation = 1).

In some of the statistical models discussed below, we also included a content knowledge for teaching English Language Arts measure (CKT-ELA). The objective behind designing the CKT-ELA measure was much the same as in mathematics: to attend not just to the knowledge that adults use in everyday life (i.e., reading text), but also to the specialized knowledge teachers use in classrooms (i.e., determining the number of phonemes in a word; assessing a piece of text and determining the best question or task to enhance student understanding). The two major content domains included on this form were knowledge of word analysis – the process of helping students actually read printed text – and knowledge of comprehension. The three major teaching domains included knowledge of the content itself, knowledge of students and content, and knowledge of teaching and content. This last category was not represented in the mathematical work, but includes items focused on ways to enhance student learning of particular pieces of text, remediate student problems with text, and so forth. This CKT-ELA measure was constructed through a similar process to the mathematics measure: item-writing by reading educators, experts, and classroom teachers; piloting in California; factor analyses; choosing items for inclusion on the study's teacher questionnaire that balance across the domain map and maximize desired measurement qualities; and IRT scoring. We here use a measure that combines all of the content and knowledge domains and that has a reliability of .92.   Details on the construction of this measure can be found in Phelps & Schilling (2004) and Phelps (2004).

School characteristics. The one school characteristic employed in this model is household poverty, or the percentage of households in poverty in the neighborhood census tract where schools were located. This measure was constructed from 1990 census data.

Statistical Models and Estimation Procedures

This paper used linear mixed models to estimate the influence of student, teacher, and school characteristics on gains in student achievement. All analyses were conducted using the PROC MIXED procedure in SAS. As described earlier, the main dependent variable was student gain scores over one year of participation in the study. The main advantage of using gain scores over the use of covariate adjustment models that regress pre- on post-test scores is that gain scores are unbiased estimates of students' academic growth (Mullens, Murnane & Willett 1996; Rogosa, Brandt, & Zimowski 1982; Rogosa & Willett 1985). However, gain scores can be subject to unreliability, and as a result, the reader is cautioned that the effects of independent variables on the outcome measure are undoubtedly underestimated (Rowan, Correnti & Miller 2002).

We elected to exclude consideration of a number of factors from our statistical models for simplicity of results and discussion. One such factor was instructional practice, as reported on the daily mathematics log. Another was the mathematics curriculum materials used by each school, including whether the school was using the mathematics program recommended by the school reform program. A third was the improvement program selected by the school. Although each of these is a potentially important influence on student achievement, results from initial models suggested the effects of these factors on gains in student achievement were complex – interactive with student background characteristics, for instance, as well as grade level. Notably, however, participation in a Comprehensive School Reform program had little independent main effect on students' achievement gains, a finding that makes sense given that the programs under study focused mainly on instructional improvement in English Language Arts.

As discussed earlier, there was substantial student attrition and missing data on key variables. First graders without spring-spring data and third graders without fall-fall assessment data were necessarily excluded from the analyses. Also, teachers were excluded from the analysis if they did not return any of the three teacher questionnaires, thus providing no information on their preparation for teaching, years of experience, or content knowledge for teaching mathematics. When teachers did return questionnaires but did not answer enough content knowledge for teaching (CKT) items to reasonably generate a person-level score, we imputed their score. This resulted in roughly 10% of first grade teachers and 20% of third grade teachers with imputed scores using mean imputation. Teachers who did not log their mathematics instruction had their mean mathematics instructional time and absence rate imputed as well. The use of mean imputation is one standard method for dealing with missing cases, but an unfortunate side effect is that the actual covariances between variables are not maintained in the data set. To correct for this problem, we include an indicator (dummy) variable that indexes whether or not a teacher had missing data on a given variable.

In summary, a number of data issues exist in the study, including the small number of students with complete data within each classroom, missing data on many variables, a lack of complete alignment between the teacher and student mathematics assessments, and student attrition. As discussed, the first three problems would tend to bias results conservatively (i.e., against finding positive teacher/classroom effects in our models). For example, the limited number of students per classroom makes it more difficult to reliably discriminate academic growth rates across classrooms, in turn making it more difficult to detect the effects of classroom variables on student achievement gains. The use of mean imputation procedures can reduce the amount of observed covariation between inputs and outcomes, making effects more difficult to detect. And, the lack of perfect alignment across student and teacher assessments produces additional unreliability in the analyses (see Leinhardt & Seewaldt, 1981; Barr & Dreeben, 1983; and

Berliner, 1979 for arguments about overlap). As we have seen, the fourth problem (student attrition) seems neutral with respect to bias, especially since there is little evidence of selection bias in the data.

Results

Table 3 shows pre-standardization sample means and standard deviations for variables included in this analysis. Several of these descriptive statistics have substantive interpretations and implications. As Table 3 shows, the average first grader gained nearly 58 points on the Terra Nova scale, while the average third grader gained 39 points. This is a two-grade snapshot of the often-observed trend toward decelerating academic growth rates in longitudinal studies of student achievement. Other interesting findings are that five percent of first graders and four percent of third graders were reported as absent more than 20% of the time. Finally, roughly 70% of the students in our study sample were non-Asian students of color.

Several teacher-level descriptive statistics also stand out. Because we averaged reports of mathematics methods and content courses, and because teachers report such courses as ranges (e.g., 1-3 courses, 4-6 courses), the measure representing these reports has no easy substantive interpretation. However, it may help the reader to know that 12% of teachers reported never having taken a mathematics content or methods course, 15% reported taking between 1 and 3 such courses, and 27% reported taking between 2 and 6 courses. In many colleges of education, mathematics methods courses are taught by education school faculty, and typically cover the use of manipulatives and other representations for content, problem solving, classroom organization, and designing and teaching math lessons. Mathematics content courses are often taught by a member of the mathematics department and usually cover mathematical topics in the K-6 curriculum — whole numbers and fractions, place value, probability, geometry, combinatorics, and, often, problem solving. Some other required mathematics content courses may be the same as those taken by mathematics majors.

