

Effects of Teacher Qualifications, Practices, and Content on 4th and 5th Grade Mathematics Achievement in High- and Low-Poverty Classes¹

Robert G. Croninger
University of Maryland

John C. Larson
University of Maryland

Clare E. VonSecker
Montgomery County Public Schools

April 4, 2006

DRAFT: PLEASE DO NOT CITE OR QUOTE WITHOUT AUTHORS' PERMISSION

¹ This work was supported by a grant from the Interagency Educational Research Initiative (IERI # 0115389), a combined effort of the National Science Foundation, the U.S. Department of Education, and the National Institutes of Health. The opinions expressed in this manuscript are our own and do not reflect the positions and policy of the National Science Foundation, the U.S. Department of Education, or the National Institutes of Health. Please direct questions about the manuscript to Robert G. Croninger at croninge@umd.edu.

Abstract

This paper examines the effects of teacher qualification, teaching practices and lesson content on mathematics achievement. We draw on our own work and the work of colleagues engaged in a longitudinal study of high-quality teaching in 4th and 5th grade classes. We use data collected during the 2004-05 school year, including student academic records and roughly 500 observed mathematics lessons in 66 mathematics classes. Multilevel models of mathematics achievement indicate no effects for teacher qualifications after controlling for students' prior achievement but effects for teaching practices and content. Moreover, results suggest that students in high-poverty classrooms benefit most from teaching practices that promote greater cognitive demand and the management of instructional activities, while students in low-poverty classrooms benefit most from lesson content that encourages students to link mathematics concepts and procedures. We conclude with a discussion of these results and possible implications for policy and future research.

Background

Through a range and mix of “policy tools” (Floden, 2003), policymakers have sought – sometimes successfully, sometimes unsuccessfully – to increase student achievement, ensure efficiency in the operation of schools, and address inequities in the distribution of educational goods (Hannaway & Woodroffe, 2003). The recent reauthorization of the Elementary and Secondary Education Act, No Child Left Behind, provides such an example. It sets unprecedented goals for public education, including guarantees that all students attending Title I schools be taught by a “highly-qualified” teacher by the end of the 2005-06 school year (NCLB, 2001, Part A, Section 1119). Such a mandate is justified by research that portrays “teacher effects” as one of the more powerful in-school factors contributing to student learning (Sanders & Horn, 1998; Sanders & Rivers, 1996; Rice, 2003), as well as a group of studies that portrays these effects as especially important for students from historically disadvantaged populations (Darling-Hammond, 1997, 2000; Goldhaber & Anthony, 2003). Proponents of this line of research argue that improving teaching is possible (Ball & Cohen, 1999; Cohen & Hill, 2000; Darling-Hammond, 2000) and that teachers might make more of a difference than structurally-focused educational reforms (Sanders & Horn, 1998; Hawley & Rosenholtz, 1984; Sanders & Rivers, 1996).

While research provides some evidence that promoting teacher excellence can be an important policy tool in improving student achievement, it does not provide much guidance about how education leaders and policymakers can do so or what exactly constitutes “quality” in teaching. Teacher qualifications, such as experience, certification status in content areas, and professional coursework, are one set of proposed “quality” indicators. Although such teacher characteristics provide convenient indicators for administrative hiring and placement decisions

(Croninger, Rice, Rathbun, & Nishio, in press; Rice, 2003), examinations of the relationship between teacher qualifications and student achievement are mixed and inconclusive, especially at the elementary-grade level.² Studies that have documented a positive effect of teacher degree level on elementary student achievement (e.g., Ferguson & Ladd, 1996) have been overshadowed by studies that have found either no discernable effect (Murnane, & Phillips, 1981) or even negative effects of elementary teachers holding an advanced degree on their students' achievement (Croninger et al, in press; Murnane, 1975; Rowan, Correnti, & Kiesling, 2002). There is also little evidence that certification status, perhaps the most widely advocated indicator of quality, affects student achievement in the elementary grades (Croninger et al., in press; Rowan et al., 2002), though some studies have found a link between coursework preparation in a specific subject area and student achievement (Croninger et al., in press).

Another perspective on quality examines teacher practices to determine what teachers do to enhance student learning (see Brophy & Good, 1986; Shulman, 1986; Floden, 2001 for reviews of this literature). Early studies focused on discrete behaviors, such as the use of homework as an instructional tool, how to effectively provide feedback to students or how to manage time and organize lessons; more recent studies have focused on how teachers mediate student learning by the manner in which they understand and portray subject matter to students (Hill, Rowan, & Ball, 2005), engage students in thinking more deeply about topics (Knapp, 1995; Smith, Desimone, & Ueno, 2005), link ideas to students' prior knowledge or schemas, motivate students through praise and feedback and create learning environments conducive to high expectations for achievement (American Psychological Association Board of Educational

² An exception is the growing consensus that subject-specific education, certification and advanced degrees make a difference at the secondary level of schools. See Goldhaber and Anthony (2003); Goldhaber & Brewer (1997a,b, 1998, 2000) and Rowan, Chiang, and Miller (1997) for examples and reviews of this research. There is no comparable consensus, however, about what matters among elementary school teachers.

Affairs, 1995; Alexander, & Murphy, 2000). Although research into teacher practices has identified generic practices and provided broad principles by which to guide teaching, less is known about how these principles should be used to mediate instruction for specific content and contexts (Brophy & Good, 1986; Floden, 2001; Shulman, 1986; Stodolsky, 1988).

Yet a third perspective on quality urges researchers and policymakers to consider the content of lessons posed to students. Research from this perspective notes that irrespective of teacher qualifications or practices students are unlikely to master content unless it is presented to them (e.g., Gamoran, 2003; Porter 2002; Porter, Floden, Freeman, Schmidt, & Schwille, 1986). Surveys have suggested that teachers and schools vary widely in the mathematics curriculum that they present to students, especially in the United States, where curriculum is determined locally (Porter, 1989; Stevenson & Baker, 1991). Moreover, studies suggest that even in the same school there is substantial variability between classes in the manner in which elementary-school teachers structure, order and present mathematics content to students (Rowan, Harrison & Hayes, 2004). Some researchers have suggested that aligning content coverage with instructional goals may be one of the more powerful policy tools available to policymakers interested in ensuring quality teaching (Gamoran, 2003; Porter, 2002). While alignment itself is a worthy aim for policy, whether it produces quality instruction depends in part on the nature of the goals that educators and policymakers have for student learning (Hiebert et al., 2005).

