

Studying High-Quality Teaching in a Highly-Charged Policy Environment

Linda Valli, Robert Croninger, Daria Buese,
University of Maryland

Abstract

This paper examines ways in which the current policy context influences teaching as well as challenges these influences pose for research on teaching. Drawing on a study of 4th and 5th grade reading and mathematics instruction, we argue that educational policy produces certain types of change in teaching, and the context of teaching, that have not been widely discussed in the literature. To illustrate this potential policy impact, we focus on three core dimensions of teaching: the teacher, time, and teaching quality. Our findings indicate under-reported, multiple influences on student learning due to both local and federal policies; unexpected changes in what is taught and what kind of teaching occurs within the school day, throughout the year, and across years; and narrowing conceptions of teaching quality. Policy influences on these core dimensions challenge researchers to more carefully specify who is responsible for teaching whom, maintain flexibility in data collection, retain research participants across years, and promote alternative conceptions of quality teaching.

DRAFT: PLEASE DO NOT CITE OR QUOTE WITHOUT AUTHORS' PERMISSION

Studying High-Quality Teaching in a Highly-Charged Policy Environment

Policymakers have long been interested in education as a mechanism for addressing social problems and promoting the public good. Through a range and mix of “policy tools” (Floden, 2003), policymakers attempt, with varying degrees of success, to influence components and outcomes of the educational system, including teacher qualifications and expertise; educational programs, materials, and resources; and student achievement (Hannaway & Woodroffe, 2003). What has changed in recent years is what some have argued is an unprecedented shift in control over public schools (Epstein, 2004). Once almost the exclusive domain of local school districts, elementary and secondary schools are affected to ever greater degrees by outcome-oriented policies determined at the federal and state levels (Croninger, Valli, & Price, 2003).

The federal No Child Left Behind Act (NCLB, 2001), for example, requires that schools make adequate yearly progress (AYP) in students’ standardized achievement test scores and that teachers meet its definition of being highly qualified, while the state determines standardized test content and criteria for proficiency. At the same time, local school districts continue to develop their own curriculum, tests, and programs. In some instances these three policy levels function in concert, providing school administrators and teachers with a coherent picture of expectations and education in the public interest. But frequently, local, state, and federal policies collide, putting conflicting pressures on school personnel and pulling them in different directions (Cohen, 1995).

These conflicting messages help explain the difficulty policymakers and reforms have in effecting lasting and substantive change in instruction (Cohen & Spillane, 1992). But, because of the increasing pressures from local, state, and federal levels—particularly

policies with high-stakes consequences—schools and teachers are no longer insulated from external mandates. Our study of 4th and 5th grade reading and mathematics instruction, conducted in the context of rapidly changing and increasingly powerful policies, suggests that the current policy context has, indeed, changed the practice of teaching. While the scope and significance of these changes are not yet clear, they warrant scrutiny.

In this paper, we examine ways in which the current policy context influences teaching and the challenges these influences pose for research on teaching, especially research about the elusive notion of quality in teaching.¹ We argue that educational policy produces certain types of change in teaching—and the context of teaching—that have not been widely discussed in the literature. To illustrate this potential policy impact, we focus on three core dimensions of teaching namely:

- *agency*—who is doing the teaching,
- *time*—when the teaching of particular content occurs, and
- *quality*—what characteristics of teaching are thought to be worthwhile.

Challenges posed for research by this new policy environment are practical, methodological, and ethical. We need to understand more clearly: just who is responsible for student learning; how teaching varies during the school day, throughout the school year, and across years; and what sustains rich, complex views of teaching qualities and goals.

RESEARCH METHODOLOGY

¹ This paper is based on research supported by the Interdisciplinary 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.

This paper draws on data from a four-year study (2001-2005) of 4th and 5th grade classes in reading and mathematics. The purpose of that study was to learn more about what teachers and programs do to assist students who are struggling to acquire foundational skills in these subject areas. Although foundational skills are typically associated with primary grades, we selected 4th and 5th grades because there is limited research in this area (Chambliss & Graeber, 2003) and yet, as indicated by NAEP results, large percentages of student in these grades continue to score below basic levels. Because of our particular interest in teacher success with low-achieving students, we located the study in one of the largest and most diverse school districts in the country; approximately 70 teachers from 11-18 schools participated each year. We selected schools with moderate to high levels of poverty (30-90%) that had higher than expected levels of student achievement, that were nominated by knowledgeable insiders, or with whom we had professional ties.

