

A Multi-method Look at Teaching Quality: Insights and Quandaries from One Study of Teaching<sup>1</sup>

by

Robert G. Croninger,  
University of Maryland

Daria Buese,  
University of Maryland

&

John Larson,  
University of Maryland

April 3, 2006

**DRAFT: PLEASE DO NOT CITE OR QUOTE WITHOUT AUTHORS' PERMISSION**

---

<sup>1</sup> This work was supported by 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](mailto:croninge@umd.edu).

## **Abstract**

This paper examines the challenges of combining multiple methods as a strategy for representing the complexities of teaching, such as the complexities associated with potentially competing perspectives on “quality” and the conditional nature in which “quality” manifests itself in teaching. We draw on our own work and the work of colleagues engaged in a longitudinal study of teaching in 4<sup>th</sup> and 5th grade classes in public elementary schools. We use data collected during the 2004-2005 school year, including student academic records, roughly 500 observed mathematics lessons in 66 mathematics classes, teachers’ logs, and qualitative interviews with teachers about the challenges that they face in providing quality instruction to their students. We describe our research study and methods, provide an example of an exploratory analysis that combines methods and identifies the conditional nature of quality, and discuss the possibilities and limitations of using these different streams of data to characterize teaching. We conclude with a discussion of possible implications for future research (both our own and that of others).

## **Introduction**

Teaching is a complex and sometimes baffling process. Teachers and students face a range of specific and ambiguous objectives that they are to accomplish through daily interactions within classes and schools. Although observers of these interactions can identify well-rehearsed routines and processes, there is far less agreement about what teachers and students actually do to promote learning (Ball & Lampert, 1999).

Descriptive phrases such as “teaching as performance” (Sarason, 1999), “teaching for understanding” (Cohen, McLaughlin, & Talbert, 1993), “teaching as problem solving” (Lampert, 2001) or “effective and responsible teaching” (Oser, Dick, & Patry, 1992) identify what might be considered essential aspects of teaching, but they do not identify the ontological and epistemological boundaries that surround teaching as an educational and social phenomenon.

The study of teaching becomes even more vexing when education researchers move from describing teaching to characterizing the nexus between teaching and learning (Floden, 2001). Although there is mounting evidence that “teacher quality” is one of the more powerful in-school factors for improving student learning (Darling-Hammond, 1997; Rice, 2003; Sanders & Horn, 1998), especially for students from historically disadvantaged populations (Darling-Hammond, 1997; Goldhaber & Anthony, 2003), there is little guidance about what either policymakers or practitioners should do to actually bring about desirable outcomes for students. Despite a varied and well-established history of investigations into teaching, the research base is more provocative than prescriptive (Ball & Rowan, 2004).

The desire to provide useful, research-based information to policymakers and teachers poses a series of problems for education researchers. First, if teaching *is* a complex process, such complexities do not lend themselves to straight-forward interpretations of what teachers should do to help students learn or what policymakers should do to encourage quality teaching. Although there have been numerous attempts to distill research into “best practices,” such litanies provide only generic guidance and may even be counter-productive to learning in specific instructional contexts or situations. As a result, education researchers are faced with the challenge of finding the right balance between complexity and simplicity in studies of teaching, particularly when such studies are meant to provide useful information to policymakers and practitioners.

Second, assessing the quality of teaching is not only a methodological challenge but a political or moral challenge (Fenstermacher & Richardson, 2005). Because what teachers are to accomplish in the classroom is “culturally contestable,” to use Bruner’s phrase (1999, p. 401), reaching agreement about any one perspective or set of educational objectives is highly doubtful (and perhaps not even desirable in a democratic society). Consequently, education researchers are faced with the challenge of incorporating multiple perspectives about what constitutes quality teaching into their research designs so as to provide more robust and balanced findings for policymakers and practitioners.

Finally, not only is the appraisal of quality dependent upon an evaluative perspective, but it is also dependent upon the various contexts in which teaching occurs. The subject matter or lesson being taught, the resources available to teachers in and outside of the classroom, students’ prior knowledge, motivation, and academic experiences, all are factors that influence teaching, including the qualities of teaching that

lead to desirable outcomes for students (Alexander & Murphy, 1999; Brophy & Good, 1986; Fenstermacher & Richardson, 2005; Floden, 2001). A third challenge faced by education researchers, therefore, is how to adequately acknowledge the conditional nature of “quality” and still provide information that is generalizable to educationally meaningful situations.

In this paper, we address these enduring challenges and discuss our own experiences as participants in a multi-year study of teaching in 4<sup>th</sup> and 5<sup>th</sup> grade classes in public elementary schools. We describe the overall approach that we adopted to these challenges, explain the design of our research study, discuss the instruments and methodological strategies that we used to capture different aspects of teaching, present an example of how we are seeking to analyze and integrate data from the study, and discuss what we consider to be some of the possibilities and limitations associated with our approach. We conclude with a brief discussion of methodological approaches to address the complexities of teaching, potentially conflicting perspectives on quality, and the conditional nature of what constitutes both “good” and “effective” teaching in schools (Fenstermacher & Richardson, 2005).

### **Background**

The history of research on teaching reflects the challenges and the methodological decisions that researchers have made to address them (Floden, 2001). Some studies have emphasized preserving the complexities of teaching by focusing on in-depth cases and narratives so as to develop rich portraits of what teachers do in their classes (e.g., Dillon, 1989; Ladson-Billings, 1994; Lampert, 2001), while other studies have emphasized large-

scale surveys so as to develop broad-brushed characterizations of teaching representative of larger populations (see Rowan, Correnti, & Miller, 2002, for a review and critique). Still other studies have promoted the use of teacher logs to examine the content presented to students (Porter 2002), teacher and student interviews to consider beliefs and understandings (e.g., Allington & Johnson, 2002), student work to investigate what students learn or can do (e.g., Newmann & Associates, 1996), and classroom observations to examine what students and teachers actually do in classrooms (e.g., Good & Brophy, 2003).

Yet another approach has been to focus less on the means of teaching and more on strengthening statistical techniques to isolate achievement gains that might be attributed to teachers' interactions with students. A renewed interest in the use of random experimental designs to identify practices, typically in the form of scripted instruction or well-specified interventions, represents an effort to identify reliable practices and programs that can be successfully "scaled up" to other teachers as a strategy for promoting teaching quality (Boruch & Mostellar, 2001; Whitehurst, 2002). Value-add modeling, less concerned about what teachers actually do and more concerned about identifying the variability between teachers in achievement gains, provides a metric for policymakers to calibrate "quality" teaching to rewards and possible sanctions (Rivkin, Hanushek, & Kain; Sanders & Horn, 1998). Although the hazards and potential difficulties associated with these techniques are debated (see, e.g., McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Raudenbush, 2005; Valli, Croninger, & Walters, 2005), they have become important methodological strategies for identifying teaching quality and informing policy (Lissitz, 2005).

Although the use of these approaches to study teaching has produced a rich body of findings and multiple perspectives on teaching, this multiplicity has yet to provide a satisfying portrayal of what constitutes “quality” teaching or what practitioners can do to promote it in classrooms (Ball & Rowan, 2004). Spurred by this literature and the literature on mixed-methods approaches to research (Bryman, 1988; Greene, 2001), we designed an interdisciplinary study of teaching that we believed could capture the complexity of teaching, generate useful information, and explore the conditional nature of quality – that is, provide a more nuanced portrayal of what constitutes quality in teaching.<sup>2</sup> To do so we developed a multi-methods design that incorporated both quantitative and qualitative modes of inquiry. In the paragraphs that follow we describe our efforts, as well as the efforts of our colleagues,<sup>3</sup> to implement a mixed-methods study capable of examining and enhancing our understanding of the practices and processes of teaching from multiple methodological and normative perspectives.

### **Description of the Study**

The High-Quality Teaching (HQT) Study, a multi-year, federally funded, interdisciplinary examination of teaching in 4<sup>th</sup> and 5<sup>th</sup> grade classes, was designed to provide a multifaceted, nuanced examination of what constitutes quality teaching. The purpose of the HQT study is to investigate what teachers do to help 4<sup>th</sup> and 5<sup>th</sup> grade students succeed in reading and mathematics, as well as to investigate how various

---

<sup>2</sup> Our goal was not to create an exhaustive or completely satisfying portrayal of teaching, for doing so would assume that the methodological challenges we describe are indeed resolvable; rather, our goal is to develop a portrayal that captures the complexity of teaching and explores the conditional nature of what constitutes quality teaching in the upper-grades of public elementary schools.