Nearly 90% of the teachers in the sample were certified, and the average teacher was in her twelfth year of teaching. The average teacher reported spending just under an hour per day on mathematics instruction: 55.6 minutes for first graders and 50.3 minutes for third graders. These figures include days on which mathematics was not taught due to an assembly, field trip, test preparation, or similar interruption. Finally, the average teacher reported being absent on 5-6% of logging days, or for roughly 9 days of a 180 day school year. This figure doubtlessly includes professional development days in addition to other absences.

The household poverty variable shows that roughly one in five households in the neighborhoods surrounding schools in this study were below the poverty line. Inclusion of the household poverty variables in these analyses is intended to capture the additional effect of poverty concentration within schools on student achievement, net of students' SES.

Tables 4 and 5 show the correlations among the teacher preparation, experience, and CKT-M variables. The size and strength of these relationships was similar at these two grades, and several relationships stood out. Note first the modest positive correlations of years of teaching experience with certification and with methods and content courses. This is consistent with the observation that teachers continue to take mathematics methods and content courses as they continue in their careers, and that uncertified teachers are also less experienced than certified teachers. In contrast, note that our measures of teachers' mathematical content knowledge for teaching was not significantly correlated with any of the teacher preparation or experience variables at grade 1, and showed only a very small correlation with teacher certification at grade 3. We cannot draw any firm conclusions about causation from these correlations, but this pattern of findings suggests that neither ensuring teacher certification nor increasing teachers' subject-matter or methods coursework (two common approaches to improving teacher quality), ensures a supply of teachers with strong content knowledge for teaching mathematics. Finally, teachers' mathematics CKT and language arts (ELA) CKT measures are correlated, but not as strongly as one might expect: .39 and .37 in the first and third grades, respectively.

Table 6 shows results of an unconditional model that decomposes variance in student gain scores into that which lies among schools, among teachers within schools, and among students within classrooms for the third grade data. The largest amount of variance (85%) lies among students within classrooms. This statistic is in line with findings from other studies, and includes not only the influence of native intelligence, motivation, behavior, personal educational history, and/or family support for educational outcomes, but also variance due to errors in measurement. Given the large amount of variance within classrooms, only a small amount of the remaining variance can lie among teachers – roughly 8% for first grade, and 2% for third grade. Again, this estimate is probably artificially low because of unreliability in measurement of student achievement and student gains in achievement, as well as the small student samples per classroom. To determine whether teacher-level effects can be further modeled, we conducted a likelihood ratio test for variance components; this test rejected the null hypothesis that there was no meaningful variance among teachers. Finally, 6% and 7% of the variance was among schools in the first and third grades, respectively.

Table 7 shows the estimates derived from the two statistical models estimated for first and third grade data. All independent variables were standardized before entry into these analyses, making coefficients easily interpretable as the effect of a one standard deviation increase in each independent variable on gains in students' IRT mathematics scale score over a one year interval. Student level variables, which remained the same in both statistical models, were the strongest predictors of gain scores by this metric. Initial mathematics Terra Nova scores, for example, were strongly and negatively related to gain scores. In other words, students who performed well at the initial assessment, tended to regress to more average performance on the second assessment. Family socio-economic status (SES) was also a strong predictor of gain scores; for every standard deviation increase in socio-economic status, students gained an additional 2 to 4 points. Missing family SES data was not related to student gains at the first grade, but was negatively related to student gains at the third grade, where the proportion of missing SES data was higher. This suggests that families of third graders who did not respond to the phone interview had students who gained less over the course of the year. Female students had nearly two points more in annual growth than did males in the third grade, but there were no gender effects in first grade. Non-Asian minority students had lower gain scores in the first grade and, more marginally ($p=.11$), in the third grade. Students who were absent on more than 20% of days (high absence) also gained less than students with lower absence rates in the third grade model; this effect was close to significant ($p<.10$)

in the first grade model as well. Though these models are not fully enough specified to explore the subtle effects of race, culture, and SES on student achievement, the results are consistent with other research in this arena (Lee & Burkam, 2002; Phillips, Brooks-Gunn, Duncan, Klebanov, & Crane, 1998). Thus, we are satisfied that key student covariates have been captured, thereby allowing the teacher-level modeling we discuss below.

Teachers' content knowledge for teaching mathematics was a significant predictor of student gains in both models at both grade levels. The effect was strongest in Model 1, where students gained an additional two and a quarter points on the Terra Nova for every standard deviation difference in teachers' mathematics content knowledge. Expressed as a fraction of average monthly student growth in mathematics, this translates to roughly 1/2 to 2/3 of a month of additional growth per standard deviation difference on CKT-M. CKT-M was the strongest teacher-level predictor in these models, larger than teacher background variables, and greater than the average time spent on mathematics instruction each day. In third grade, its effect size rivaled that of SES and students' ethnicity and gender while in the first grade models, the effect size is not far off. This suggests that knowledgeable teachers can positively and substantially affect student learning of mathematics, and the size of this effect is, at least in this sample, in league with the effects of student background characteristics.