The present study examines what constitutes quality from all three perspectives. We rely on classroom observations and teacher self reports of their qualifications to examine teaching-related factors associated with mathematics achievement in 4th and 5th grade classes. Observation data utilized in this paper focus on the measurement of classroom instructional practices and processes, student responses to teacher practices and processes, and the content of

observed lessons during the 2004-05 school year. Drawing on the work of others who have examined poverty as an important context for instruction (e.g., Knapp, 1995), we investigate whether teacher qualifications, practices, and content have similar or different effects on mathematics achievement in low-poverty and high-poverty classes. We focus specifically on examples of practices and content consistent with current efforts to reform mathematics instruction, especially reform efforts meant to foster greater cognitive engagement with mathematics by elementary-school students (NCTM, 2000).

Methods

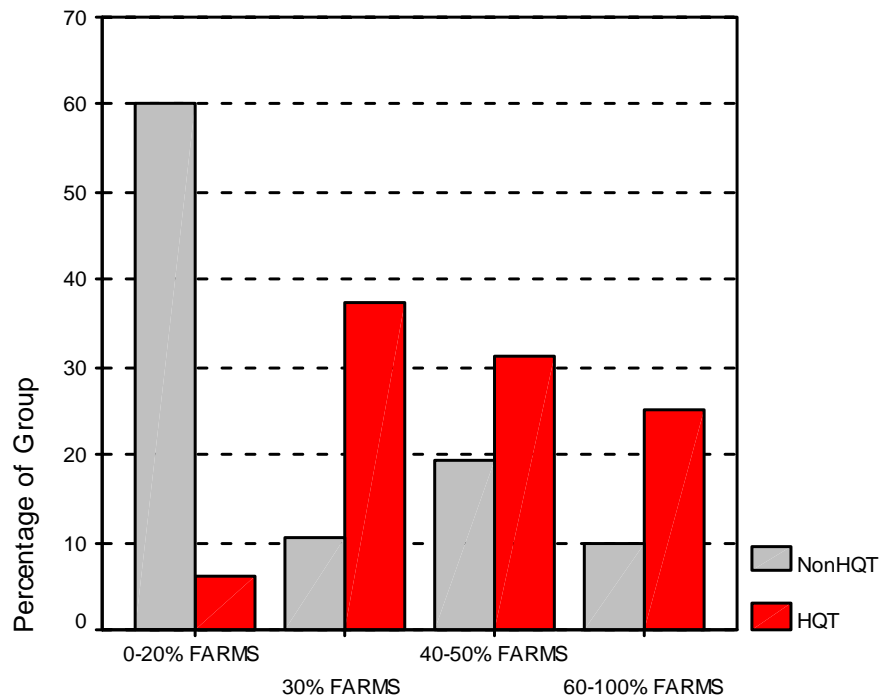
Data for this study come from a four-year study of high-quality teaching (HQT) funded by an Interagency Education Research Initiative grant, a joint program of the U.S. Department of Education, the National Institutes of Health, and the National Science Foundation. The purpose of the study is to investigate what teachers do to help 4th and 5th grade students succeed in reading and mathematics, as well as to investigate how various educational policies and organizational factors influence the ability of teachers to scale up and sustain good teaching practices over time.

Study Population

Data for this study come from classroom observations conducted during the 2004-05 school year. During that year 16 elementary schools, all from a single, large metropolitan school district, participated in the study. The schools in the study are among 125 elementary schools within a district that enrolls nearly 140,000 students (K-12), making the district one of the 20 largest school districts in the country. Roughly one-quarter of students in the district participate

in a free-and-reduced-price meals service (FARMS); however, for the purposes of this study, we recruited schools with higher levels of poverty. Nearly two-thirds of the schools in the study have FARMS concentrations of over 40%, and slightly over half of the students (58%) in the study schools are either African American or Hispanic. See Figure 1 for a comparison of poverty enrollments at participated schools (i.e., HQT schools) and elementary schools not included in the study.

Figure 1 FARMS enrollments at participating (HQT) and non-participating schools.



Beside the focus on schools with moderate to higher levels of poverty, we sought to recruit schools where we would have an opportunity to observe high-quality teaching in 4th and 5th grade classes. During the initial year of the study, we used district achievement data to identify moderate- to high-poverty schools with higher than expected achievement gains in one

or more subject areas and smaller than expected achievement gaps between FARMS and non-FARMS students. When a school agreed to participate in the study, we then sought to recruit all of the 4th and 5th grade teachers in the school to participate in the study, particularly those teachers who taught mathematics and reading in standard classes. Participation was completely voluntary and some schools and teachers declined to participate in the study. When teachers or schools declined to participate or withdrew from the study, we attempted to replace them with similar teachers or schools in the district.