Primary forms of data collection included classroom observations with time-sampling protocols, teacher daily logs of curriculum coverage, interviews with principals, teachers, and educational specialists, and statewide tests of student performance. As discussed later in the paper, standardized protocols were informed by, but not limited to, our conceptions of high-quality teaching derived from research on reading and mathematics instruction (Chambliss & Graeber, 2003), cognitive psychology (Alexander & Murphy, 1998), and policy-practice relationships (Croninger, Valli, & Price, 2003). One construct— *cognitively-demanding instruction*—is used to illustrate the ways in which policy influences practice. We distinguished between high and low cognitive demand in teacher activities, student activities, and lesson content, by using data from the

time sampling observation instrument. *Student simple response*, for example, was counted as low demand while *student conjecture* was high demand.

We also tracked policy changes that affected the teachers in our study. During the four years of data collection, significant policy changes occurred at all governance levels. The federal government passed the No Child Left Behind (NCLB) Act of 2001, which mandates annual testing of every student (Grades 3-8) in reading and mathematics and imposes clear sanctions for lack-of-compliance. The state replaced its own performance-based testing program—which reported results at the school, rather than the student, level and was limited to just three grade levels—with high-stakes tests for grades 3-8 that complied with NCLB guidelines. It also established proficiency standards, set AYP benchmarks, and developed a voluntary state curriculum aligned with the state-test content at each grade level. In the meantime, the school district developed its own new curriculum frameworks in reading and mathematics, required schools to give the unit tests in mathematics at specified times during the school year, and offered professional development opportunities linked to the new curricula. As seen below, this confluence of policy changes affected core dimensions of teaching: who was teaching whom, what was taught when, and what was regarded as important characteristics of teaching.

THE TEACHER DIMENSION

Who is allowed to teach which students is a matter of public interest. As public institutions, schools try to guarantee local communities that their children will be safe and taught by competent professionals by enacting standards for teacher licensing and assignments. Each state sets licensing standards and school districts attempt to limit out-of-field assignments, recognizing that different qualifications are needed to teach

different subject matters to students of different ages. As a recent manifestation of the concern for teacher quality, the federal No Child Left Behind Act of 2001 established incentives for states to limit out-of-field assignments and to enable increasing numbers of teachers “to meet certification, licensing, or other requirements needed to become highly qualified” (NCLB, 2001, Sec. 2113).

Research bolsters this public interest in teacher quality. With a growing body of evidence suggesting the importance of teacher quality for student learning (Darling-Hammond, 1997; Goldhaber & Anthony, 2003; Rice, 2004; Sanders & Horn, 1998), policymakers have promoted teacher evaluation and accountability systems linked to student test scores (Valli, Croninger, & Walters, 2004). The National Governors Association has urged states to “consider measurable student achievement as a principal outcome on which teachers are evaluated” (Goldrick, 2002, p. 4). At least 14 states, as well as numerous schools districts, use or have explored using student performance data as part of teachers’ performance reviews (Boser, 2001; Bradley, 1999; Goldrick, 2002; Schemo, 2004; Wainer, 2004).

These teacher accountability policies are based on the assumption that teaching is an individual endeavor in which one teacher has sole responsibility for a group of students, what Lortie (1975) called the “egg-crate” classroom. They further assume that statistical methods can clearly sort out what is implicitly considered to be rare and easily identified patterns of co-responsibility such as team teaching. While numerous scholars have questioned the statistical assumptions underlying teacher accountability systems, (Ballou, 2002; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Raudenbush, 2004),

less attention has been paid to the assumption that classroom teaching is still an individual enterprise.

But data we collected suggest that a convergence of policies, such as accountability, data-based decision making, inclusion, and professional development, has changed the traditional relationship between teacher and students, with multiple adults now influencing student learning within core subject areas. Teaching has often become a collective rather than an individual enterprise in all three phases of teaching: planning, instruction, and assessment. Students are temporarily or permanently re-assigned teachers during the school year; they can stay in their original classroom but be taught by someone other than the teacher of record; and they can receive supplemental lessons, from a different teacher during the day, week, or marking period. These instructional arrangements have implications for both teacher accountability policy and teacher research.