<sup>3</sup> Other members of the research study include Drs. Linda Valli, Marilyn Chambliss, Anna Graeber, and Patricia Alexander from the University of Maryland, and Dr. Jeremy Price, from Montclair State University.

educational policies and organizational factors influence the ability of teachers to scale up and sustain good or quality teaching over time. Five basic research questions guide the study:

1. What do 4<sup>th</sup> and 5<sup>th</sup> grade teachers do to help students acquire foundational skills in reading and mathematics?
2. What do these teachers do to help close the achievement gap between traditionally high- and low-performing students in these subject areas?
3. How do successful teachers change their pedagogical practices to respond to new educational challenges and priorities?
4. How do various education policies and organizational factors influence the quality of teaching that occurs in 4<sup>th</sup> and 5<sup>th</sup> grade classes?
5. What are the attributes of high-quality teaching in different contexts and subject areas? How do these qualities converge and diverge with different perspectives on high-quality teaching?

Using multiple methods, investigators follow teachers over time to compare teaching in reading and mathematics, in classes comprised of different types of students, in elementary schools with different organizational structures, and in years with different policy pressures. The study design called for the development, piloting, and modification of instruments during the 2001-2002 school year; data collection continued through the 2002-2003, 2003-2004, and 2004-2005 school years. Throughout this period of time more than 170 teachers and 25 elementary schools participated in the study. During the final year of the study, we have focused our attention on archiving data and developing analytic models and strategies to create multifaceted representations of teaching in different contexts and subject areas (i.e., reading and mathematics).

Data gathered as part of the study include classroom observations, lesson artifacts (e.g., handouts), teachers' daily logs, student achievement and academic records,

interviews with teachers and principals, and in-depth case studies of individual teachers and schools. For this paper, we draw on data from classroom observations, teachers' daily logs, focus-group and individual teacher interviews, and student achievement records. We describe each of these data streams next.

### *Observation Protocols*

As one perspective on teaching quality, we developed a time-sampling observation protocol for observing reading and mathematics instruction in the 4<sup>th</sup> and 5<sup>th</sup> grades. (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 (3<sup>rd</sup> through 6<sup>th</sup> grades), instructional standards promoted by professional organizations (e.g., the National Council of Teachers of Mathematics, 2000, and the International Reading Association and National Council of Teachers of English, 1996), the research literature on high-quality teaching in reading and mathematics, and lessons learned from the initial piloting of the instrument.

To facilitate data collection and coding, we developed software for recording observations, minimizing entry errors by observers, and compiling observations into database systems. Using the software, observers respond to a set of questions presented to them on an automated screen on a laptop computer every three minutes. 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, types of materials, and student attentiveness. Each question provides the observer with a

hierarchy of possible answers. Roughly half of the answers are identical for the reading and mathematics protocols, permitting comparisons of practice by subject area.

The study design called for 6-8 observations per class per year. Teachers typically provided us with access to a “typical” reading class and a “typical” mathematics class. In a few schools, where individual teachers specialized in a subject area, teachers granted us access to either two typical reading classes or two typical mathematics classes. Each observation lasted roughly 60 minutes, providing about 20 data collection episodes per lesson (60 minutes divided by 3 minutes per episode) and 120-160 data collection episodes per class across the course of a year (20 episodes times 6-8 observations). All observers participated in pre-observation training to ensure interreliability; observers were also periodically paired with a second observer during the year to monitor potential drift in reliability. During biweekly meetings, observers discussed potential coding problems and resolved coding differences. (Examples of the observation protocols can be found at <http://education.umd.edu/EDCI/hqtstudy/HQTInstruments.htm>.)

### *Teacher Logs*

Because a number of studies have encouraged the use of teacher logs as a method for gauging quality teaching, we developed a daily survey for teachers to complete about one of the students in their class.<sup>4</sup> The logs were recorded on PDAs (Personal Data Assistants) given to participating teachers when they first agreed to be a part of the study. Each PDA included a roster of the students in the class and a survey designed to identify various aspects of curriculum coverage, including broad topics (e.g., reading, writing,

---

<sup>4</sup> We programmed each PDA with the student names assigned to each class. Teachers selected one student per day, selecting a student from the list alphabetically, and re-selecting students only after cycling through all of the students in the class.

working independent of text, etc.), specific details about text (e.g., narrative, expository, or poetry), specific details about the goals for activities (e.g., comprehension, decoding, spelling, etc.), and the use of various technologies (e.g., internet browser, interactive software, etc.).<sup>5</sup> As with the observation protocols, we based the content of the surveys on the district's formal curriculum (3<sup>rd</sup> through 6<sup>th</sup> grades), curricular standards promoted by professional associations, the research literature, and our early experiences piloting the instrument. Although the teachers' logs provided much greater detail about curriculum coverage, there was some overlap, at the most general levels, with information collected through the observation protocols; the logs also provided opportunities to compare aspects of curriculum coverage for reading and mathematics.

We provided teachers with a detailed glossary and instructions about how to complete the daily logs. We also provided teachers with training, used their input from these sessions to improve the wording of items and increase the instrument's validity and reliability, and assigned several members of the research team to answer questions, supply technical support, and provide periodic reports to teachers regarding the number of log records that they had successfully uploaded to date. (Examples of the daily log can be found at <http://education.umd.edu/EDCI/hqtstudy/HQTInstruments.htm>.)

### *Student Academic and Achievement Records*

At the beginning of the school year (late September, early October), participating teachers provided us with the student rosters for the classes that we would observe as part of the study. During the third and fourth year of the study, we verified these rosters in the

---

<sup>5</sup> Each PDA contained a survey about mathematics curriculum coverage and reading curriculum coverage. The examples given in the text are for the reading survey.

spring to determine changes in student-teacher assignments. At the end of the school year, we requested the most recent achievement data for students (either the most recent 4<sup>th</sup> or 5<sup>th</sup> grade achievement score on the state-mandated assessment), as well as the students' prior achievement record (2<sup>nd</sup> through 4<sup>th</sup> grade for 5<sup>th</sup> graders; 2<sup>nd</sup> through 3<sup>rd</sup> grade for 4<sup>th</sup> graders) for every student on a class roster. We also requested demographic information (e.g., race, gender, Free and Reduced-Price Meal Service FARMS) and related academic histories for students (e.g., English for Speakers of Other Languages [ESOL] status, Special Education Status, and time in the district.). We linked these records to the observation and log data compiled in each class.

### *Case Studies and Interviews*

We designed the qualitative and quantitative dimensions of the study to complement each other – that is, to measure “overlapping but distinct facets” of teaching in the classes and schools that participated in the study (Greene, 2001, p. 253).<sup>6</sup> The qualitative data include observational case studies of classroom teaching, focus-group interviews, individual interviews with principals and teachers, and year-long case studies at selected schools. We also attempted to interview teachers before and after each observation to establish teachers' expectations for a lesson and gather teachers' assessments of “how the lesson went.” However, we found these pre- and post-interviews difficult to complete, as the press of school day seldom provided an opportunity to speak at length with teachers before or after an observation.

Selected observational case studies of math or reading lessons provided an

---

<sup>6</sup> Although “complementarity” was the primary relationship between the quantitative and qualitative streams of data that we collected, we also used qualitative data to develop or probe specific observations in classes and expand our understanding of policy influences through the school case studies.

opportunity to examine teachers' instructional rationales in connection to teacher practices. Along with the time sampling protocol, observers compiled field notes and interviewed teachers about their teaching beliefs, knowledge, and expectations. The in-depth case studies of schools, one the other hand, were designed to provide greater insight into how principals, teachers, and key support staff (e.g., staff developers) implemented local, state, and federal policies, including high-stakes accountability policies like No Child Left Behind, and how they thought these policies influenced teaching in their schools. We conducted the observational lesson case studies during the second and third years of the study, while the school case studies were the primary focus for qualitative data during the fourth year of the study.