An important question is whether the effect of teachers' content knowledge on growth in student achievement is linear – that is, whether the gain of slightly over two points per standard deviation of teacher CKT is constant across the range of teacher knowledge. Perhaps only the most knowledgeable teachers deliver highly effective mathematics instruction; alternatively, it may be that only the least knowledgeable teachers have any effect on students' mathematics achievement. To investigate this question, we divided teachers into deciles by their CKT-M score, with the lowest decile (1) representing the least knowledgeable teachers. We replaced the linear measure of CKT-M in Model 1 with this new ten-category demarcation of teachers, and show the results – estimated student gains per CKT-M decile – in Figures 2 and 3. Teachers in the lowest two deciles (0-20%) of the first grade CKT-M distribution taught students who gained, on average, nearly 10 fewer points than students in the highest category, which was the referent. However, above the lowest two deciles there appears little systematic relationship between increases in teacher knowledge and student gains. A statistical test for difference of means (in SAS, the "lsmeans" test) confirmed that significant differences occurred only between the lowest 20% of teachers and other categories. In the third grade data (Figure 3), the significance test suggested that teachers in the first three deciles (0-30%) significantly impacted their students' achievement vis-à-vis the top four deciles. Yet here, the non-linear effect is less pronounced. One possible explanation is the difference in first and third grade content. Perhaps only the very lowest-scoring first grade teachers had difficulties teaching the content at this level, whereas even moderately-scoring third grade teachers may have found the more difficult third-grade content challenging to teach.

Despite success in identifying a positive relationship between mathematical knowledge for teaching and student gain scores, the possibility remains that general knowledge of or aptitude for teaching, not content-specific knowledge for teaching, produced this finding. We have no measure of general knowledge or aptitude for teaching, and therefore cannot directly address this issue. However, we did include in our analyses a measure of content knowledge for teaching English Language Arts that was similar in intent to the CKT-M measure, but designed to measure teachers' knowledge of and ability to teach word analysis and reading comprehension. If the ELA and mathematics measures both draw heavily on general knowledge of teaching, they should be moderately to highly correlated, and should share the positive relationship to student achievement seen in Model 1. In Model 2, we included this CKT-ELA measure and found that although it was positively related to student gains at the first and third grade, it was not statistically significant. Further, it had only a small effect on the absolute size and significance of the CKT-M variable. This suggests that the effect of teachers' knowledge on student achievement is at least content-specific, and in mathematics, reflects more than just general knowledge of teaching.

Our models showed other significant or near-significant findings. For example, the average length of a teachers' mathematics lesson was significantly related to third grade student gains, with a one-standard deviation in daily mathematics lesson length – about 14 minutes – yielding an additional 1.8 points. This translates to roughly an additional 2 weeks of instruction per year for a classroom that receives the additional 14 minutes per day. Teachers' mathematics preparation, that is, the average number of content and methods courses taken in pre-service or graduate training, positively predicted student gains in the third grade, but was just outside of traditional significance ($p = .06$). The effects of another commonly argued policy solution, teacher certification, also were insignificant in this particular sample of teachers and students. Although certification was mildly related to teachers' knowledge of content in the third grade (Table 5), it had no independent influence on student gain scores. This may reflect a true null effect, or occur because non-certified teachers have taken a comparable number of math methods and content courses as certified teachers (see Table 5). Thus non-certified teachers may be en route to traditional certification, or transfers into new schools from other states (Darling-Hammond, Berry & Thoreson, 2001) or mathematics-intensive professions. It could also reflect the fact that certification requirements vary across the states included in the study.

Years of teaching experience, measured linearly, shows no relationship to first grade student achievement, and a marginally significant ($p = .11$) positive relationship in the third grade. Some studies, however, have suggested that it is teachers in the first several years of their career who negatively impact student achievement. We created a dummy variable representing teachers in their first or second years of teaching and entered it into the models in place of the linear measure. The significance of this variable in the third grade model did not change, but this measure of novice teachers did become marginally significant in the first grade model ($b = -5.3$, $p < .10$).

We checked these models in several ways: adding and deleting variables to check for model stability; using pre-on-post models rather than gain score models[9]; creating dummy variables to check for linearity. Overall significance of key variables held firm, and residuals were normally distributed.

Conclusion

The analyses just presented have clear limitations, including the sample of students, missing data, and a lack of alignment between our measure of teachers' mathematical knowledge and student achievement. Because many of these problems would bias the effect size coefficients of our content knowledge for teaching variable toward zero, however, we feel confident that the positive effects we see in our analyses are robust and, if anything, underestimated. However, we are less confident in any borderline or null results, such as those found for the teacher preparation measures. Therefore, we focus our concluding discussion mainly on the effects of the content knowledge variable on students' achievement.

We found that teachers' mathematical knowledge for teaching positively predicted student gains in mathematics achievement during the first and third grade. We were modestly surprised to see this first grade effect, since we had expected the CKT-M measure to have its effects mainly at grades with more complex content – e.g., at grade levels where multidigit addition or multiplication, functions, fractions, and decimals were being taught. That it also positively affects student gains in the first grade suggests that teachers' content knowledge plays a role even in the teaching of very elementary mathematics content. Many kindergarten and first grade teachers explain their choice of grade level by referencing both their love of young children and their lack of mathematics knowledge. Our analyses suggest that mathematical knowledge for teaching is important, however, even at these earliest grade levels.

An important feature of our analyses was that we measured mathematical knowledge for teaching, not just teachers' computational facility or course-taking. Although scholars from John Dewey (1904) to Joseph Schwab (1964) to Lee Shulman have observed that teachers' responsibilities for teaching specific subject matter require special knowledge of the content being taught, the nature of this special knowledge has not been elaborated. Consequently, it has been difficult to measure reliably or validly on a large scale. Our work built on these scholars' theories about relationships of subject matter and pedagogy to design a measure of teachers' mathematical knowledge for teaching, and we can report here that this more task-sensitive measure is positively related to student achievement.