TRACKING MULTIPLE TEACHER INFLUENCES

To provide a more systematic picture of the nature and extent of the multiple-teacher influences on student learning, we summarize data from the 18 schools and 67 classroom teachers that participated in the third year (2003-04) of the study.² We begin with some “regularities” of schooling—teacher absences, student mobility, and annual testing schedules—and then take a close look at collaborative instructional designs.

Teachers, like all workers, get sick or are absent from the classroom for other personal or professional reasons. Based on daily logs data, the average amount of time someone other than the assigned teacher had responsibility for instruction due to absences was roughly 7%. Student mobility rates, which track student enrollment in more than one

² We focus on 2003-04 because of additional, relevant resource data we were able to collect that year.

school during a school year, also attenuate the relationship between individual teachers and students. During the 2003-04 school year, the average student mobility rate in these 18 schools was 20%.³ If the students in our 4th and 5th grade sample were similar to the general population of students in these schools, one out of five had at least one other teacher, in a different school, who was responsible for their learning during that academic year. Although not policy-related, these normal fluctuations in classroom participation are rarely accounted for in teacher accountability models.

In addition, because of public notification requirements introduced by NCLB, the district we studied scheduled end-of-year tests in March so that parents could be provided with test results prior to the beginning of the next school year. In other words, every 2003-04 teacher in the study shared instructional responsibility for achievement gains with at least one 2002-03 teacher. Given the district's 9.5-month school calendar, this amounts to roughly one-quarter of students' "tested" instructional time. These percents—7% time due to teacher absences, 20% student mobility rates, and 25% instructional time received in the previous year—begin to create a picture of significant multiple, rather than singular, influences on student learning.

To account for multiple influences on student learning due to collaborative instructional design models, we created teacher and student databases indicating the types of school personnel besides the assigned classroom teacher who had instructional responsibilities for the students in our study. As indicated on Table I, more than two-thirds of the students in the reading classes (67%) and more than half the students in the mathematics classrooms (55%) received instructional assistance from more than one

³ Mobility rate was determined by the % of students who transferred into and out of the school in relation to total student enrollment.

educator. This assistance could be provided by individuals with quite varied degrees of expertise during class instruction or in addition to regular class instruction. Students could also be reassigned to a different classroom. All but three of the 67 teachers in this analysis reported having students who received some sort of additional instructional influence. These practices are not insignificant. Instead, this ensemble of instructional agents embodies a phenomenon of distributed responsibility for instruction.⁴ When viewed cumulatively and in relation to the teacher absence, student mobility, and test schedule data, one begins to get a sense of the complexity of teacher-student linkages and the unrealistic assumptions that characterize many teacher accountability proposals (Valli, Croninger & Walters, 2004).⁵

Table 1
Types of Personnel Providing Instruction

<i>Personnel Type</i>	<i>Reading</i>	<i>Mathematics</i>
In-class para-educator influence	30%	27%
In-class student teacher influence	17%	19%
In-class staff developer/ resource teacher	15%	13%
In-class, individualized specialist assistance	8%	7%
Out-of-class assistance	18%	6%
Re-assigned to other teacher	7%	1%
One or more of the above	67%	55%

Furthermore, with the school district emphasizing data-based decision-making, teachers worked collectively on planning lessons and assessing student learning, sharing responsibility for teaching and learning with each other and the school’s professional staff. Although this sharing took place within increasingly standardized expectations to

⁴ We are not making any claims about coherence by using this concept.

⁵ An additional source of influence on student learning is summer school, which varied across schools and students. Because we were not able to track individual participation, we did not include that source of influence in our analysis. Similarly, pull-outs for “specials” such as band practice, have a variable influence on the links between teacher performance and student achievement. Our anecdotal data suggest that this does not occur evenly across teachers, giving some teachers an additional instructional burden.

follow the district's curriculum, the amount and quality of shared work varied across teachers and schools. In some classrooms, para-professionals were of great assistance; in others, they interfered with the smooth delivery of instruction. In some cases, even within the same school, new teachers had high-quality mentors and/or grade-level teams to help them plan and learn to assess student work. But in a few cases, these supports were missing.