During the first three years of the study, we interviewed principals at participating schools, asking about school goals, schedules, assignments, programs, student/teacher characteristics, teacher support, resources, and leadership styles. During this same time period we used focus-group meetings with participating teachers to capture the language teachers use to describe quality teaching, what they regard as supports for quality teaching, and what they view as constraints to their own efforts to achieve high levels of quality in their classes. These qualitative data, along with the lesson and school case studies, illuminate salient elements of teaching, variation within teaching, and the relationship of classroom-level events to school and district level events (Bogdan & Biklen, 1992; Erickson, 1986; Merriam, 1998; Strauss, 1987; Stake, 1995).

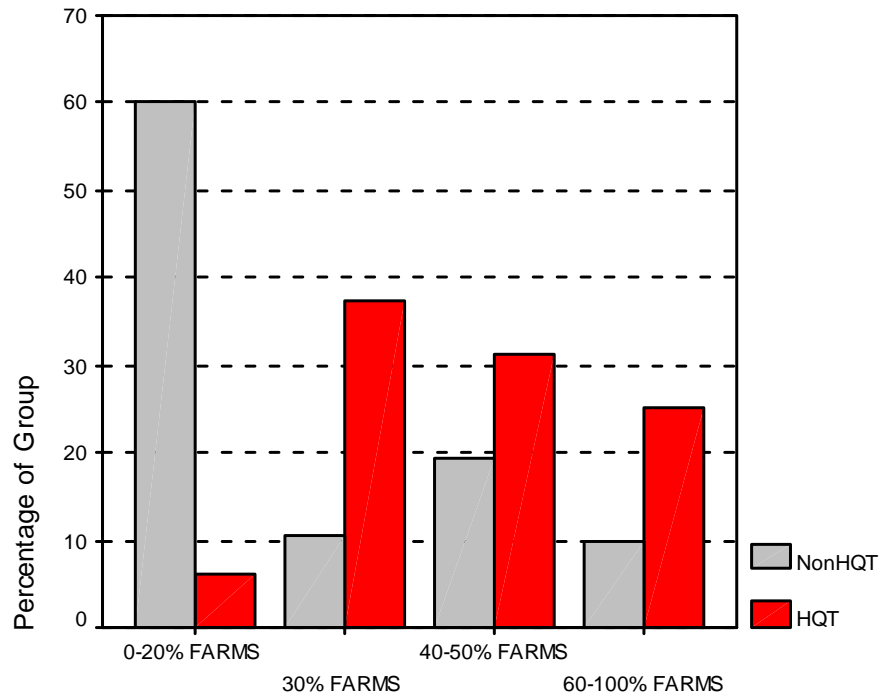
### *Study Site*

The schools in our study come from one of the 20 largest school systems in the nation. The school district employs more than 19,000 employees and enrolls nearly 140,000 students (K-12). Like many large, metropolitan school districts, the district faces a wide range of educational challenges associated with the dramatic changes in its student population and heightened expectations for student achievement. In the past two decades, the enrollment of low-income students has doubled. More than 30% of students receive FARMS assistance and 20% are currently, or have been at one time, enrolled in ESOL programs. We selected the school district because of its racial and economic diversity, its reputation for collecting a range of reliable data about students and schools, and the district's willingness to partner with us in conducting the research study.

The study design called for us to identify a group of moderate-to-high poverty schools with greater than expected achievement gains in the district and then to follow these schools and their 4<sup>th</sup> and 5<sup>th</sup> grade teachers for three to four years. 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 the 4<sup>th</sup> and 5<sup>th</sup> grade teachers to participate, particularly those teachers who taught mathematics and reading in regular 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.

Data for this paper come from classroom observations and teachers logs completed during the 2004-05 school year. We also rely on individual case study interviews with selected teachers conducted in the spring of 2004, as well as focus-group and individual interviews conducted in 2002-03, the year that the district launched a new mathematics curriculum, and focus-group and individual interviews conducted in 2004-05. During the 2004-05 school year 16 elementary schools participated in the study. Nearly two-thirds of the schools in the study had FARMS concentrations of over 40%, and slightly over half of the students (58%) in the study schools were either African American or Hispanic. See Figure 1 for a comparison of poverty enrollments at the HQT schools and elementary schools not included in the study.

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



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.<sup>7</sup> 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 teachers curriculum logs for 46 mathematics classes. The average number of entries was 65 days or roughly 3 months of lessons. 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 4<sup>th</sup> and 5<sup>th</sup> grade students who had data for the state-required mathematics assessment in 2005 and 2004, we had a total of 1,074 students, taught by 63 different teachers, in 66 different classes.<sup>8</sup>

### **An Example**

To facilitate the integration of data, we developed a conceptual framework based on our interests in exploring the complexities of teaching, particularly the potential conditional nature of quality. Although it is always helpful to approach an analysis with a conceptual framework, Greene (2001, p. 256) argues that such frameworks are especially important in mixed-methods studies for they provide “an organized set of conceptual starting places for integrative analysis.” The framework (see Figure 2), like those proposed by others (e.g., Lampert, 2001; Schwab, 1973), focuses on the interactions of students and teachers around subject matter in specific contexts or educational milieus. We portray learning as occurring within a network of interactions

---

<sup>7</sup> Three teachers provided us with information about two of their mathematics classes, giving us a total of 66 classes for the year.

<sup>8</sup> Roughly three-quarters of the students (78%) included on the class rosters.

that includes teacher practices and characteristics (e.g., qualifications), curricular content (e.g., subject matter), and context (e.g., students' prior knowledge).

-----  
Figure 2 about here  
-----

In this example, we integrate four streams of data – student mathematics achievement records, observational data for mathematics classes, teachers' log data for their mathematics classes, and qualitative data drawn from case studies, individual interviews, and focus-group interviews.<sup>9</sup> We begin with a quantitative analysis of the first three data streams, and then follow this analysis with an examination of qualitative data drawn from archived case study and interview files using NVivo© software for qualitative analysis. Of course, there is no reason why we could not start with a qualitative analysis and then proceed to a quantitative analysis. We envision future analyses and integration of data to proceed more iteratively, with the quantitative and qualitative analyses informing each other.

### **Quantitative Analysis**

We focus on a small set of measures to examine the potential interplay of teacher practices, content, and context on learning. We describe the key measures used in the quantitative analysis, relating them to the conceptual framework; characterize these variables in terms of the context for instruction; and present the results of a multilevel analysis that serves as the springboard for the qualitative analysis and interpretation.

---

<sup>9</sup> We integrate multiple streams of data for mathematics in this paper, and only for data collected during the 2004-05 school year. In subsequent analyses, we will attempt to integrate data and our findings across subjects and across years.

### *Key Variables*

Learning. We use student achievement on a state-mandated mathematics assessment as a measure of learning. The assessment is linked to the state's learning goals and includes a set of items that permit national comparisons of student performance. Mathematical knowledge covered by the assessment includes algebra, geometry, statistics and probability, number sense, and mathematical procedures. We use student scale scores in 2005 as a dependent variable in a multilevel analysis and students' 2004 scale scores as a control for prior achievement. While we do not believe that assessment data are the only criteria by which to measure quality or even what Fenstermacher and Richardson (2005) describe as "effective teaching", they are criteria of considerable importance to policymakers and practitioners.<sup>10</sup> (See the Appendix for a fuller description of measures relevant to the quantitative analysis.)

Teaching Construct. We use 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

---

<sup>10</sup> See the other papers in this symposium for discussions about epistemological and political issues surrounding the use of achievement data to measure learning or as an indicator of quality.

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.

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.<sup>11</sup> We then combined these three measures to estimate the ratio of linking content to the sum of conceptual and procedural content. This measure, along with the measure of teacher practice, might be considered as indicators of what Fenstermacher and Richardson (2005) refer to as “good teaching” in mathematics.

We also used data from teachers’ logs to characterize curriculum coverage in classes. Teachers who completed the logs designated the amount of time that they spent on one of seven possible topics: algebra, geometry, statistics and probability, computation, measurement, assessment, and reasoning. We calculated the standard deviation across all possible topics as a possible measure of curriculum balance. High scores indicate greater variability in the amount of time spent of topics (some topics

---

<sup>11</sup> 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.

received much more attention than others); low scores indicate greater balance in the amount of time spent on topics.