This modifies findings from earlier studies exploring the effect of teachers on student achievement (for summaries, see Begle, 1979; Greenwald, Hedges & Laine 1996; Hanushek, 1996). For one, it confirms Shulman's (1986) important critique of the process-product literature, namely, that studying teacher impact in light of subject-specific behavior is important. Moreover, our findings help envision a new generation of process-product studies designed to answer questions about how teachers' mathematical behavior – in particular their classroom explanations, representations, and interactions with students' mathematical thinking – might affect student outcomes. It also informs findings from the educational production function literature, first by pointing out that a direct measure of teachers' content knowledge for teaching trumps proxy measures such as courses taken or experience, and then by suggesting that measures of teacher knowledge should be at least content-specific, and even better, specific to the knowledge used in teaching children.

Our findings both support and challenge recent policy initiatives. If successful, efforts to improve teachers' mathematical knowledge through content-focused professional development and pre-service programs will work to improve student achievement, as intended. Such programs include California's Mathematics Professional Development Institutes, the National Science Foundation/Department of Education's Math-Science Partnerships, and many other local efforts throughout the U.S. Yet our results suggest that those who may benefit most are teachers in the lowest third of the distribution of knowledge, and that efforts to recruit teachers into professional development and pre-service coursework might focus most heavily on those with weak subject matter knowledge for teaching. However, without ways to differentiate and select such teachers, and without strong incentives for bringing such teachers into content-focused professional development, the intended effects of these major programs may be lost. Moreover, without conceptual and analytic tools for examining whether and what teachers learn from such professional development, efforts to develop the quality and effectiveness of programs designed to improve teaching will be impeded.

Another key question generated by our results concerns equity, namely the intellectual resources available to students across race and socio-economic status (see Cohen, Raudenbush & Ball, 2003 for a discussion of such resources). In the first grade, teachers' mathematical knowledge for teaching in this data set was distributed fairly evenly across students of different socio-economic status, but there was a negative relationship between student minority status ($r = -.16$, $p < .01$) and teachers' mathematical knowledge for teaching. In the third grade, the relationship between student SES and teacher knowledge was significant ($r = .11$, $p < .05$) in this data set and the relationship between minority status and teacher knowledge increased in comparison to first grade ($r = -.26$, $p < .0001$). These results are similar to those found elsewhere with other samples of schools and teachers (Hill & Lubienksi, under review; Loeb & Reininger, 2004). This problem of inequitable distribution of teaching knowledge across different socioeconomic and ethnic groups is particularly pressing if the relationship of teachers' mathematical knowledge to instructional quality is nonlinear, as our analyses suggest. A portion of the achievement gap on the National Assessment of Educational

Progress and other standardized assessments might result from teachers with less mathematical knowledge teaching more disadvantaged students. One strategy toward closing this gap, then, could be investing in the quality of mathematics content knowledge among teachers working in disadvantaged schools. This suggestion is underscored by the comparable effect sizes of teachers' knowledge and students' socio-economic status on achievement.

Three additional lines of inquiry grow naturally from the study presented here. The first calls for examining the effects of mathematics instructional methods and curriculum materials (texts) on student performance. A key component of this analysis will involve examining interactions between teacher knowledge and instructional method/uses of texts. A second line of inquiry should parse more precisely different theoretically- and empirically-grounded distinctions in content knowledge for teaching and investigate their relationships, separately and in combination, to student achievement. The analyses reported here do not make such distinctions, and it is possible that effects may differ across types of knowledge — e.g., common knowledge (CCK), specialized knowledge of content (SKC), as well as knowledge of students and content (KSC) and knowledge of content and teaching (KCT) (see Hill, Schilling & Ball, 2004).

Finally, a third line of inquiry could focus on investigating whether and how the instructional practices of mathematically knowledgeable and less knowledgeable teachers differ. Teachers do not improve student learning simply by scoring well on our multiple-choice assessment. However, what knowledgeable teachers <u>do</u> in classrooms – or how knowing mathematics affects instruction – has yet to be studied and analyzed. Does teachers' knowledge of mathematics affect the decisions they make? Their planning?

How they work with students, or use their textbooks? How they manage student confusions or insights, or how they explain concepts? Previous research on teachers' content knowledge suggests knowledgeable teachers may provide better mathematical explanations, construct better representations, better "hear" student methods and have a clearer understanding of the structures underlying elementary mathematics and how they connect (e.g., Ball, 1993; Borko et al., 1992; Carpenter et al., 1989; Leinhardt & Smith, 1985; Ma, 1999; Thompson & Thompson, 1994). However, analyzing the practice of knowledgeable teachers may also surface new aspects of the mathematical knowledge that matters for teaching: how mathematical and everyday language are bridged, for example, or how representations are deployed, or how numerical examples are selected. Ongoing research on teaching, on students' learning, and on the mathematical demands of high quality instruction can contribute to increasing precision in our knowledge of the role of content knowledge in teaching.

References

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, MI.

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

Ball, D. L. (1991). Teaching mathematics for understanding: What do teachers need to know about subject matter? In M. Kennedy (Ed.), Teaching academic subjects to diverse learners (pp. 63-83). New York: Teachers College Press.

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., Camburn, E., Correnti, R., Phelps, G., & Wallace, R. (1999). New tools for research on instruction: A web-based teacher log. Working paper, Center for Teaching Policy. Seattle, WA: University of Washington.

Barr, R. & Dreeben, R. (1983) How schools work. Chicago: University of Chicago Press.

Begle, E. G. (1972). Teacher knowledge and student achievement in algebra (SMSG Rep. No. 9). Palo Alto, CA: Stanford University.

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

Benson, G. (2002) Study of Instructional Improvement school sample design. University of Michigan, Ann Arbor: Institute for Social Research.