One indication of the potential benefit of teacher collaboration outside the classroom is the similarity in teaching practices irrespective of years experience. If planning, instruction, and assessment were singular activities, we would expect, as previous studies have shown (Rice, 2003), that neophytes would demonstrate less expertise than experienced teachers. That was not the case for the participants in our study. Using cognitively-demanding instruction as a proxy for expertise, we mapped the percentage of time mathematics teachers made requests of students that were judged to be of high or low cognitive demand along with students' responses. As seen in Figures 1, the profiles of teachers with only 1-5 years of experience were quite similar to those of the experienced teachers. Having observed the joint planning and mentoring that occurred in these schools, our best hunch is that these collaborative activities were at least partially responsible for the similarities in cognitively-demanding instruction.

****Figure 1 about here****

CHALLENGES FOR RESEARCH

The challenge these complex instructional arrangements pose for researchers is to determine just who is responsible for teaching. In order to untangle teacher and school effects, researchers need to suspend assumptions about "egg crate" schools and more

carefully specify complicated influences on student learning. This will help them determine whether these variations are just marginal noise, peripheral to the assigned teacher's influence on student learning, or of such significance that they call into question basic research assumptions and measures of teacher effects. If the patterns and degree of multiple influence we found, especially in the higher-poverty schools, are at all typical, we suspect that a considerable proportion of school effects are being attributed, positively or negatively, to individual teachers. But because of informal arrangements made among teachers and deviations from formal expectations, researchers should be skeptical of using school or school-district level databases on teacher-student assignments. Although gathering these data from teachers is more difficult, our study has persuaded us that it is a necessary task.

As we will discuss later, we are not arguing that good teachers do not matter, only that good teaching is distributed among students in more complex ways than is expressed in policy language and proposals. The use of research to shape educational policy, especially policy that has high-stakes consequences for individual teachers, places a strong ethical responsibility on researchers to specify carefully multiple in-school influences in models that are designed to predict student outcomes.

To begin to account for these multiple variations and influences we developed a web-based system for teachers to roster their students for classes in reading and mathematics. We used this system (and more conventional paper-and-pencil checklists) to take two "snapshots" of students assigned to teachers during the year. We also asked teachers to notify us of any changes in their rosters; to provide us with information about routine instructional assistance that they received from staff developers, para-educators,

or student teachers; and to identify individual students who received regular resource help in or out of the classroom. Because teachers were keeping daily logs for the study, we were also able to account for days each teacher was absent from the classroom. This compilation of data enabled us to track, with a certain degree of confidence, the network of influences on students' learning (Valli, Croninger, & Walters, 2004).

But even if greater precision in separating teacher effects from school effects is possible, our study has made us cautious about the use of these techniques for purposes teacher accountability. While teacher-level research can serve teachers, schools, and students by identifying individual teachers' professional development needs, it can inadvertently support accountability systems that undermine school-based, collaborative learning communities. Data showing benefits of collaboration to new teachers in this study, as well as other studies of teacher collaboration suggest that this might not be in the best interest of either teachers or students (Darling-Hammond, 1997; Little, 1999; Smylie, 1994). If teachers are held individually, rather than collectively, responsible for student learning, they would have little incentive to work together as teams, to develop shared goals and ways of analyzing student work, or to serve as mentors and peer coaches.

THE TEMPORAL DIMENSION

Similar to the complexity involved in determining agency, we found considerable complexity in school schedules. We could not assume that mathematics and reading would be taught in, or only in, the designated part of the school day, or that teaching practices would be similar at different times during the school day. We also found unanticipated changes in teaching practices during the school year and across school

years. The four-year time frame for data collection, enabled us to observe changes in teaching practice and content coverage related to the changing policy context. This context influenced the way we thought about the temporal dimension of teaching, making us even more convinced of the importance of studying teaching over time and at strategic times.

Given research on the stability of teaching, this focus on change might seem like an odd, even distracting choice. Numerous scholars have claimed that teaching changes only in non-substantive ways and that changes do not persist (Cohen, 1988; Cuban, 1993; Tyack & Cuban, 1995). Another tradition, however, asserts that teachers change all the time, in their thinking, teaching and learning, and that “these changes may be prompted, promoted, or supported by discussions with other teachers, an evaluation by an administrator, a workshop, experience with an often-tried activity that no longer works, an article in a practitioner or research journal, a new grade level or population of students, etc.” (Richardson, 2001b, p. 908).