Context. The primary context of interest for these analyses is whether a class can be characterized as a high-poverty class or a low-poverty class. Classes with fewer than 50% of the students receiving FARMS we classified as low-poverty classes; classes with half or more of the students receiving FARMS we classified as high-poverty classes. We are especially interested in determining whether the relationship with achievement for the measures of reform-oriented practices and content (ratio of high to low practices and the ratio of linking to conceptual and procedural) are contingent upon whether instruction occurs in high-poverty or low-poverty classes.

#### *Characterization of Context*

Table 1 presents descriptive statistics for the key 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 classes is 30%, whereas the average percentage of FARMS students in the high-poverty classes is 69%.

The low-poverty and high-poverty classes in our study population differ in selected class characteristics, particularly in terms of mathematics achievement. There are slightly more 5<sup>th</sup> grade classes amongst the low-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 in part to student mobility (classes with more mobile students 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. The largest differences between the low-poverty and high-poverty classes, however, are in average student achievement. The 4<sup>th</sup> and 5<sup>th</sup> graders in the low-poverty classes have substantially higher levels of achievement than the 4<sup>th</sup> and 5<sup>th</sup> 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).

-----  
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 of activities or the management of materials or behavior. 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 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.

More than two-thirds of the mathematics classes have teachers' log data for 2004-05. Somewhat more high-poverty classes than low-poverty classes are missing these data, though the difference is not statistically significant. Overall, there was slightly less variability in the time spent on topics in low-poverty classes (-.1 SD) compared to high-poverty classes (.1 SD), but, again, this difference is not statistically significant.

### *Multilevel Analysis*

Table 2 presents the results of our fully conditional model.<sup>12</sup> The within-class model indicates that FARMS students scored lower than non-FARMS students by .09 SD on the 2005 state mandated mathematics assessment, while students with higher prior

---

<sup>12</sup> We ran several models for arriving at the model presented in Table 2. Earlier models included a dummy-coded variable for fifth-grade (1 = 5<sup>th</sup> grade, 0 = 4<sup>th</sup> grade), a dummy-coded variable for missing log data (1 = Yes, 0 = No), and an interaction term between the SD of topics covered and High FARMS. We also experimented with a control for the number of log entries for a class (1 = fewer than 30, 0 = greater than 30). None of these variables proved statistically significant ( $p < .10$ ), so we dropped them from the final model. Moreover, the more parsimonious model explained more of the variance in the intercept than the fuller model, providing additional justification for the deletion of these variables from the analysis.

achievement scored higher than students with lower prior achievement (the difference being .67 SD on the 2005 assessment for a student who scored 1 SD above the mean on the 2004 assessment). To control for differences between classes in students' average prior achievement, we grand-mean centered this variable in the within-class model. As a result, effects reported for variables in the between-class model are net of differences between classes in students' 2004 assessment scores.<sup>13</sup>

-----  
Table 2 about here  
-----

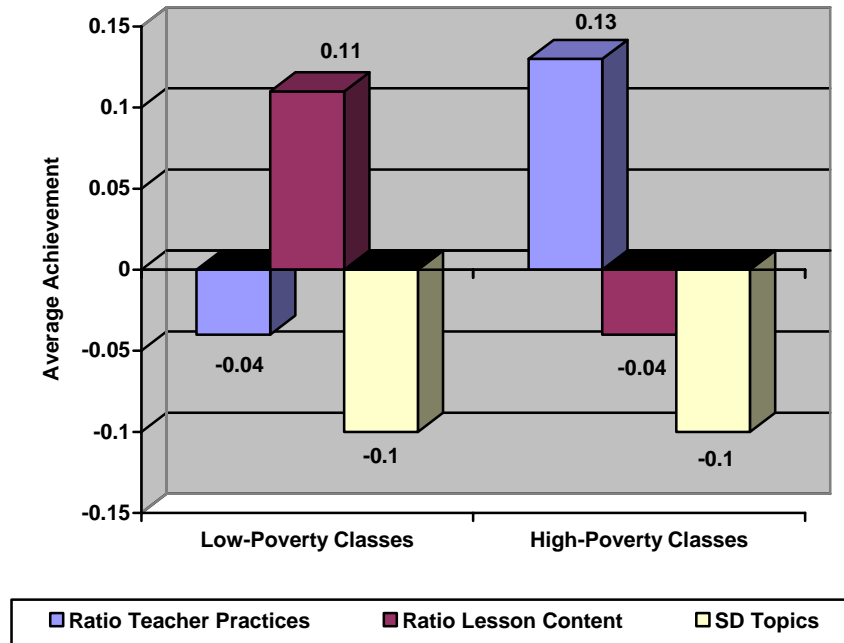
The between-class model indicates that students in high-poverty classes have lower levels of average achievement than students in low-poverty classes (-.18 SD). More importantly, though, it also indicates that the effects of some of our measures of teacher practices and content are contingent on the poverty status of the class. While students in classes where teachers reported greater variability in topic coverage had lower levels of achievement (-.10 SD), regardless of the poverty status of the class, the effects of the ratio of high to low teacher practices increased achievement only in high-poverty classes (-.04 + .17 = .13 SD), whereas the ratio of linking content to procedural and conceptual content only increased achievement in low-poverty classes (.09 SD). Figure 3 presents these interaction effects graphically, along with the effects associated with variability in coverage of mathematics topics.<sup>14</sup>

---

<sup>13</sup> Classes differed substantially in the prior achievement of students, as indicated by the differences in 2004 assessment scores between low-poverty and high-poverty classes in Table 1. Controlling for differences between classes in students' prior achievement explained nearly two-thirds of the difference between classes in students' 2005 assessment scores.

<sup>14</sup> The bars in Figure 3 can be derived directly from Table 2. For example, the effect of the ratio of teacher practices is -.04 SD in a low-poverty class and .13 SD in a high-poverty class (-.04 + .17); similarly, the effect of curriculum content is .09 SD in a low-poverty class and -.04 SD in a high-poverty class (.09 - .13).

Figure 3 Effects of teacher practices, lesson content and curriculum coverage in high-poverty and low-poverty classes.



### Qualitative Analysis

Although the findings from the quantitative analysis capture some of the complexities of teaching, particularly the notion that what constitutes quality may vary with context, they do not shed much light on why the poverty status of classes might mediate these effects. Nor does the quantitative analysis tell us much about why classes where teachers report greater variation in coverage of mathematics topics would have lower levels of achievement than classes where teachers report less variation, regardless of poverty status. While the findings are intriguing, they are not especially enlightening. We turn to our fourth stream of data, the qualitative data, to add greater meaning to these findings. We first describe the information sources used for the qualitative analysis, and then we relate these data to the quantitative findings presented in Figure 3.

### *Characterization of Data*

We began our qualitative analysis by examining archived data using NVivo. For the purposes of this paper, we retrieved data from coding nodes containing examples of policy influences on teaching mathematics, the enactment of the district's mathematics curriculum, the influence of district and state testing on teachers' work in the classroom, concerns teachers had about student learning in mathematics, and how the school climate, organizational structures, and programs influenced teachers' work. We also looked at how these nodes overlapped and interacted to reveal the density of the influences and complexity of contexts in which teaching mathematics occurred.

We first examined data from four focus group interviews with teachers conducted in the spring of 2003. In those interviews, we questioned teachers on their personal beliefs about quality teaching and asked them how they believed their teaching was supported or constrained by the district mathematics and reading curricula, organizational factors within their schools, and professional development provided to them by their schools or the district. These data are particularly relevant to our understanding of the mathematics achievement data because teachers were nearing the end of their first year of a new, district-designed mathematics curriculum. This curriculum was a significant departure from the way the teachers were accustomed to teaching mathematics, and it was a difficult transition year for nearly every teacher we spoke with.

We then examined data from case studies of two lessons, one each taught by a teacher who had participated in the study in 2004-05. Mr. DiLorretto<sup>15</sup> taught mathematics to 5<sup>th</sup> graders in a low-poverty class and school, while Mr. Forrest taught

---

<sup>15</sup> All names mentioned in the paper are pseudonyms.

mathematics to 5<sup>th</sup> graders in high-poverty class and school. Each teacher was selected for a case study during the 2003-04 school year because each had been identified by observers as engaging in more reform-oriented practices than other teachers being observed in the study that year (e.g., posing demanding problems to students, requests for reflection, multiple representation of concepts); each also had relatively high values on the measure of teacher practices used in the quantitative analysis.<sup>16</sup> Data from these cases include field notes, which provide greater details about the lesson than the observation protocol, and individual interviews with the teachers.