During the 2004-05 school year, a total of 73 teachers participated in the study. Of these teachers, 63 participated in the mathematics observation component of the study and provided us with the student rosters for one of their typical mathematics classes.³ We observed roughly eight mathematics lessons in each class during the year. The typical duration of an observation was 60 minutes. Along with the observation data we collected copies of handouts and instructional materials, as well as attempted to clarify with teachers their personal expectations (before an observation) and assessments of a lesson (after an observation). At the end of the year, the district provided us with the achievement and academic records of students based on the class rosters. After limiting the data to 4th and 5th grade students who had data for the state-required mathematics' assessment in 2004 and 2005, we had a total of 1,074 students, taught by 63 different teachers, in 66 different classes.⁴

Observation Instruments

We developed a series of data collection instruments for the high-quality teaching study, including a time-sampling observation protocol for observing mathematics instruction in the 4th

³ Three teachers provided us with information about two of their mathematics classes, giving us a total of 66 classes for the year.

and 5th grades.⁵ (See Brophy & Good, 1986, and Good & Brophy, 2003, for a discussion of the use of classroom observations to study teaching.) We based the protocol on the district's formal curriculum (3rd through 6th grades), instructional standards promoted by the National Council of Teachers of Mathematics (2000), the research literature on high-quality teaching in and mathematics, and lessons learned from the initial piloting of the instrument. Every three minutes observers respond to a set of questions presented to them on an automated screen on a laptop computer. Questions ask observers to record what the teacher is doing, what students are doing, current content and context for the lesson, classroom organization, as well as the use of technology, specific materials, and student attentiveness. Each observation created roughly 20 data collection episodes (60 minutes divided by 3). All observers participated in pre-observation training to ensure interreliability; observers were also periodically paired with a second observer during the year to monitor interrater reliability and identify potential coding problems.

Measures

Although the study provides a broad range of information about teaching, the context for teaching, and student achievement, we focus on a smaller set of data to explore the relationship between teacher qualifications, practice, content, and achievement in low-poverty and high-poverty classes. We describe briefly the central measures used in this study. (See the Appendix for a fuller description of measures relevant to the analysis and paper.)

Student achievement. We use the overall scale scores on a state-mandated mathematics assessment to measure achievement. The assessment is linked to the state's curriculum and includes a set of items that permit national comparisons of student performance. Mathematical

⁴ Roughly three-quarters of the students (78%) included on the class rosters had achievement records for both years.

knowledge covered by the assessment includes algebra, geometry, statistics and probability, number sense, and mathematical procedures. We use the students' 2005 assessment score as the dependent variable and their 2004 score to control for prior achievement.

Teacher qualification. The first year that teachers participate in the study they complete a brief questionnaire about their preparation, degrees, and years of experience. We used this information to create two measures of teacher qualifications: whether teachers had an advanced degree (MA/MEd) or not and years of experience in 2004-05. Nearly two-thirds of the teachers (65.2%) have an advanced degree; the average years of experience teaching is 9.

Teacher practices. We used data from the observation protocol to create measures of practices thought to promote deeper engagement in mathematics content and greater focus on learning as an indicator of quality (Knapp, 1995; NCTM, 2000; Smith, Desimone, & Ueno, 2005). We calculated the average percentage of time across observations that teachers made higher cognitive demands on students (e.g., requested an alternative answer or posed a higher-order problem or task to students) or managed an instructional activity (e.g., used manipulatives to demonstrate a mathematical concept). Observers witnessed these practices, on average, 17% of the time.

We also calculated the average percentage of time across observations that teachers made lower cognitive demands on students (e.g., accepted simple answers from students or posed a lower-order problem or routine task to students) or managed materials (e.g., handed out or collected papers) or student behavior (e.g., discipline). Observers witnessed these practices, on average, 32% of the time. We combined these two measures to estimate the ratio of "higher" to "lower" practices observed across lessons in each mathematics class.

⁵ A time sampling protocol was developed for reading, as well. This paper, however, only makes use of data gathered using the mathematics protocol. We plan to expand these analyses to include reading at a future date.

Lesson content. We used a similar logic to construct a measure of lesson content thought to promote a deeper understanding of mathematics as an indicator of quality – namely, the amount of time that lessons focused on linking students’ procedural and conceptual knowledge (Hiebert et al, 2005; Hiebert & LeFevre, 1986; NCTM, 2000). On average, 6% of the lesson time that we observed involved some form of linking content, whereas 25% involved conceptual content and 45% involved procedural content.⁶ We then combined these three measures to estimate the ratio of linking content to the sum of conceptual and procedural content.

Analytic Models and Estimation

Because the effects of teacher qualifications, practices and content on student mathematics achievement are nested within classes, we use multilevel modeling in our analysis. (See Raudenbush & Bryk, 2002 for a full description of hierarchical linear models.) We use a two-level random intercept model to examine the effects of our various indicators of quality on 4th and 5th graders achievement in low-poverty and high-poverty classes.⁷ The model partitions variance in the outcome Y_{ij} into two components: variance between students within a mathematics class (level 1) and variance between classes in average mathematics achievement (level 2). At level 1, we model student i in class j as a function of a vector of 4th and 5th graders’ characteristics (X_{ij}) and random student error (r_{ij}):

$$Y_{ij} = \beta_{0j} + \beta_{qj}X_{qij} + r_{ij},$$

where

⁶ Percentages do not add to 100% because observers were not always able to distinguish content, teachers were managing non-instructional materials, or no instruction occurred during a specific episode.

⁷ We considered fitting a three-level model to these data (students nested within classes nested within schools). However, our preliminary models found greater proportions of variance between classes than schools, so we decided to fit the simpler model to these data. Moreover, given the limited degrees of freedom available to use within schools (roughly 4 classes per school) and between schools (16 elementary schools), we decided that a more parsimonious model would be more appropriate for this analysis. We hope to pool data across years in subsequent analyses to see if a more complex, three-level model is more informative.

Y_{ij} is the achievement of student i taught in class j ;

β_{0j} is the average achievement of students in class j ;

X_{qij} is the vector of $q = 1, \dots, q$ student characteristics;

β_{qj} are the level-1 coefficients that measure the effects of individual student characteristics on individual student achievement; and

r_{ij} is the random error or unique effect of student ij on achievement.

We measure first-grade achievement (Y_{ij}) using students' 2005 mathematics assessment score. Variables that tap the characteristics of students are the students' FARM status and their 2004 mathematics assessment score. We group-mean center FARM status and grand-mean center students' 2004 scale score in the level-1 model. By grand-mean centering students' 2004 achievement score, we adjust the level-2 model for differences between classes in students' prior achievement.