Berliner, D. (1979) Tempus Educare. In P. Peterson and H. Walberg (Eds.), Research on teaching: Concepts, findings, and implications (pp. 120-135). Berkeley, CA.: McCutchan.

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

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.

Brophy, J. & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), Handbook of Research on Teaching, Third Edition (pp. 328-375). New York: Simon & Schuster McMillan.

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

Carnoy, M. & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. Educational Evaluation & Policy Analysis, 24, p. 305-31.

Carpenter, T. P., Fennema, E., Peterson, P.L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. American Educational Research Journal, 26, 499-531.

Cohen, D.K., Raudenbush, S. & Ball, D.L. (2003) Resources, instruction and research. Educational Evaluation & Policy Analysis, 25, 119-42.

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. Educational Evaluation and Policy Analysis, 23, 57-77.

Dean, C. A. & Goffney, I. M. & Hill, H.C. (2005). Technical report on the validity of CKT-M items. Ann Arbor, Michigan: University of Michigan.

Dewey, J. (1902). The child and the curriculum. Chicago: University of Chicago Press.

Doyle, W. (1977). Paradigns for research on teacher effectiveness. Review of Research in Education, 5, 163-198.

Ferguson, R.F. (1991). Paying for public education: New evidence on how and why money matters. Harvard Journal on Legislation, 28, 458-498.

Gage, N. (1978). The scientific basis of the art of teaching. New York: Teachers College Press.

Greenwald, R., Hedges, L.V., & Laine, R.D. (1996). The effect of school resources on student achievement. Review of Educational Research, 6, 361-396.

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

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

Hanushek, E.A. (1981). Throwing money at schools. Journal of Policy Analysis and Management, 1, 19-41.

Hanushek, E.A. (1996). A more complete picture of school resource policies. Review of Educational Research, 66, 397-409.

Harbison, R.W. & Hanushek, E.A. (1992). Educational performance for the poor: Lessons from rural northeast Brazil. Oxford, England: Oxford University Press.

Hill, H.C., Schilling, S.G., & Ball, D.L. (2004). Developing Measures of Teachers' Mathematics Knowledge for Teaching. Elementary School Journal, 105, 11-30.

Hill, H.C. & Lubienski, S.T. (in progress). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. Ann Arbor, MI: University of Michigan.

Lee, V. & Burkham, D. (2002). Inequality at the starting gate: Social background difference in achievement as children begin school. Washington, DC: Economic Policy Institute.

Leinhardt , G., & Seewaldt, A.M. (1981). Overlap: What's tested, what's taught. Journal of Educational Measurement, 18, 85-95.

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

Loeb, S. & Reininger, M. (2004). Public policy and teacher labor markets: What we know and why it matters. East Lansing, MI: The 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.

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

Mullens, J. E., Murnane, R. J., Willett, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: a multilevel analysis of longitudinal evidence from Belize. Comparative Education Review, 40, 139-57.

Phelps, G. & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. Elementary School Journal, 105, 31-48.

Phelps, G.C. (2004). Content knowledge for teaching reading: Technical report on reading measures used on the Study of Instructional Improvement teacher questionnaire. Consortium for Policy Research in Education, Study of Instructional Improvement, Research Report. Ann Arbor, MI: University of Michigan.

Phillips, M., Brooks-Gunn, J., Duncan, G.J., Klebanov, P., & Crane, J. (1998). Family background parenting practices, and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), The Black-White test score gap (pp. 103-145). Washington D.C.: Brookings.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rogosa, D.R. & Willett, J.B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50, 203-228.

Rowan, B., Chiang, F., & Miller, R.J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. Sociology of Education, 70, 256-284.

Rowan, B., Correnti, R. & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the *Prospects* study of Elementary Schools. Teachers College Record, 104, 1525-1567.

Rowan, B., Harrison, D., & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. Elementary School Journal, 105, 103-127.

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, 4-14.

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

Strauss, R.P. & Sawyer, E.A. (1986). Some new evidence on teacher and student competencies. Economics of Education Review, 5, 41-48.

Summers, A.A. & Wolfe, B.L. (1977). Do schools make a difference? American Economic Review, 67, 639-652.

Tatto, M.T., Nielsen, H.D., Cummings, W., Kularatna, N.G. & Dharmadasa, K.H. (1993). Comparing the effectiveness and costs of different approaches for educating primary school teachers in Sri Lanka. Teaching and Teacher Education, 9, 41-64.

Thompson, P. & Thompson, A. (1994). Talking about rates conceptually, Part I: A teachers' struggle. Journal for Research in Mathematics Education, 25, 279-303.

Wayne, A.J. & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. Review of Educational Research, 73, 89-122.

Wilson, S. M. & Wineburg, S. (1988). Peering at history through different lenses: The role of disciplinary perspectives in teaching history. Teachers College Record, 89, 525-39.

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, England: Holt, Rinehart & Winston.

Notes

The authors thank Robert J. Miller, Geoffrey Phelps, Stephen G. Schilling and Kathy Welch for their assistance. Errors remain the property of the authors. The research reported in this paper was supported in part by grants from the U.S. Department of Education to the Consortium for Policy Research in Education (CPRE) at the University of Pennsylvania (Grant # OERI-R308A60003) and the Center for the Study of Teaching and Policy at the University of Washington (Grant # OERI-R308B70003); the National Science Foundation's Interagency Educational Research Initiative to the University of Michigan (Grant #s REC-9979863 & REC-0129421), The William and Flora Hewlett Foundation and The Atlantic Philanthropies. Opinions expressed in this paper are those of the authors, and do not reflect the views of the U.S. Department of Education, the National Science Foundation, the William and Flora Hewlett Foundation or the Atlantic Philanthropies.