A final source of data is the in-depth case studies of three schools that we conducted during the 2004-05 school year; data from these cases provide insight into the third year of the enactment of the mathematics curriculum (the same year of data used in the quantitative analysis). We interviewed staff developers and math specialists, individually, to learn more about the curriculum, local resources, professional development and teacher learning. In May 2005 we interviewed focus groups of these same specialists, and also, teachers in their grade level teams. We asked how federal, state, local and school policies affected teaching practices, teachers' thinking about teaching, their roles, responsibilities, and students, and how policies supported, constrained, or changed their perspectives on high-quality teaching or helped them address the achievement gap.

---

<sup>16</sup> As suggested in Table 1, many of these practices were relatively rare in the lessons that we observed; however, they were less rare in the lessons taught by some teachers compared to the lessons taught by other teachers.

### *A More Complex Tale*

During the 2002-03 school year, the district implemented a new mathematics curriculum designed to align instruction across grades, ensure greater uniformity in curriculum coverage and create a stronger foundation for students to pursue more advanced coursework in later grades. The newly implemented curriculum was fast-paced, content dense, and placed a greater emphasis on establishing a deeper understanding of mathematics. Although teachers still had textbooks to use in class, the textbooks were not aligned with the instructional guide or newly mandated unit tests (roughly 5-6 per year). Nor was the new curriculum fully aligned with the states' learning objectives and mandated assessment in mathematics. The initial reaction was one of frustration, anger and resistance. Naomi, a mathematics teacher in a low-poverty school, provides an example from the spring 2003 focus group interview with teachers:

I can already tell you that once I become familiar with the new math curriculum I'm not going to be teaching all these objectives in the same order they're being presented by the county because I have never done that before. If it doesn't make sense to the children, it doesn't make sense to me... The unit two is geometry and unit five is geometry. Teach it all at the same time. Give them a chance to master it.

While teachers in high-poverty classes and schools expressed similar concerns about the new curriculum, they also talked about the greater emphasis on reading, writing, and conceptual knowledge. Many felt that their students lacked the foundational skills required to master the units included in the curriculum, particularly given the order in which the units were scheduled and the pace which provided little opportunity to repeat a lesson. Nadine, a mathematics teacher in a high-poverty class, voiced these concerns at the end of the first year of the curriculum's implementation:

... it's very hard to teach a fifth grade curriculum because the curriculum changed and so it was too big of a leap to ask the kids to make... it's very hard if kids

cannot read and you have to read everything to them... I mean, one thing we did one day was in decimals and they didn't understand "least to greatest." I think the new vocabulary, the new curriculum now is really big on words that nobody really uses in their everyday vocabulary. ...but... when they go to the assessment, I have to read greatest and least.<sup>17</sup>

Much of what we heard in the end-of-year focus-group interviews might be attributed to the normal stresses associated with a major change in the curriculum. Teachers talked about needing to go back and review and deepen their own mathematics knowledge. The new curriculum placed greater cognitive and pedagogical demands on them, just as it placed greater cognitive and performance demands on their students. More importantly, though, these early interviews suggest that the challenges faced by teachers in low-poverty and high-poverty classes were somewhat different. In low-poverty classes, teachers needed to understand the new curriculum better and find ways to adjust the sequence and pacing of units to suit their students' learning; in high-poverty classes, however, the challenge was much greater. Teachers had to grapple with the same challenges as the teachers in low-poverty classes but also develop strategies for providing their students with the vocabulary and foundational skills required by the new curriculum.

Effects of variation in topic coverage. One explanation for the negative effects associated with greater variation in curriculum coverage may be the ability of teachers to adjust the sequencing and pacing of the new curriculum to their students' learning. By the spring of 2005, there was less dissatisfaction with the curriculum and more familiarity with the underlying logic of the district's curriculum guide. Some of the teachers who participated in the end-of-year focus groups in our three case study schools, though not

---

<sup>17</sup> The demanding vocabulary presented in the curriculum was a concern in both high- and low-poverty classrooms. This concern was particularly acute, and created greater challenges for teachers, in classrooms with high ESOL populations.

all, felt that new curriculum provided greater freedom than they initially realized, freedom to be creative and tailor their mathematics lessons to their own strengths and student learning. When asked whether teachers were becoming more comfortable with the mathematics curriculum, Ms. Seibert, a mathematics specialist in a high-poverty school, told us:

Some are [becoming more comfortable]... [but] its different as to the group of kids that you have. I mean some teachers are feeling more comfortable with the curriculum in that they are using their personal judgment, whereas the first year or two they would really fight me... They were using the instructional guide more as a Bible than as a guide... You don't have to do the lesson in the guide; you have to teach that objective.

Although more teachers expressed comfort with the curriculum in the spring of 2005 than in the spring of 2003, most teachers still felt the pressure associated with the fast pace of the curriculum, a pressure that was directly linked to getting through the curriculum by March, when their students had to take the state mandated assessment in mathematics. As schools approached the scheduled assessment, many teachers dropped the district curriculum and began to focus on topics not yet covered but likely to be on the state-mandated assessment. Anita, a mathematics teacher in a low-poverty school, explained:

I think we would all agree the most frustrating of all was around the [assessment] time. Because things were so out of order. I mean, literally, day to day, we were changing concepts on them, because [of] the [state-mandated assessment] pressure and the way that the curriculum is not aligned with [the assessment]... by the time the [assessment] comes up, there's still a lot of things that we have not taught, because it (a tested topic) was at the end of the year, according to the curriculum.

There was considerable pressure on teachers to cover a breadth of curriculum, and to cover it rapidly so that students were at least "exposed" to topics before the assessment. As Anita described the pressure, "that was no mastery at all, we were just,

‘kids... this is what you do, move on.’” Such pressures may have been especially acute for teachers who felt uncomfortable with the curriculum, who did not understand its logic or have the ability to tailor the curriculum to their own students. The more teachers succumbed to these pressures, the greater the variation in topic coverage in their classes. If students had an insufficient understanding of these topics by the time of the assessment, it would help to explain the negative effects of curriculum coverage identified by the quantitative analysis.

Differential effects of lesson content. We observed linking in mathematics classes only about 6% of the time, roughly the same amount of time in low-poverty classes as high-poverty classes. When it did occur, however, it promoted higher levels of learning only in low-poverty classes. One explanation for this relationship may be that students in high-poverty classes simply lacked the foundational knowledge to benefit from the content. An exchange during a focus group interview in the spring of 2005 between Mr. Wilson and Ms. Gabriel, 5<sup>th</sup> grade teachers in a low-poverty and high-poverty class, revealed similar goals for students but very different starting points. Mr. Wilson praised the new curriculum because “I think it is rigorous and I think it does challenge the kids to really think about what they are doing...” Ms. Gabriel agreed, “They have to think... they got to think.” But the two provide very different examples of “thinking” in their classes.

In Mr. Wilson’s class, he talked about students linking a particular procedure to the underlying concept of balancing equations:

...when I was in school, you know, I was told when you’re balancing an equation you circle the number, you pull it over, you know... it changes sides when it goes over... What did you do? Well, I hopped the fence. And that doesn’t make sense to everybody.

He then went on to explain how important it is to explain the underlying concept of equations, the logic of isolating variables and balancing equations, so that his students “have that underlying understanding of what’s going on in... math.” Ms. Gabriel, however, while agreeing, posed a very different challenge. She wanted her students to understand “basic math concept[s]:”

I’m hoping that in the years to come that when... you look at reasonable answers and, you know, you have 245 cans and you have to divide them. ...and each of five classes has to share them and they’ll put 2,000... look at your answer and see if it is reasonable. How could it be reasonable? ...And that kind of thing, the basics of basics, is what they don’t have.

While Mr. Wilson is linking procedures for solving equations with the concept of balancing equations, Ms. Gabriel is linking division with number sense while performing basic operations.

Many of the high-poverty teachers with whom we spoke talked about the importance of helping low-achieving students grasp “basic math concepts,” a concern not emphasized by teachers in low-poverty classes. As indicated in Table 1, building conceptual knowledge in mathematics was more of a priority in high-poverty classes than in low-poverty classes. We do not interpret this exchange, or the results of the quantitative analysis, to indicate that students in high-poverty classes do not benefit from content that links concepts and procedures; however, the benefits may be greater, especially as measured by achievements tests, for students who have greater foundational skills and better conceptual understanding of mathematics.