At level 2, we model the average mathematics achievement of students in class j as a function of a vector of teacher qualifications, observed practices, observed content and class characteristics (W_{sj}) and random class error (u_{0j}):

$$\beta_{0j} = \gamma_{00} + \gamma_{0s}W_{sj} + u_{0j},$$

where

β_{0j} is the average mathematics achievement of students in class j ;

γ_{00} is the average class achievement across all classes;

W_{sj} is the vector of $s = 1, \dots, s$ teacher and class characteristics;

γ_{0s} are the level-2 coefficients that measure the effects of teacher and classroom characteristics on average student achievement; and

u_{0j} is the random error or unique effect of class j on achievement.

The level-2 model includes a series of teacher qualifications (MA/MEd degree and years of experience), a measure of average teacher practice (the ratio of the average percent of high to

low practices observed in lessons) and a measure of average lesson content (the ratio of average percent of lesson focused on linking versus conceptual and procedural knowledge). We also include a dummy-coded measure for the concentration of FARM students in a class (1 = 50% or greater, 0 = less than 50%) and a set of interaction terms to determine whether the effects of quality varied with the concentration of FARM students in classes.⁸ We enter all variables at this level uncentered.

Results

We present first descriptive statistics for variables and bivariate analyses of differences in class characteristics, teacher characteristics, teacher practices, and lesson content between low-poverty and high-poverty classes. Next we present the results of our multilevel analysis in which we model average achievement in mathematics classes as a function of teacher qualifications, teacher practices and lesson content.

Descriptive and Bivariate Analyses

Table 1 presents descriptive statistics for the major variables used in the analysis. The last column in the table provides means and percentages for all 66 classes, while the second and third columns provide these data for low-poverty and high-poverty classes respectively.

Roughly half of the classes fall into our low-poverty classification (48%) and half into our high-poverty classification (52%). The average percentage of FARMS students in the low-poverty

⁸ We examined the possibility of interactions between the measures of teacher qualifications and the concentration of FARM students in classes, as well as possible interactions between the measures of teacher qualifications, practice, and content. Because none of these interactions were statistically significant, we do not discuss them in this paper.

classes is 30%, whereas the average percentage of FARMS students in the high-poverty classes is 69%.

As indicated by Table 1, the low-poverty and high-poverty classes in our study differ substantially in class characteristics, particularly in terms of our measures of mathematics achievement. There are slightly more 5th grade classes amongst the low-poverty classes than the high-poverty classes, though the difference is not statistically significant. The low-poverty classes tend to be larger than the high-poverty classes (18 v. 14 students). Smaller class sizes for high-poverty classes are due partially to student mobility (classes with more mobile students will have fewer students with a full range of assessment scores) but also instructional decisions by teachers and principals to reduce class size in classes with high concentrations of low-performing students.⁹ Indeed, the largest differences between the low-poverty and high-poverty classes in the study are differences in average student achievement. The 4th and 5th graders in the low-poverty classes have substantially higher levels of achievement than the 4th and 5th graders in the high-poverty classes. The difference in 2004 scores is more than two-thirds of a standard deviation (.7 SD), while the difference in 2005 scores is more than three-quarters of a standard deviation (.8 SD).

The teachers responsible for instruction in the low-poverty classes have more years of experience and a greater likelihood to possess an advanced degree than the teachers in the high-poverty classes, though this latter difference is not statistically significant. On average, the teachers of low-poverty classes have 11 years of experience, while the teachers of high-poverty classes have 8 years of experience. Because our measure of experience is somewhat skewed, it understates the true difference in experience between teachers of low-poverty and high-poverty

classes. While 13% of the teachers of low-poverty classes have three or fewer years of experience and one-quarter have five or fewer years of experience, half of the teachers of the high-poverty classes have three or fewer years of experience and nearly two-thirds (65%) have five or fewer years of experience. These are strikingly large differences for classes within the same school district.

Table 1 about here

Table 1 also displays differences between low-poverty and high-poverty classes in teacher practices. What we describe as teachers making higher cognitive demands on students occurs only about 7% of the time in the mathematics lessons that we observed; more frequently teachers make lower cognitive demands of their students (21%). Practices that encourage higher cognitive demands occur more frequently in low-poverty than high-poverty classes (9% v 6%), whereas practices that encourage lower cognitive demands occur more frequently in high-poverty than low-poverty classes (23% v. 18%). There is no difference between low-poverty and high-poverty classes in the management activities of teachers. The ratio of practices that we use in our analysis is higher in the low-poverty than the high-poverty classes (.7 v .5).

As with teacher practices, the content that we describe as promoting a deeper understanding of mathematics is a relatively rare event in the mathematics lessons that we observed (6%); more frequently the content of lessons is procedural knowledge (45%) or conceptual knowledge (25%). Although there is no difference between low-poverty and high-poverty classes in the presentation of linking content, the lessons in low-poverty classes have a

⁹ As part of the overall study, we interviewed principals and teachers about the allocation of instructional resources. Among the strategies mentioned by principals and teachers was a reduction in class size for students with lower

slightly greater focus on procedural knowledge than the lessons in high-poverty classes (48% v 42%), whereas the lessons in high-poverty classes have a slightly greater focus on conceptual knowledge than the lessons in low-poverty classes (27% v 23%). There is no difference between low-poverty and high-poverty classes in the ratio of content that we use in our analysis.

Multilevel Models

We specify a fully unconditional model first and then a within-class model that includes individual student's FARMS status and 2004 mathematics achievement (not shown). The intraclass correlation is $\rho = .49$, indicating that roughly half of the variance in students 2005 mathematics assessment score is within classes and half is between classes. The average reliability of the estimates of β_{0j} , the estimate of average achievement in each class, is $\lambda = .93$. The within-class model indicates that FARMS students have lower levels of achievement than non-FARMS students (-.09 SD) and that students with higher levels of prior achievement have substantially higher levels of 2005 achievement than students who began the year behind their peers (.67 SD). These two variables explain roughly half of the within-class variance in achievement, and prior achievement alone explains more than three-quarters of the between class variance. In other words, much of the variance between classes in mathematics achievement is due to the clustering of students by prior achievement levels.

Table 2 about here

levels of achievement.