[1] The BNSE measures student proficiency at age 14, the equivalent in the U.S. of an end-of-eighth-grade exam.

[2] We note, too, that "pedagogical content knowledge" itself has yet to be precisely defined and mapped. See Ball, (1988), Shulman (1986), Shulman (1987), Grossman (1990) and Wilson & Wineberg (1988) for different potential organizations of this knowledge.

[3] The sampling technique used conditioned school selection on geographic location, year of entry into CSR program, and an index of community disadvantage. The last ensured comparable schools within each CSR program. For more detail on the sampling process, see Benson (2002).

[4] This table does not compare the exact set of students in our analyses to ECLS students, but uses instead all students who were in kindergarten at the time our study began. Many, but not all, of these students appear in the first grade cohort reported upon here. Students

leaving the study were also replaced by randomly sampled new students, whose information was not included in Table 1.

[5] We are grateful to the schools, teachers, and students participating in this study for allowing the collection of this data.

[6] Briefly, many of these items misfit in item response theory models; factor analyses indicated multidimensionality, as some items drew on mathematics knowledge, some on knowledge of students, and some on both jointly; as a set, they were also too "easy" for the average teacher; cognitive tracing interviews suggested teachers' multiple-choice selections did not always match their underlying thinking. All four problems resulted in our projecting low reliabilities for the number of items that could be carried on the SII TQ. We are continuing to develop theory and measures in an effort to address these results.

[7] "Difficulty" describes the relationship among items, differentiating between those that are easier for the population of teachers as opposed to those that are more difficult. Here, item difficulty is used to ensure that the SII assessment had both easier items – which would allow differentiation among lower-knowledge teachers – and harder items, which would allow the differentiation among higher-performing teachers.

[8] Two-parameter models take into account both the difficulty of an item and the correctness of a response in scoring. Two teachers who both answer 4/5 items correctly, for instance, may have different scores if one correctly answered more difficult items than the other. Missing data in this sample makes 2-parameter models attractive because of this feature. Results in Table 7 were similar with the 1-parameter scoring method.

[9] Results from covariate adjustment models are similar to those obtained with gains models; the content knowledge for teaching mathematics effect size and significance does not change in the first grade model, and increases in the third grade model.

Table 1

SII vs. ECLS students

	SII (n=1,616)	ECLS (n=21,116)
Household income		
UNDER $5,000	3.5%	3.3%
$5,000 - $9,999	8.3%	4.2%
$10,000 - $14,999	9.8%	7.7%
$15,000 - $19,999	9.5%	6.8%
$20,000 - $24,999	9.0%	7.9%
$25,000 - $29,999	8.4%	6.4%
$30,000 - $34,999	7.6%	7.0%
$35,000 - $39,999	6.6%	5.6%
$40,000 - $49,999	9.1%	10.3%
$50,000 - $74,999	18.9%	20.0%
$75,000 - $99,999	5.6%	9.5%
$100,000 - $199,999	4.3%	8.8%
$200,000 or more	3.3%	1.95%
Mother's educational background	n=1,840	n=19,809

Did not complete high school	18.3%	14.3%
High school diploma or equivalent	34.7%	30.6%
Some college or vocational school	34.9%	31.7%
Bachelor's degree	9.3%	14.6%
Master's degree or attended professional school	2.4%	5.9%
Ph.D. or other advanced degree	0.4%	1.4%
Father's educational background	n=1,205	n=16,066
Did not complete high school	14.4%	11.2%
High school diploma or equivalent	34.1%	26.0%
Some college or vocational school	29.1%	20.8%
Bachelor's degree	12.1%	13.0%
Master's degree or attended professional school	4.3%	5.7%
Ph.D. or other advanced degree	1.0%	3.2%
Family structure	n=1,900	n=18,962
Biological mother/father present in household	52.2%	63.8%
Parent with stepparent or partner	6.5%	9.5%
Single parent	40.7%	22.6%
Student race	n=2,130	n=21,190
White	26.2%	57.0%
Black	47.5%	16.4%

Hispanic	14.3%	18.9%
American Indian/Alaskan Native	0.5%	1.8%
Asian or Pacific Islander	5.1%	3.4%
Hispanic	16.4%	20.6%
Other	4.3%	2.4%

Table 2

Instrument response rates

	2000-2001		2001-2002		2002-2003	
	sample / completed rate	Pct.	sample / completed rate	Pct.	sample / completed rate	Pct.
Self-Administered Questionnaires						
Teacher Questionnaire (TQ)	2874/1806	69%	4043/2969	73%	3751/2861	76%
Teacher Logs						
Teacher sample-math log	178/172	97%	570/519	91%	n/a[b]	
Completed logs –filtered [a]	9025/8216	91%	31414/28560	91%	n/a	
Parent Interview						
Parent Questionnaire (PQ)	2343/1999	85%	3777/2877	76%	n/a	
Student Instruments						
Terra Nova (TN) - Fall	1289/1247	97%	3845/3690	96%	4868/4638	95%
Terra Nova (TN) - Spring	2313/2220	96%	5080/4897	96%	4743/4595	97%

[a] Log samples filtered by teacher refusal, student move-out, student ineligible, and parental refusal.