Differential effects of teacher practices. We observed what we characterized as high-quality versus low-quality practices in classes no more than 17% of the time, with some of these practices being somewhat more predominant in low-poverty than high-poverty classes. Nonetheless, when these practices did occur, the quantitative analysis

indicated that they benefited students in low-poverty classes more than they benefited students in high-poverty classes. One explanation may be the greater emphasis on vocabulary in the new mathematics curriculum, an emphasis felt especially by teachers in high-poverty classes. Ms. Breman, a staff developer in a high-poverty school, explained:

I mean one of the things that we know is really, really important is discussion. Discussion in mathematical projects, discussion of the leading concepts and stories, and it is really difficult when the language is not there.

The need for building students' mathematical vocabulary was stated to us in the spring of 2003, when the curriculum was first implemented, as well as two years later in our case study schools.

Mr. DiLorretto, a 5<sup>th</sup> grade teacher in a low-poverty class, made no mention in his interview of language problems or the need to develop students' mathematical vocabulary. He "strove to have students understand concepts at a high level;" engaged students orally in discussions of mathematical concepts, and relied heavily on the textbook for his handouts and worksheets. He explained to us:

Well, the math has a lot of reading involved... I think the reading's definitely involved with the math and I try to involve the math with the reading as much as I can through activities, various novel activities.

By contrast, Mr. Forrest, a 5<sup>th</sup> grade teacher in a high-poverty class, spoke frequently about the challenges that language posed in implementing the mathematics curriculum in his classroom. When we observed Mr. Forrest's lesson, he constantly re-worded problems so his students would understand them, linked math concepts to everyday knowledge that he thought students possessed, and created his own handouts and worksheets for students. "I move more slowly than we're supposed to," he told us, and "integrate concepts from other units into the one I'm doing." He emphasized

understanding the meaning of vocabulary and basic concepts of mathematics throughout the case lesson.

Well planned classroom interactions do not necessarily mean that students and teachers are engaged in highly demanding, advanced mathematical work. As a 4<sup>th</sup> grade teacher in a high-poverty class lamented, “You try inquiry but it never... I won’t say never, but rarely did that construction or that inquiry lesson lead them to applying it.” But the more successful teachers, even when they “repeat, and repeat and repeat,” are representing mathematical concepts in multiple ways and clarifying vocabulary for students that are essential to the curriculum and likely to be on the assessment. While students in low-poverty classes may also benefit from these practices, they may not be as dependent on their teachers to mediate the curriculum as students in high-poverty classes. If so, we would expect these practices, which emphasize greater oral exchanges with teachers and multiple activities to represent mathematical concepts, to have a greater influence on achievement in high-poverty classes than low-poverty classes.

### **Discussion**

We began this paper by discussing the challenges of studying teaching. We argued that education researchers must strike a balance between complexity and simplicity in the portrayal of teaching, address the potential conditional nature of what constitutes quality in teaching, and appreciate the multiple perspectives by which quality in teaching might be judged. We also described the manner in which we sought to address these challenges in our own study, primarily by using multiple methods to examine teaching, and provided an example of an analysis that integrates data across

methods, both quantitative and qualitative. We reflect on the analysis and the challenges to studying teaching next.

### **Striking a Balance**

Integrating multiple methods into studies of teaching, increases the likelihood that education researchers will strike a better balance between an accurate portrayal of the complexity of teaching and the desire to provide useful information to practitioners and policymakers. In the example that we presented, observational data provided a window into teachers' lessons, particularly the interactions between students and teachers around subject matter; log data provided an opportunity to incorporate aspects of curriculum coverage into our models; and student achievement records provided the possibility of linking these dimensions of teaching to achievement gains. Most importantly, though, the qualitative data, gathered through individual interviews, focus-group interviews, and field notes of selected lessons and schools, provided insights into the underlying meanings of the relationships identified by the quantitative analysis. The integration of these various data streams permitted a richer portrayal of teaching, more consistent, we believe, with the pedagogical challenges that teachers face in the classroom.

Although the analysis presented in this paper is preliminary and exploratory, it is suggestive of the possible insights that might be garnered from a multi-method look at teaching quality. First, what constitutes quality teaching, at least as judged by achievement gains, may be conditional in nature. Students in high-poverty classes are more dependent on their teachers to mediate the curriculum and provide multiple representations of mathematics, whereas students in low-poverty classes may be better

able than students in high-poverty classes to access mathematical knowledge through textbooks, standard worksheets, and more complex lesson content. One possible explanation is that students in high-poverty classes lack the mathematical vocabulary and basic conceptual knowledge required to understand the lesson content without substantial mediation by their teachers. In Vygotsky's terms (1978), the content might be characterized as beyond students' "zone of proximal development," without, that is, substantial mediation by teachers.

Second, teaching exists in a policy environment that helps to shape the learning opportunities presented to students and the nature of teachers' actions. In the analysis that we presented, the implementation of a more demanding mathematics curriculum, along with the pressure to prepare students for the state-mandated assessments in March, created additional pressures on teachers in terms of curriculum coverage. Although many teachers felt more comfortable with the curriculum two years after its implementation, the fast-pace of the curriculum and the "march" to March remained a challenge for most teachers. As suggested by our analysis, an important indicator of quality may be the ability of teachers to navigate successfully the policy environment for their students— in this case, the ability to meet the demands of the curriculum and assessment timeline and still present students with a coherent and appropriate set of lessons.

A major challenge for us, as well as any study that seeks to integrate data across multiple methods, will be to *maintain* a balance between an accurate portrayal of the complexity of teaching and a desire to provide useful information to practitioners and policymakers. While the volume and detail of data that we have collected provide considerable opportunity to examine teaching from multiple perspectives, it also

threatens to overwhelm our ability to integrate these data into a coherent set of propositions about what constitutes quality in teaching. Multiple sets of analyses may make it more difficult to meaningfully synthesize results, or they may identify a more comprehensive framework by which to discuss findings. As we move forward with our analyses, we plan to continue to integrate multiple data streams so as to better understand the linkages presented in Figure 2. Our plan is to develop models within mathematics and reading and then compare these models to identify common themes and subject-specific themes pertinent to our research questions. Although this will not guarantee that we will be able to maintain the balance we desire, keeping the challenge at the forefront of our analyses should increase the likelihood that we will do so.

### **Conditional Quality**

Our analysis highlighted the conditional nature of quality, a phenomenon brought into greater focus by the use of multiple methods. Although the interaction terms in Table 2 identify possible differential effects for the ratio of teacher high-low practices and the ratio of linking to conceptual and procedural knowledge, these interactions, while statistically significant, failed to take on meaning until we examined the qualitative data for possible explanations. Moreover, not all aspects of quality were conditional, at least not as measured by the achievement status of students at the end of the year. Both students in low-poverty and high-poverty classes benefited from having teachers who were successful at navigating the district's new mathematics curriculum and the state's mathematics assessment timeline to present them with coherent lessons.

These conditional qualities, however, will not be identified or understood unless they are specifically examined in studies of teaching. The associations between students' achievement gains and our selected measures of teacher practices and content are essentially zero without examining the possible interactions that these measures have with poverty status of classes. The conditional qualities of teaching have long been asserted by education researchers but not well understood or systematically examined in the study of teaching (Alexander & Murphy, 1998; Ball & Lampert, 1999; Brophy & Good, 1986; Cohen Raudenbush, & Ball, 2003; Fenstermacher & Richardson, 2005; Floden, 2001; Lampert, 2001). Studies that pursue these conditional qualities will enrich not only our knowledge of teaching but also our knowledge of policies and practices that promote and perhaps even inhibit desirable outcomes for students.

### **Good and Effective Teaching**

Our analysis suggests a possible convergence of notions of “good” teaching and “effective” teaching as defined by our measures of practice, content and learning, though the convergence is more conditional than it is sometimes portrayed in the literature. Although the convergence is reassuring, the relative scarcity of these practices is troubling. Many of the practices promoted by reform advocates, such as inquiry-based learning or cognitively demanding dialogue, did not occur frequently in the mathematics lessons that we observed; more frequent were occurrences of direct instruction, even amongst teachers who characterized inquiry-based learning as characteristic of quality teaching. The press to raise achievement in schools made little distinction between good and effective teaching. On the contrary, good teaching in mathematics *was* effective

teaching – teaching that raised students’ test scores on the state-mandated mathematics assessment.