Table 2 presents the results of our fully conditional models. We present the within-class model and the between-class model for two specifications: a full model that includes all of the control variables and quality indicators and a trimmed model that includes only statistically significant variables and variables associated with the interaction terms. The full model indicates no difference between the 4th and 5th grade classes in average achievement, after controlling for differences in prior achievement and other class characteristics, nor differences between classes in the qualifications of teachers given the other variables in the model. The average achievement of high-poverty classes, however, is approximately one-fifth of a standard deviation (-.19 SD) below that of the average achievement of students in low-poverty classes. Although neither the measure of teachers practices nor the measure of lesson content is statistically significant ($p < .10$), the interaction terms associated with these variables indicate that their effects vary for low-poverty and high-poverty classes. To better understand the nature of the interactions, we “trimmed” the model to include only statistically significant variables and variables associated with the two interaction terms.¹⁰ The coefficients for the trimmed model are essentially the same as the coefficients for the full model, though the error terms are reduced slightly leading to slightly greater t-ratios and lower p values.

¹⁰ We examined possible interactions with grade level and teacher qualifications but none of these interactions proved statistically significant ($p < .10$). Moreover, the trimmed model explains more of the variance between classes than the full model, providing additional justification for excluding from the model non-significant variables and variables not associated with the interaction terms.

Figure 2. Differential Effects of Teacher Practices and Content in Hi- and Low-poverty Classes.

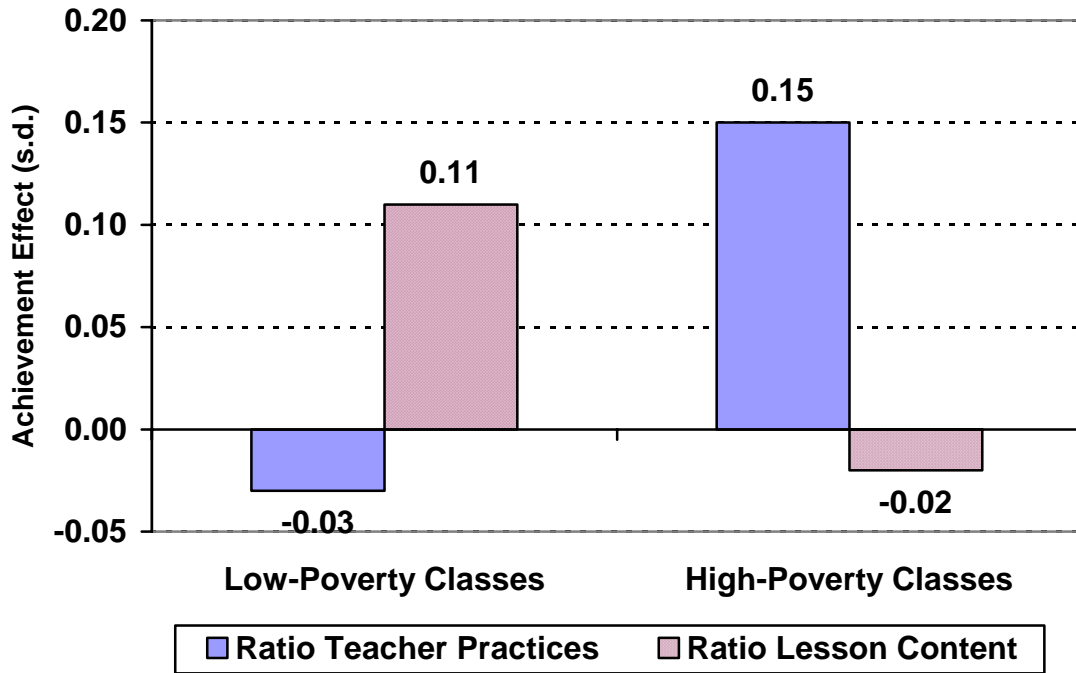


Figure 2 presents graphically the interaction effects. The coefficient for the ratio of high to low teacher practices in low-poverty classes is non-significant – that is, the ratio of teacher practices has no effect on achievement given the other variables in the model. However, the interaction between high FARMS and teacher practices is statistically significant, indicating that while teacher practices does not affect achievement in low-poverty classes, it does affect achievement in high-poverty classes ($-.03 + .18 = .15$ SD). The results are just the opposite for lesson content. The coefficient for the ratio of linking to conceptual and procedural content is a little more than a tenth of a standard deviation (.11 SD), indicating that students in low-poverty classes benefit from more demanding content. However, these same effects do not occur in high-poverty classes ($.11 -.13 = -.02$). The effects are essentially zero.

Discussion

These results are suggestive of potential indicators of what constitutes quality teaching of mathematics in upper-grade elementary classes. These analyses demonstrate what might be referred to as a convergence of “good teaching” and “effective teaching” (Fenstermacher & Richardson, 2005) – that is, practices and content consistent with the reforms advocated by NCTM produce higher levels of mathematics achievement in 4th and 5th grade classes. However, what constitutes one dimension of quality, effective teaching, varies by context, by whether teachers are responsible for instruction in low-poverty or high-poverty classes. Students in low-poverty classes benefit most from more demanding lesson content, as defined by the ratio of linking content to conceptual and procedural content, whereas students in high-poverty classes benefits most from teacher practices that encourage greater cognitive demand and focus on instruction activities. Moreover, the differences in these effects may actually be greater than represented by Figure 2 given that the low-poverty classes in our study have higher levels of poverty than the district’s overall FARM enrollment (see Figure 1).

One explanation might be that students in high-poverty classes do not have sufficient background knowledge of mathematics to take full advantage of the linking content presented to them in lessons. By definition, linking involves connecting conceptual knowledge and procedural knowledge (Hiebert & LeFevre, 1986; NCTM, 2000). If students have limited knowledge in either area, they lack the foundational knowledge required by the content. Given the large differences between low-poverty and high-poverty classes in students’ prior achievement, this seems like a plausible explanation. We plan to further examine this possibility

by developing additional measures of content using the observation protocols and teachers' logs¹¹ to see if the effects vary for different types of mathematics content.