[b] n/a indicates data from this year not used in this paper

Table 3

Sample means and standard deviations

		Grade 1			Grade 3		
Label	Description	Mean	SD	N	Mean	SD	N
Student variables							
Average gain	Spring K – Spring 1st Fall 3rd– Fall 4th	57.6	34.6	1190	39.4	33.1	1773
Initial math score	Initial math Terra Nova score	466.6	41.5	1190	563.7	36.2	1773
SES	Family socio-economic status	-.01	.74	1190	-.05	.66	1773
SES missing	No data on family socio-economic status	.07	.26	1190	.23	.42	1773
High absence	Marked 1 if student's absence rate exceed 20%	.05	.22	1190	.04	.09	1773
Female	Marked 1 if student is female	.51	.50	1190	.53	.50	1773
Minority	Marked 1 if student is non-Asian minority	.68	.47	1190	.70	.46	1773

<u>Teacher/classroom variables</u>							
Math methods & content	Math methods and content courses taken	2.56	.95	334	2.50	.91	365
Certified	Marked 1 if teacher is certified	.89	.31	334	.90	.25	365
Years experience	Years experience reported in Year 2 of study	12.21	9.53	334	12.85	9.45	365
CKT-Mathematics	Content knowledge for teaching mathematics	.03	.97	334	.05	.89	365
CKT-M Missing	Missing content knowledge for teaching mathematics	.09	.29	334	.19	.39	365
CKT-ELA	Content knowledge for teaching English Language Arts	.14	.74	334	.07	.64	365
CKT-ELA Missing	Missing content knowledge for teaching English Language Arts	.08	.27	334	.18	.38	365
Math lesson length	Average length in minutes of mathematics class	55.6	13.4	334	50.3	14.4	365
Teacher absence rate	Percent of logs on which teacher reports own absence	.05	.22	334	.06	.05	365

Log data missing	No information on math lesson length, teacher absence or student absence	.06	.24	334	.10	.31	365
Pct class minority	Percent minority in a classroom, initial time point	.47	.32	334	.64	.35	365
<u>School-level variables</u>							
Household poverty	Percent of households in poverty	.18	.13	115	.19	.14	115

Table 4

Correlations between teacher preparation, experience, and mathematical knowledge for teaching – 1st grade

	Math methods & content	Certified	Years experience	CKT-Math	CKT-ELA
Math methods & content	1.0	.10	.18*	.00	-.07
Certified		1.0	.20**	.07	.04
Years experience			1.0	.00	.01
CKT-Mathematics				1.0	.39**
CKT-ELA					1.0

* Significant at $p < .05$

** Significant at $p < .001$

Table 5

Correlations between teacher preparation, experience, and mathematical knowledge for teaching – 3st grade

	Math methods & content	Certified	Years experience	CKT-Math	CKT-ELA
Math methods & content	1.0	.03	.19**	-.08	-.05
Certified		1.0	.15*	.11*	.02
Years experience			1.0	-.09	.05
CKT-Mathematics				1.0	.37**
CKT-ELA					1.0

* Significant at $p < .05$

** Significant at $p < .001$

Table 6

Variance Components

	Grade 1	Grade 3
Teachers	99.2	24.4
Schools	77.4	79.3
Residual	1028.3	990.27
Total	1204.9	1093.97
AIC	11774.8	17386.3

Table 7

Student gain score models

	Model 1		Model 2	
	Grade 1	Grade 3	Grade 1	Grade 3
Intercept	57.6	39.3	57.6	39.3
	(1.31)	(.97)	(1.31)	(.97)
Student variables				
Initial math score	-19.5***	-17.0***	-19.5***	-17.1***
	(1.21)	(.84)	(1.21)	(.84)
SES	3.96***	2.13**	3.95***	2.12**
	(.94)	(.76)	(.94)	(.76)
SES missing	.15	-1.80*	.15	-1.80*
	(.72)	(.73)	(.73)	(.73)
Female	-.55	1.80**	-.56	1.79**
	(.87)	(.70)	(.87)	(.69)
Minority	-4.15**	-1.86	-4.14***	-1.84

	(1.43)	(1.15)	(1.43)	(1.15)
High absence	-1.51	-.74*	-1.51	-.74*
	(.88)	(.38)	(.88)	(.38)
<u>Teacher/classroom variables</u>				
Math methods & content	.53	1.64	.55	1.70
	(1.00)	(.92)	(1.01)	(.92)
Certified	.23	-.34	.24	-.33
	(.89)	(.73)	(.90)	(.72)
Years experience	.72	1.02	.72	.95
	(1.14)	(.64)	(1.15)	(.66)
Background variables missing	-.22	-.61	-.21	-.57
	(.96)	(.81)	(.95)	(.80)
CKT-Mathematics	2.22*	2.28**	2.12*	1.96**
	(.91)	(.75)	(1.00)	(.77)
CKT-ELA			.26	.82

			(1.18)	(.87)
CKT missing	-.64	-.31	-.64	-.22
	(1.25)	(1.00)	(1.27)	(1.02)
Math lesson length	-.11	1.77*	-.11	1.82*
	(1.04)	(.87)	(1.05)	(.88)
Teacher absence	-1.01	-.37	-1.00	-.36
rate	(.92)	(.88)	(.94)	(.88)
Log data missing	-1.80*	.75	-1.81*	.70
	(.91)	(.81)	(.91)	(.83)
Pct class minority	2.29	-2.22	2.34	-2.20
	(1.37)	(1.28)	(1.41)	(1.28)
School-level				
variables				
Household poverty	-1.60	-1.59	-1.60	-1.64
	(1.33)	(1.02)	(1.33)	(1.02)
Variance				
components				

Teacher	80.63	13.8	84.6	14.7
School	82.40	53.2	79.88	52.6
Residual	730.89	774.5	730.65	774.11
AIC	11342.4	16836.0	11340.1	16833.6

* Significant at $p < .05$

** Significant at $p < .01$

*** Significant at $p < .001$

Figure 1

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 x 25 125 +75 875	35 x25 175 +700 875	35 x 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