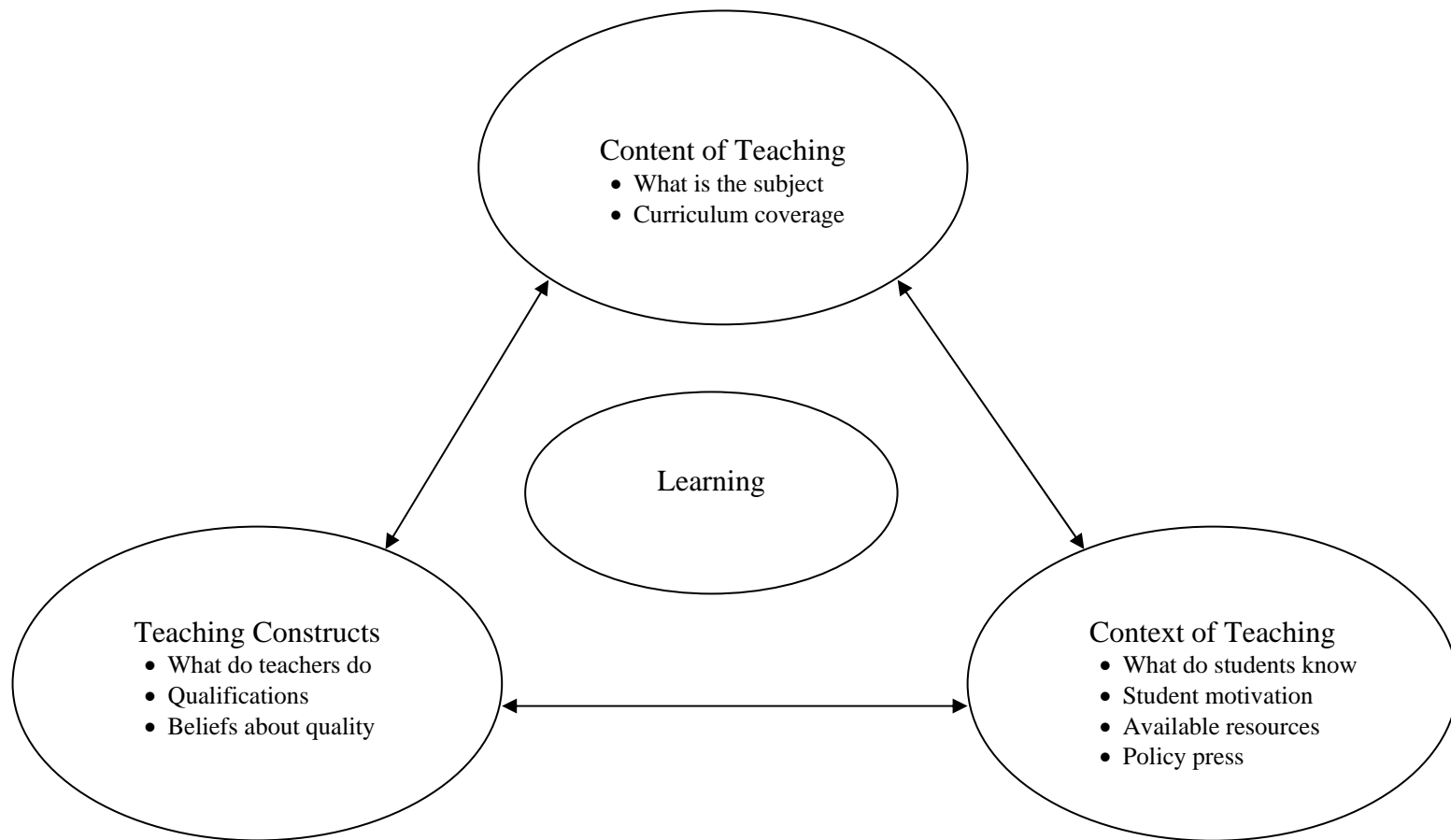
Our analysis does very little to challenge this assertion about teaching quality. By using the state’s mandated assessment as an indicator of learning, we have privileged the mathematical knowledge and skills tested by the assessment. A different assessment or a different set of student outcomes might show an even greater convergence between reform-oriented practices and student learning (see Newmann & Associates, 1996, for an example in mathematics and Allington & Johnson, 2002, for an example in reading). Without an examination of these possibilities, good teaching becomes raising test scores, especially in a high-stakes environment where teachers face increasingly severe sanctions for failing to meet mandated levels of achievement. In all of the classes and schools that we observed, the press to raise achievement test scores was considerable, especially in classes and schools that enrolled students with the greatest likelihood of failure.

Counter balancing the dominance of state-mandated assessments in determining quality teaching is probably the most difficult challenge faced by educational researchers. The state-mandated assessments are ubiquitous and they have real-life consequences for teachers, students, and schools. Moreover, developing alternative assessments is technical difficult, expensive and ethically problematic, particularly given the amount of assessment time that already interferes with instruction and student learning. Furthermore, gaining consent from parents and administrators to collect data directly from students can be a daunting and difficult task, too often ending with spotty participation and uncomfortably large amounts of missing data. Developing strategies to

address these challenges is of paramount importance if good teaching is not to become defined simply as teaching that produces higher test scores.

In our own study, we plan to examine multiple measures of learning to broaden, at least potentially, the criteria used to gauge teaching quality. The state-mandated assessment used by the schools in the study provides a series of subscales that may permit a more detailed examination of the relationship between teaching and the acquisition of different domains of mathematical knowledge (e.g., Algebra, Geometry, Statistics/Probability, Number Sense, and Computation). A second possibility is examining not the scale score but the proficiency levels designated by the state. These types of analyses might reveal greater convergence between notions of good and effective teaching depending on the type of knowledge being assessed or the proficiency level attained by students. A final possibility is not to examine the link with assessment data but to examine alternative outcomes thought to indicate deeper engagement in mathematics, such as the extent to which students respond with alternative explanations, elaborate on an answer, or pose a hypothesis as part of a lesson. If these and similar student behaviors are thought to constitute a positive outcome for students, in and of themselves, then they are worthy of being examined more closely as an alternative indicator of quality teaching.

**Figure 2 Initial Conceptual Framework for Analyses**



**Table 1** Descriptive statistics for low- and high-poverty classes (n = 66)<sup>a</sup>

<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 <sup>***</sup>
Pct. Fifth Grade	62.5	52.9	57.6
Number of Students	18.3	14.4	16.3 <sup>**</sup>
Ave. 2005 Math Score <sup>bc</sup>	0.3	-0.5	0.0 <sup>**</sup>
Ave. 2004 Math Score <sup>c</sup>	0.3	-0.4	0.0 <sup>**</sup>
<i>Teacher Practices</i>			
Ave. Pct. High Cog.	8.5	6.1	7.2 <sup>*</sup>
Ave. Pct. Low Cog.	18.0	23.3	20.7 <sup>**</sup>
Ave. Pct. High Mng.	10.2	9.7	9.9
Ave. Pct. Low Mng.	10.9	11.2	11.1
Ratio High to Low <sup>c</sup>	0.7	0.5	0.6 <sup>**</sup>
<i>Lesson Content</i>			
Ave. Pct. Linking	6.1	6.2	6.2
Ave. Pct. Conceptual	22.6	26.7	24.7 <sup>~</sup>
Ave. Pct. Procedural	48.1	42.1	45.0 <sup>*</sup>
Ratio Linking to Con. & Proc. <sup>c</sup>	0.1	0.1	0.1
<i>Curriculum Coverage</i>			
Pct. No Log Data	25.0	35.3	30.3
SD Major Topics Covered <sup>c</sup>	-0.1	0.1	0.0

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

<sup>a</sup> We defined a low-poverty class as having less than 50% of the students receiving 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.

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

<sup>c</sup> We use a standardized version of these variables in our analyses (M = 0, SD = 1).