Scholars have addressed these competing claims by examining more carefully the object, type, and context of change being discussed. A growing body of evidence indicates that content coverage, or *what* gets taught, is more easily changed than instructional practices, or *how* teachers engage students with that content (Floden, 2001); that incremental change happens, but not radical change (Cuban, 1988); and that change is more likely to persist when initiated, or at least owned, by teachers themselves—when it is naturalistic (Richardson, 2001b). The problem with these distinctions is that they fail to capture the way change occurs in the everyday world of schools. We contend that in the current policy environment, with multiple mandates and demands placed on teachers,

what is taught profoundly influences *how* it is taught; that the line between incremental and radical change is hard to determine, especially in its cumulative impact on students; and that even if teachers do not *own* the change effort, policy pressures are now so intense that teachers can no longer insulate themselves from their influence.

In the following sections, we describe some of the changes we observed in the teaching of reading and mathematics during the school day, throughout the school year, and across years. We discuss the multiple policy influences on those changes and the implications for research design.

STUDYING TEACHING DURING THE SCHOOL DAY

Teaching changed during the school day, mostly in response to targeting student needs as they related to testable content and skills. At the beginning of the school year, administrators typically established a simple comprehensive schedule showing what teachers teach what subjects at what grade levels at what time. We soon discovered that we could not rely on those official documents to determine the timeframe for reading and mathematics lessons. Both subjects were taught (and often quite differently) outside the designated time and often by different individuals.

It is common to think of lessons as discrete, time-bounded segments of the day, in which teachers orchestrate students' engagement with a specific subject for a set of instructional purposes. Of course, the more discrete and bounded the instructional event, the easier it is for observers to capture teacher and student engagement in a particular lesson, and the easier it is for teachers to describe clearly and unambiguously curricular coverage as part of their daily logs. This model generally fit mathematics instruction in the classrooms that we have observed, where teachers had regularly set lessons that

ranged from 60-90 minutes. However, as testing time drew near, homeroom teachers would frequently conduct “test prep” sessions at the beginning of the day that focused on test content and format. Although generally commissioned by the school administration, these supplementary lessons seldom became a formal part of the school schedule. We happened upon them by chance and did not have the resources or authority to track them as part of our research agreement.

The model of a bounded lesson fit even less well in reading/language arts classes, where instruction occurred in varying blocks of time and was often blended into other subject areas. Variations across schools and changes over time in reading/language arts instruction made consistency in identifying and recording lessons more difficult than we imagined. In some of our schools, 90-120 minutes a day were set aside specifically for instruction in reading/language arts, although students and teachers deviated from this schedule roughly once a week to accommodate “specials.” In other schools, 60 minutes were set aside for reading in the morning and another 60 minutes for writing in the afternoon (or vice versa), and in still other schools, teachers routinely incorporated reading/language instruction into social studies or science instruction, essentially extending the “official” reading/language arts block into other parts of the day.

In many of the schools students engaged in “Bell Work,” usually a review of grammar, punctuation, spelling, and sentence structure but sometimes oral reading, while beginning-of-day routines took place. And, as indicated in Table 1, numerous students (almost one out of five in Year 3 of the study) had supplementary reading instruction outside the regular reading block. As with mathematics, the amount of supplementary instruction in reading seemed to be related to assessments. Schools in greatest danger of

failing to meet AYP were most likely to add supplementary reading instruction, with “reading intervention” mandated by the school district for 18 Title I schools and strongly encouraged for others. Eight of the 18 schools in our study had some form of reading intervention, although the numbers of targeted students as well as the nature of the intervention varied considerably from school to school.

Table 2 illustrates the importance of studying instruction as it occurs outside the official time block, by showing just how different the reading intervention classes were from regular reading classes. The percents are based on data from the three schools in which we observed reading intervention as well as regular reading classes. These schools were at the high end of the poverty and ESOL range of our sample and were selected for more in-depth policy analysis. Eight of the eleven intervention classes were taught by the regular classroom teachers, the other three were taught by reading and ESOL specialists. As can be seen by comparing data down the columns, intervention classes were less likely to engage students in cognitively-demanding work.