Figure 2

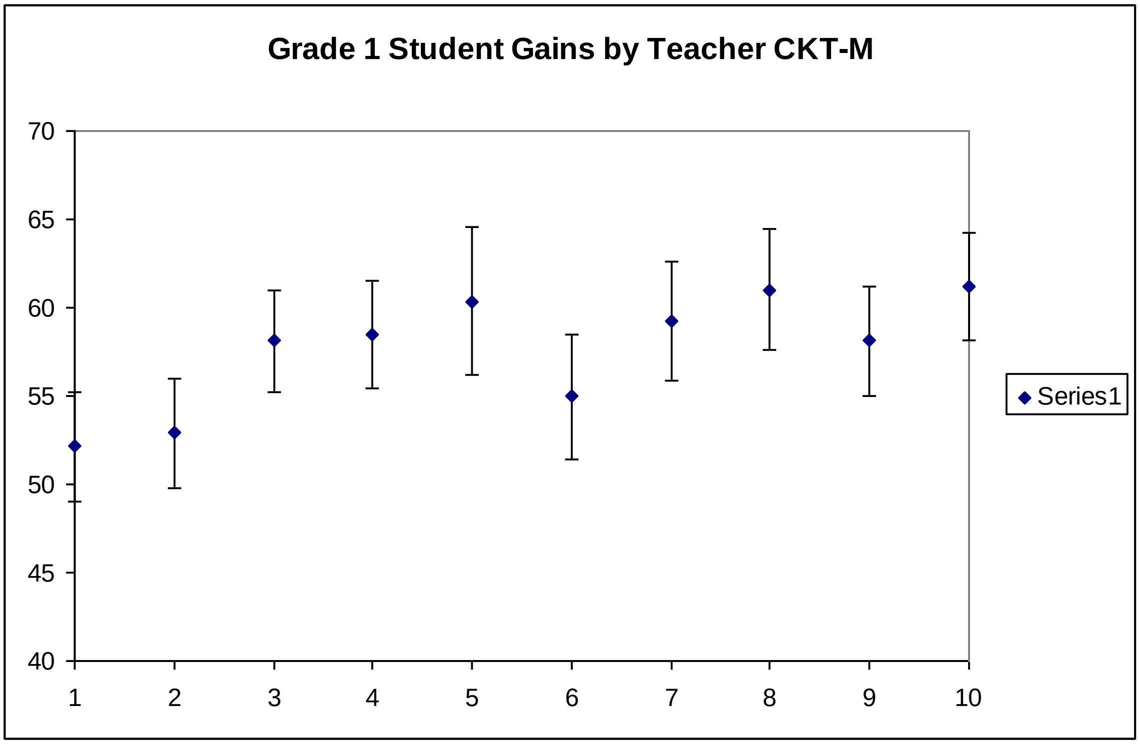
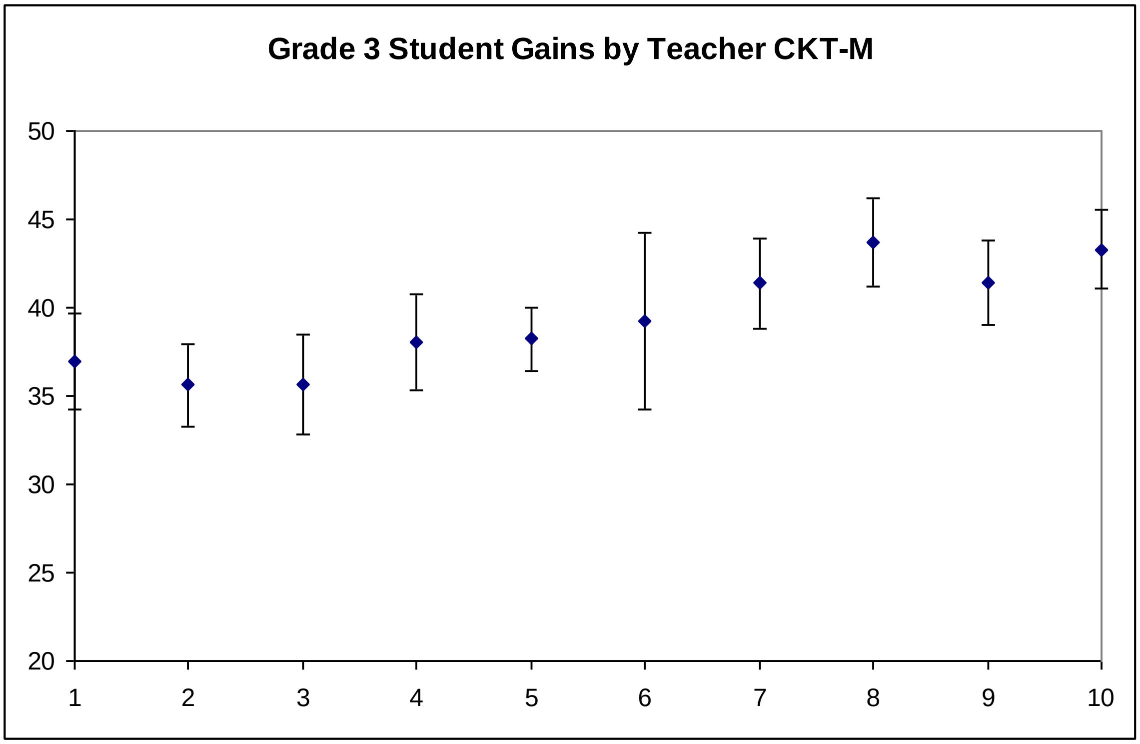

Grade 1 Student Gains by Teacher CKT-M
70
65
60
55
50
45
40
1
2
3
4
5
6
7
8
9
10
Series1

Figure 3


Grade 3 Student Gains by Teacher CKT-M
50
45
40
35
30
25
20
1
2
3
4
5
6
7
8
9
10