**Table 2** Effects of teacher practices, lesson content and curriculum coverage on math achievement (n = 1,074 students in 66 classes)<sup>a</sup>

<i>Variables</i>	<i>Final Model<sup>b</sup></i>
<i>Within-Class Model for Outcome, <sup>c</sup> Y<sub>ij</sub></i>	
Intercept, $\gamma_{00}$	0.09
FARM Status (1 = Yes), $\gamma_{10}$	-0.09 <sup>*</sup>
2004 Math Scale Score, <sup>d</sup> $\gamma_{20}$	0.67 <sup>***</sup>
<i>Between-Class Model for Intercept, <math>\gamma_{00}</math></i>	
High FARMS (1 = Yes), $\gamma_{01}$	-0.18 <sup>*</sup>
Ratio Hi/Lo Teacher Practices, <sup>d</sup> $\gamma_{02}$	-0.04
Ratio Linking/Conceptual & Procedural, <sup>d</sup> $\gamma_{03}$	0.09 <sup>~</sup>
SD of Topics Covered, <sup>d</sup> $\gamma_{04}$	-0.10 <sup>**</sup>
Interaction FARMS by Ratio Hi/Lo Teacher Practices, $\gamma_{05}$	0.17 <sup>~</sup>
Interaction FARMS by Ratio Linking/ Conceptual & Procedural, $\gamma_{06}$	-0.13 <sup>~</sup>

<sup>~</sup>  $p < 0.10$ ; <sup>\*</sup>  $p < 0.05$ ; <sup>\*\*</sup>  $p < 0.01$ ; <sup>\*\*\*</sup>  $p < 0.001$

<sup>a</sup> 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.

<sup>b</sup> A previous model included dummy-coded variables for fifth grade and no log data, as well as an interaction between SD of topics covered and High FARMS. We dropped these non-significant variables from the model and specified a more parsimonious model. The final model explains roughly two-thirds of the variance in the 2005 mathematics scale score (49% within classes, 84% between classes); the between-class model explains 19% of the variance in average achievement controlling for students prior achievement.

<sup>c</sup> Individual FARMS status is grouped-mean centered; the 2004 mathematics scale score is grand-mean centered.

<sup>d</sup>The measure is standardized (M = 0, SD = 1). See the appendix for information about the metrics of all variables.

## **Appendix**

### **Description of Measures Used in the Study**

#### **Student-level Measures (n = 1074, 423 4<sup>th</sup> graders and 651 5<sup>th</sup> graders)**

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

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

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

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

Fifth grade. We included a measure to determine whether the average 4<sup>th</sup> grade scale score differs from the average 5<sup>th</sup> grade scale score (1 = 5<sup>th</sup>, 0 = 4<sup>th</sup>). Slightly more than half of the classes (57.6%) in the study are 5<sup>th</sup> grade classes.

SD of topics covered in log. We took the percentage of time teachers spent on different topics and calculated SD across topics. We then standardized the score (M = 0; SD = 1). Classes without log data received the mean score for this measure. High scores indicate greater variability in the amount of time spent on topics; low scores indicate greater balance in the amount of time spent on topics. Topics include algebra, geometry, statistics and probability, computation, measurement, assessment, and reasoning.

No log data. We created a missing data flag to determine whether classes with and without log data differed in average 2005 achievement scores (1 = No Log Data, 0 = Has Log Data). Slightly less than one-third of the classes (30%) have no log data.

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.

Ave. 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%.

Ave. 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%.

Ave. 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%.

Ave. 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 ( $M = 0$ ;  $SD = 1$ ).

Ave. 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%.

Ave. 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%.

Ave. 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 ave. pct. lessons focused on linking to the sum of the ave. 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).

Interaction high poverty by SD of log topics. To examine possible interactions between log topics and the percentage of FARMS students in classes, we computed the interaction term (High poverty x SD of log topics).

## 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.
- Allington, R.L., & Johnson, P.H. (2002) *Reading to learn: Lessons from exemplary fourth-grade classrooms*. New York: Guilford Press.
- Ball, D.L., & Lampert, M. (1999). Multiples of evidence, time, and perspective: Revising the study of teaching and learning. In E.C. Lagermann & L.S. Shulman (Eds.), *Issues in education research* (pp. 371-398). San Francisco: Jossey-Bass.
- Ball, D.L., & Rowan, B. (2004). Introduction: Measuring instruction. *The Elementary School Journal*, 105 (1), 3-10.
- Bogdan, R., & Biklen, S.K. (1992). *Qualitative research for education. An introduction to theory and methods* (2<sup>nd</sup> edition). Boston: Allyn & Bacon.
- Boruch, R. & Mostellar, F. (2001). (Eds.), *Education, evaluation and randomized trials*. Washington DC: Brookings Institute.
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3<sup>rd</sup> ed., pp. 328-75). New York: Macmillan.
- Bruner, J. (1999). Postscript. Some reflections on education research. In E.C. Lagermann & L.S. Shulman (Eds.), *Issues in education research* (pp. 399-409). San Francisco: Jossey-Bass.
- Bryman, A. (1988). *Quantity and quality in social research*. London: Routledge.
- Cohen, D., McLaughlin, M. W., Talbert, J. E. (Eds.) (1993). *Teaching for understanding. Challenges for policy and practice*. San Francisco: Jossey Bass.
- 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., Larson, J.C., & VonSeker, C. (2006 April). Effects of teacher qualifications, practices, and content on 4<sup>th</sup> and 5<sup>th</sup> grade mathematics achievement in high- and low-poverty classes. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.

- Dillon, D. (1989). Showing them that I want them to learn and that I care about who they are: A microethnography of the social organization of a secondary low-track English-reading classroom. *American Educational Research Journal*, 26 (2), 227-259.
- Fenstermacher, G.D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, 107(1), 186-213.
- Floden, R. (2001). Research on effects of teaching: A continuing model for research on teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (4<sup>th</sup> ed., pp. 3-16). Washington DC: AERA.
- Goldhaber, D., & Anthony, E. (2003). *Teacher quality and student achievement*. New York: Teachers College Press.
- Good, T., & Brophy, J. (2003). *Looking in classrooms* (9<sup>th</sup> ed.) Boston: Allyn & Bacon.
- Greene, J.C. (2001). Mixing social inquiry methodologies. In V. Richardson (ed.), *Handbook on research on teaching* (4<sup>th</sup> ed., pp. 251-258). 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.
- International Reading Association (IRA) & National Council of Teachers of English (NCTE). (1996). *Standards for the English language arts*. Newark, DE & Urbana, IL: International Reading Association & The National Council of Teachers of English.
- Knapp, M. S. (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.
- Ladson-Billings, G. (1994). *Dreamkeepers: Successful teachers of African-American children*. San Francisco: Jossey-Bass.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven: Yale University Press.

- Lissitz, R.W. (Ed). (2005). *Value added models in education: Theory and applications*. JAM Press: Maple Grove, MN.
- McCaffrey, D., Lockwood, J., Koretz, D., & Hamilton, L. (2003) *Evaluating value-added models for teacher accountability*. Santa Monica, CA: Rand Corporation.
- Merriam, S.B. (1998). *Qualitative research and case study applications in education (2<sup>nd</sup> edition)*. San Francisco: Jossey-Bass.
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Newmann F., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey-Bass.
- No Child Left Behind Act, (2001). Public Law 107-110.
- Oser, F.K., Dick, A., & Patry, J. (Eds.). (1992). *Effective and responsible teaching. The new synthesis*. San Francisco: Jossey-Bass.
- Porter, A. (2002). Measuring content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.
- Raudenbush, S.W. (2005). Learning from attempts to improve schooling: The contribution of methodological diversity. *Educational Researcher*, 34(5), 25-31.
- Rice, J.K. (2003). *Teacher-quality. Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.
- Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (1998). Teachers, schools, and academic achievement, National Bureau of Economic Research, working paper 6691.
- 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(8), 1525-1567.
- 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.
- Sarason, S. B. (1999). *Teaching as a performing art*. New York: Teachers College Press.

- Schwab, J. J. (1973). The practical 3: Translation into curriculum. *School Review*, 81 (4), 501-522.
- 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.
- Stake, R.E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Strauss, A.L. (1987). *Qualitative analysis for social scientists*. Cambridge, England: Cambridge University Press.
- Vygotsky, L.S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Valli, L., Croninger, R., & Walters, K. (2005). Who is the teacher? Cautionary notes on teacher accountability systems. Unpublished manuscript.
- Whitehurst, G.J. (2002 June). Research on teacher preparation and professional development. Paper presented at the White House Conference on Preparing Tomorrow's Teachers.