Table 2
Reading Intervention and Reading Class Comparison

	<i>Regular</i> <i>N = 143*</i>	<i>Intervention</i> <i>N = 56*</i>
Teacher Hi Cognitive	12%	6%
Teacher Lo Cognitive	19%	35%
Student Response/Work Hi Cognitive	17%	10%
Student Response/Work Lo Cognitive	26%	40%
Lesson Content Hi Cognitive	28%	20%
Lesson Content Lo Cognitive	11%	24%

**N = number of lessons*

So even though all teachers and observers were given the same instructions, to record the 60-90 minute reading and mathematics lessons presented to students in a particular class, we could not be certain that we always captured the type of instruction

students received. We realized that if we were going to get a comprehensive picture of reading instruction, in particular, we would need to study what went on outside the official block designated for reading. We employed a range of strategies to address this ambiguity, particularly in our observations of reading/language arts lessons. First, when observers scheduled observations, they privileged reading instruction in determining a time to observe teachers' classes. Doing so provided some consistency across observations and increased the likelihood that we would capture that part of the reading/language arts block that focused most on reading. Second, we relied on teachers' logs, which are typically a reflection of what a teacher did during the day in a curricular area to capture aspects of a lesson that we may not have had an opportunity to observe. Third, we interviewed teachers and principals about supplementary instruction that occurred outside class time. And last, in opportune situations, we devoted resources to observing supplementary instruction in both reading and mathematics so we could understand better the nature of that instruction and how it related to students' regular class lessons.

STUDYING TEACHING DURING A SCHOOL YEAR

Just as the teaching of mathematics and reading varied within the school day (when subjects were taught, what was taught, for what purpose, and how), it also varied across the school year. As would be expected, differences in mathematics and reading lessons were driven by the school-district curriculum. Certain topics were scheduled to be taught at different times and one would expect the official curriculum policies of the school district to have a strong influence on what was taught when. So it would be no surprise if

teachers' daily logs showed that they taught fables and place value to 4th graders in October and poetry and probability to 5th graders in May.

We were surprised, however, at some of the variations in lesson content across the year in time-sampling data. In the reading protocol, the primary content categories were reading, writing, and conventions. For mathematics, the primary content categories were procedural, conceptual and linking instruction. As seen in Figures 2–5, content coverage for 4th and 5th grade showed different trends in these broad areas of literacy and mathematics instruction across the school year.

Figures 2-5 about here

Unlike the teaching of poetry, or statistics, or fables, or probability, emphasis on reading vs. writing instruction or on conceptual vs. procedural knowledge, does not seem to be driven by the official curriculum. Analysis of the curriculum frameworks provided no clue as to why reading instruction would drop off for 4th grade in November and May and climb for 5th grade in March. Nor did it help us understand why the teaching of procedural knowledge would be more dominant in February – March than the rest of the school year. Does the teaching of multiplication, division, estimation, and time (the core topics during that time period) need more procedural emphasis than the teaching of fractions, decimals, and measurement (core topics taught during conceptual peak times)? Or were factors other than curriculum topics at play? Informal observations of principal-teacher interactions suggest this was a possibility. At one school, for example, the principal explicitly told teachers to set aside the school-district curriculum in February in order to focus on the state's test topics they had not covered. Concerned that the lack of alignment between the district's curriculum and the state test would result in low test

scores and inadequate yearly progress, 4th and 5th grade teachers simply ignored the school-district curriculum, arguably a rational strategy for any school struggling to make AYP.⁶

STUDYING TEACHING ACROSS YEARS

In addition to changes during the school day and within the school year, teaching also changed across years as the new curricula and assessments replaced former ones and as high-stakes accountability became more pressing. This suggests that assertions about the stability of teaching practice might be overstated, or that teaching practice is changing more noticeably due to increased policy pressures. Given high-stake consequences, teachers may no longer be able to close their doors to a barrage of policy directives as they once did. In particular, our classroom observations indicated that teaching became more basic-skill oriented once testing was required at every grade level and had high-stakes consequences. Achieving AYP was especially challenging during the 4th year of the study, 2004-05. Compared to the previous year (2003-04), when schools were required by the state to increase the percent of students in the proficient category by only 2.5% in reading and 2.7% in mathematics, 2004-05 proficiency ratings had to increase by 11.5% in reading and 9.5% in mathematics.

Table 3 shows a three-year comparison of indicators of cognitive demand in mathematics and reading, from our total sample of teachers, 64-76 depending on the subject and year. There was a decline in every category that signified cognitively-complex instruction and either a rise or constancy in instruction that placed little cognitive demand on students. In mathematics, the change was greatest between 2003-04

⁶ Another reason to think that factors other than the official curriculum were at play is that patterns across three years varied considerably. Since the district's mathematics curriculum was the same during these years, we would have expected to see similar patterns of variation.

and 2004-05, supporting our hunch that the press to make AYP was particularly strong that year, despite teachers being in the third year of a new mathematics curriculum that should have supported cognitively-demanding instruction. In reading, changes were less dramatic, but followed the same basic pattern.