The importance of teacher practices is consistent with a group of studies that have looked at the effects of poverty on student learning. Knapp (1995), for example, in an investigation of high-poverty classes and schools, found that at-risk students achieved at higher levels in mathematics if their teachers promoted higher-order learning skills and encouraged students to reflect more deeply on their learning. These same benefits may not be evident in low-poverty classes because higher-income students may be less dependent on their teachers for guidance and motivation. Such an explanation is consistent with studies that claim historically disadvantaged children, such as the poor, are more dependent on quality teaching and school resources than students who have historically done well in school.

What is potentially disturbing about these findings, however, is that what appears to influence achievement most occurs so infrequently in mathematics lessons at the upper elementary grades. Quality, when measured in terms of the practices and content identified by our analysis, accounts for roughly 17% of lessons in terms of practices (7% high cognitive demand, 10% the management of instructional activities) and 6% of lessons in terms of content. Although this does not preclude that other, equally beneficial practices and content are not occurring during lessons, it does raise the question of what constitutes a "quality" lesson. How much cognitively demanding practice, instructional activity and content should we observe in a lesson before we identify it as of high quality? How much cognitively demanding practice, instructional activity and content should we observe across a series of lessons before we characterize teaching as high quality? Minimally, we should note that we observed less

¹¹ As part of the high-quality teaching study, teachers completed daily logs regarding the content of lessons presented to one of the students in the class. We do not use these data in this paper.

demanding practices and content more often than we observed more demanding practices and content throughout the school year.

Finally, we did not find evidence that teacher qualifications make a difference in students' mathematics achievement – neither teachers holding an advanced degree nor years of experience contributed to gains in achievement. We point out, though, that the study from which we derived our data was not designed to examine the effects of teacher qualifications on achievement. The measures of qualification that we use in this analysis are “blunt.” More “fine-grained” measures might detect more subtle effects, such as differences between teachers in the effectiveness of practices and content based on professional preparation, development or experience. Moreover, the potential effects of experience are confounded by the poverty status of classes, a methodological challenge for those who wish to study teacher qualifications and a political challenge for those who wish to equalize the distribution of educational resources in classes attended by low-income students. We encourage future research designed to tease out the moderating effects of teacher qualifications and the manner in which teacher quality might be more accessible to students from low-income families.

Appendix

Description of Measures Used in the Study

Student-level Measures (n = 1074, 423 4th graders and 651 5th graders)

2005 4th grade mathematics scale score. Fourth-grade students took a state assessment in March of 2005. We use the 4th grade mathematics scale scores in our analysis (M = 399.9; SD = 30.3). Scores range from a high of 452.4 to a low of 311.0.

2005 5th grade mathematics scale score. Fifth-grade students also took the state assessment in March of 2005. We use their mathematics scale scores in our analysis (M = 412.7; SD = 29.2). Scores range from a high of 470.9 to a low of 357.8.

2005 combined scale score. We standardized 4th and 5th grade scale scores and then standardized the scores across grades (M = 0; SD = 1). This allowed us to pool 4th and 5th grade classes in our analysis. We use this measure as the dependent variable. The combined scale score is normally distributed.

2004 3rd grade mathematics scale score. Our fourth-grade cohort took the state assessment in March of 2004 as 3rd graders. We use their 2004 mathematics scale score in our analysis as a measure of prior achievement (M = 398.7; SD = 27.3). Scores range from a high of 447.9 to a low of 337.0.

2004 4th grade mathematics scale score. Our fifth-grade cohort took the state assessment in March of 2004 as 4th graders. We use their 2004 mathematics scale score in our analysis (M = 398.6; SD = 27.6). Scores range from a high of 451.4 to a low of 338.0.

2004 combined scale score. We standardized 3rd and 4th grade scale scores and then standardized the scores across grades (M = 0; SD = 1). Again, this allowed us to pool the 4th and 5th grade cohorts in our analysis. We use this measure as a control for prior achievement. The combined scale score is normally distributed.

FARMS status. We used information about students' participation in government sponsored free-and-reduced-price meals (FARMS) service as a measure of poverty (1 = Yes, 0 = No). Roughly half (50.2%) of the students in the study participate in a FARMS program.

Class-level Measures (n = 66 classes, 60 teachers)

Grade level. We included an indicator to determine whether the average 4th grade scale score differs from the average 5th grade scale score (1 = 5th, 0 = 4th). Slightly more than half of the classes (57.6%) in the study are 5th grade classes.

High poverty. We created an indicator of whether a class had a high percentage of FARMS students (1 = 50% or greater, 0 = less than 50%). Roughly half of classes (51.5%) are high poverty classes.

Number of students. Class student samples range from a low of 5 to a high of 26. The mean number of students per class is 16.3.

MA/MEd. We used teachers' self-reports about their academic histories to create an indicator of whether teachers had an advanced degree (1 = MA/MEd, 0 = BA/BS). Nearly two-thirds of the teachers (65.2%) have an advanced degree.

Years experience. Teachers also gave us information about their years of experience teaching in the district. Teacher experience ranges from 1 to 25 years. The mean number of years is 9.4. The distribution is positively skewed.

Avg. pct. lessons teacher higher cognitive demand. We observed roughly 8 mathematics lessons per class during the 2004-05 school year. Observers used a time sampling program to record what was happening in the class every three minutes. We used this information to compute the percent of time that teachers made higher cognitive demands on the students (e.g., request that students reflect on learning, provide alternative methods or strategies, or pose higher-order problems and tasks as part of the lesson). We averaged these percentages across the lessons that we observed. Classes range from a low of 0% to a high of 17.4%. The mean percent of time is 7.2%.

Avg. pct. lessons teacher lower cognitive demand. Information from observations was used to compute the percent of time that teachers made lower cognitive demands on students (e.g., poses routine/low-order question or elaborates on a routine exercise/low-order question). We averaged these percentages across the lessons that we observed. Classes range from a low of 4.9% to a high of 44.1%. The mean percent of time is 20.7%.

Avg. pct. lessons teacher higher management. We used information from observations to compute the percent of time that teachers managed an activity and students were attentive. We averaged these percentages across the lessons that we observed. Classes range from a low of 0% to a high of 22.2%. The mean percent of time is 9.9%.

Avg. pct. lessons teacher lower management. We also used information from the observations to compute the percent of time that teachers managed student behavior or classroom materials. We averaged these percentages across the lessons that we observed. Classes range from a low of 0.7% to a high of 31.3%. The mean percent of time is 11.1%.

Ratio teacher high to low practices. Teacher high practices included both higher cognitive demand and higher management, while teacher low practices included both lower cognitive demand and lower management. The ratios range from a low of 0.06 to a high of 2.4. The mean ratio of teacher high to low practices is 0.6. The measure is near

normally distributed. We use a standardized version of this measure in our analysis ($M = 0$; $SD = 1$).

Avg. pct. lessons linking content. Information from the observations was used to compute the percent of time that the lesson content was linking procedural to conceptual knowledge. This lesson content is sometimes described as providing students with the deepest and most useful understanding of mathematics. We averaged these percentages across the lessons that we observed. Classes range from a low of 0% to a high of 22.2%. The mean percent of time is 6.2%.

Avg. pct. lessons conceptual content. We used information from the observations to compute the percent of time that the lesson content was conceptual, including conceptual learning strategies. We averaged these percentages across the lessons that we observed. Classes range from a low of 2.9% to a high of 51.3%. The mean percent of time is 24.7%.

Avg. pct. lessons procedural content. We used information from the observations to compute the percent of time that the lesson content was procedural, including procedural learning strategies. We averaged these percentages across the lessons that we observed. Classes range from a low of 19.5% to a high of 80.5%. The mean percent of time is 45.0%.

Ratio linking to conceptual and procedural. We computed the ratio of avg. pct. lessons focused on linking to the sum of the avg. pct. of lessons focused on conceptual and procedural content. The ratios range from a low of nearly 0 to a high of 0.4. The mean ratio of teacher high to low practices is 0.1. The measure is near normally distributed. We use a standardized version of this measure in our analysis ($M = 0$; $SD = 1$).

Interaction high poverty by ratio teacher practices. To examine possible interactions between practices and the percentage of FARMS students in classes, we computed the interaction term (High poverty x Ratio teacher high to low practices).

Interaction high poverty by ratio teacher content. To examine possible interactions between content and the percentage of FARMS students in classes, we computed the interaction term (High poverty x Ratio linking to conceptual and procedural).

References

- Alexander, P. A., & Murphy, P. K. (1998). The research base for APA's learner-centered psychological principles. In N.M. Lambert & B.L. McCombs (Eds.), *Issues in school reform: A sampler of psychological perspectives on learner-centered schools*. Washington DC: The American Psychological Association.
- American Psychological Association Board of Educational Affairs. (1995, Dec.). *Learner-centered psychological principles: A framework for school redesign and reform*. Washington, DC: American Psychological Association. (<http://www.apa.org/ed/lcp.html>)
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes, *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco: Jossey-Bass
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328-75). New York: Macmillan.
- Cohen, D.K., & Hill, H. (2000). *Learning policy: When state education reform works*. New Haven: Yale University Press.
- Cohen, D.K., Raudenbush, S.W., & Ball, D.L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 1-24.
- Croninger, R. G., Rice, J. K., & Rathbun, A., Nishio, M. (in press). Teacher qualifications and early learning: Effects of certification status, degree, and experience on first-grade achievement. *Economics of Education Review*.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Journal of Education Policy Analysis*, 8(1).
- Fenstermacher, G.D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, 107(1), 186-213.
- Ferguson, R.F. & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H.F. Ladd (ed.) *Holding schools accountable: Performance-based reform in education* (pp. 265-298). Washington, D.C.: Brookings.
- Floden, R. E. (2001). Research on effects of teaching: A continuing model for research on teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 3-16). Washington, DC: American Educational Research Association.

- Floden, R. E. (2003). Policy tools for improving education. In R. E. Floden (Ed.), *Review of research in education* (Vol. 27, pp. ix –xii). Washington DC: American Educational Research Association.
- Gamoran, A. (2003). *Transforming teaching in mathematics and science: How schools and districts can support change*. New York: Teachers College Press
- Goldhaber, D., & Anthony, E. (2003). *Teacher quality and student achievement*. New York: Teachers College Press.
- Goldhaber, D.D. & Brewer, D.J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-146.
- Goldhaber, D.D. & Brewer, D.J. (1998). When should we reward degrees for teachers? *Phi Delta Kappan*, 80(2), 134-138.
- Goldhaber, D.D. & Brewer, D.J. (1997a). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, 32(3), 505-523.
- Goldhaber, D.D. & Brewer, D.J. (1997b). Evaluating the effect of teacher degree level on educational performance. In W.J. Fowler (ed.) *Developments in School Finance, 1996*, Washington, D.C: U.S. Department of Education.
- Good, T., & Brophy, J. (2003). *Looking in classrooms* (9th ed.) Boston: Allyn & Bacon.
- Hawley, W., & Rosenholtz, S. (1984). "Good schools: A synthesis of research on how schools influence student achievement." Special issue, *Peabody Journal of Education*, 4, 1- 178.
- Hannaway, J., & Woodroffe, N. (2003). Policy instruments in education. In R.E. Floden (Ed.), *Review of research in education* (Vol., 27, pp. 1-24). Washington, DC: AERA.
- Hiebert, J., & LeFevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum.
- Hiebert, J., Sigler, J.W., Jacobs, J.K., Givvin, K.B., Garnier, H., Smith, M., Hollingsworth, H., Manaster, A., Wearne, D., & Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 Video Study. *Educational Evaluation and Policy Analysis*, 27(2), 11-132.
- Hill, H.C., Rowan, B., & Ball, D.L. (2005). Effects of teachers' mathematics knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371-406.

- Knapp, M. S. (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.
- Murnane, R.J. (1975). *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.
- Murnane, R.J. & Phillips, B.R. (1981). Learning by doing, vintage and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review* 1(4), 453-465.
- NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- No Child Left Behind Act, (2001). Public Law 107-110.
- Porter, A. (1989). A curriculum out of balance: The case of elementary school mathematics. *Educational Researcher*, 18 (5), 9-15.
- Porter, A. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.
- Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1986). *Content determinants* (Research Series 79). East Lansing, MI: Michigan State University.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods* (2nd edition). Thousand Oaks, CA: Sage
- Rice, J.K. (2003). *Teacher quality: Understanding the effects of teacher attributes*. Washington, D.C.: Economic Policy Institute.
- 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 (October), 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 student of elementary schools. *Teachers College Record*.
- Sanders, W., & Horn, S.P. (1998). *Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.
- Sanders, W.L. & Rivers, J.C. (1996). *Cumulative and residual effects of teachers on future academic achievement*. University of Tennessee Value-Added Research and Assessment Center.

Shulman, L.S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3-36). New York: Macmillan.

Smith, T.M., Desimone, L.M., & Ueno, K. (2005). Highly qualified to do what? The relationship between NCLB teacher quality mandates and the use of reform-oriented instruction in middle school mathematics. *Educational Evaluation and Policy Analysis*, 27, 75-109.

Stevenson, D.L., & Baker, D.P. (1991). State control of curriculum and classroom instruction. *Sociology of Education*, 64, 1-10.

Stodolsky, S. (1988). The subject matters: Classroom activity in math and social studies. Chicago: University of Chicago Press.

Table 1 Descriptive statistics for high- and low-poverty classes (n = 66)^a

<i>Variables</i>	<i>Low Poverty (Less than 50% FARMS) n=32</i>	<i>High Poverty (50% or greater FARMS) n=34</i>	<i>Total n=66</i>
<i>Class Characteristics</i>			
Pct. FARMS	30.4	68.9	50.2 ^{***}
Pct. Fifth Grade	62.5	52.9	57.6
Number of Students	18.3	14.4	16.3 ^{**}
Avg. 4 th Grd. 2005 Math Score	416.2	385.7	399.9 ^{**}
Avg. 5 th Grd. 2005 Math Score	426.4	398.9	412.7 ^{**}
Avg. Comb. 2005 Math Score ^{bc}	0.3	-0.5	0.0 ^{**}
Avg. 4 th Grd. 2004 Math Score	415.3	384.4	398.7 ^{**}
Avg. 5 th Grd. 2004 Math Score	411.8	358.3	398.6 ^{**}
Avg. Comb. 2004 Math Score ^c	0.3	-0.4	0.0 ^{**}
<i>Teacher Characteristics</i>			
Yrs. Experience ^c	11.3	7.7	9.4 [~]
Pct. MA/Med	68.8	61.8	65.2
<i>Teacher Practices</i>			
Avg. Pct. High Cog.	8.5	6.1	7.2 [*]
Avg. Pct. Low Cog.	18.0	23.3	20.7 ^{**}
Avg. Pct. High Mng.	10.2	9.7	9.9
Avg. Pct. Low Mng.	10.9	11.2	11.1
Ratio High to Low ^c	0.7	0.5	0.6 ^{**}
<i>Lesson Content</i>			
Avg. Pct. Linking	6.1	6.2	6.2
Avg. Pct. Conceptual	22.6	26.7	24.7 [~]
Avg. Pct. Procedural	48.1	42.1	45.0 [*]
Ratio Linking to Con. & Proc. ^c	0.1	0.1	0.1

*** $p < .001$; ** $p < .01$; * $p < .05$; ~ $p < .10$

^a We defined a low-poverty class as having less than 50% of the students on FARMS and a high-poverty class as having 50% or more. We used t-tests to compare the means between the low- and high-poverty classes and report the results in the total column.

^b We standardized the 4th and 5th grade scale scores for 2005 and 2004 and then combined the scores by year. We further standardized the combine scores for our analyses. The study population includes 28 4th grade classes (13 low- and 15 high-poverty classes) and 38 5th grade classes (19 low- and 19 high-poverty classes).

^c We use a standardized version of these variables in our analyses ($M = 0$, $SD = 1$).

Table 2 Effects of teacher qualification, practices, and content on math achievement (n = 1,074 students in 66 classes)^a

<i>Variables</i>	<i>Full Model</i>	<i>Trimmed Model</i>
<u>Within-Class Model for Outcome^b, Y_{ij}</u>		
Intercept, γ_{00}	0.10	0.10 [~]
FARMS Status (1 = Yes), γ_{10}	-0.09 [*]	-0.09 [*]
2004 Math Scale Score ^c , γ_{20}	0.66 ^{***}	0.66 ^{***}
<u>Between-Class Model for Intercept, γ_{00}</u>		
Fifth Grade, γ_{01}	-0.03	
MA/MEd Degree, γ_{02}	0.03	
Years of Experience ^b , γ_{03}	0.02	
High FARMS (1 = Yes), γ_{04}	-0.19 [*]	-0.20 [*]
Ratio Hi/Lo Teacher Practices ^c , γ_{05}	-0.02	-0.03
Ratio Linking/Conceptual & Procedural ^c , γ_{06}	0.10	0.11 [~]
Interaction FARMS by Ratio Hi/Lo Teacher Practices, γ_{07}	0.18 [~]	0.18 [~]
Interaction FARMS by Ratio Linking/Conceptual & Procedural, γ_{08}	-0.14 [~]	-0.13 [*]

[~] $p < 0.10$; ^{*} $p < 0.05$; ^{**} $p < 0.01$; ^{***} $p < 0.001$

^a Coefficients can be interpreted as effect sizes – that is, as the percentage change in the standard deviation of the dependent variable attributable to a unit change in the independent variable. The overall model explains roughly two-thirds of the variance in the 2005 mathematics scale score (49% within classes, 82% between classes); the trimmed model explains 9% of the between-class variance in the scale score, controlling for students prior achievement..

^b Individual FARM status is grouped-mean centered; the 2004 mathematics scale score is grand-mean centered.

^c The measure is standardized (M = 0, SD = 1). See the appendix for information about the metrics of all variables.