Table 3
Incidence of High Cognitive Demand Across Years (*Lesson Level*)

	<i>Mathematics</i> ⁷			<i>Reading</i> ⁸		
	2002-03 (<i>N</i> = 266)	2003-04 (<i>N</i> = 606)	2004-05 (<i>N</i> = 539)	2002-03 (<i>N</i> = 266)	2003-04 (<i>N</i> = 552)	2004-05 (<i>N</i> = 536)
Teacher High Cognitive Demand	11%	11%	7%	14%	10%	10%
Teacher Low Cognitive Demand	21%	20%	20%	13%	17%	17%
Student Response/Work High Cognitive Demand	18%	17%	10%	20%	18%	16%
Student Response/Work Low Cognitive Demand	33%	32%	39%	24%	24%	26%

Two distinct factors might account for the difference between mathematics and reading: (a) domain differences, reflected in our construct decisions, in which cognitive demand had more coding choices in reading and (b) the introduction of the new reading curriculum in 2004, which required more frequent use of expository text, classified as high cognitive in our constructs. Thus, our coding decisions and the new reading curriculum, might be mitigating the effects of high-stakes testing on changes in cognitive-demand in reading classes. Table 1, which compares regular reading vs. reading intervention classes might provide a clearer picture of what happens to literacy

⁷ Mathematics teacher and student high cognitive demand 2002 and 2003 are significantly higher than 2004 at .01 level; student low cognitive demand 2004 is significantly higher than 2002 and 2003 at .01 level.

⁸ Reading teacher high cognitive demand 2002 is significantly higher than 2003 and 2004 at .01 level; student high cognitive demand 2002 is significantly higher than 2004 at .01 level; teacher low cognitive demand 2003 and 2004 are significantly higher than 2002 at the .01 level; student low cognitive demand 2004 is significantly higher than 2003 at the .05 level.

instruction when it is driven by high-stake expectations. We return to this issue in reading classes in the next section of the paper.

CHALLENGES FOR RESEARCH

Studies of teaching are increasingly conducted in an environment characterized by high-stakes accountability. School personnel are under intense pressure for students to do consistently better on standardized achievement tests and to close the persistent gap between higher and lower achieving groups. These pressures impact the work of researchers in many ways: teachers are apprehensive about being observed, concerned about the uses and abuses of data, protective of their time, and want immediate feedback for their efforts. In addition, demands for accountability can result in dramatic changes in expectations for teachers and in the formats for assessing outcomes, creating dilemmas about instrumentation and proposed analytic models (Valli, Croninger, Alexander, Chambliss, Graeber, & Price, 2004).

For researchers, changes in curriculum coverage during the school year pose challenges in both the timing of data collection and the level at which data are collected and analyzed. Understanding how much time students spend reading, for example, is of major interest to reading researchers. This study indicates that the answer to that question would vary considerably, depending on the time of year a study was conducted. Similarly, there is much interest among mathematics researchers in teaching that helps students link conceptual understandings to the mathematical procedures they employ. But classroom researchers would get very different estimates of procedural vs. conceptual vs. linking approaches to teaching mathematics if they collected data in November rather than February.

An additional research challenge suggested by these data is tracing the sources of change in teaching. Are changes due to policy expectations embedded in curriculum frameworks, to principal expectations, to state test expectations—or a combination of the above? Although we expected the policy environment to have an effect on the teachers in our study, we did not fully appreciate the manner in which the policy environment would intensify some of the challenges associated with a longitudinal study of teaching in moderate to high-poverty schools. While longitudinal studies seem to be necessary to understand the impact of policy on teaching, the current policy context makes those studies more difficult to accomplish. Given accountability pressures, principals and teachers are reluctant to commit over an extended period of time. Even if willing, they often change schools or grade levels (with frequent moves out of high-stakes contexts), making the study of changes in teaching most difficult in the contexts about which we need to learn the most. Knowing in advance to expect some turnover didn't prepare us for the magnitude of the turnover rate. The original research design called for following roughly 60 4th grade teachers during Years 1-3 and 60 5th grade teachers during Years 2-4. This would enable us to follow 4th graders across two school years.

Although we had a total of 25 schools participate across the four years, only 5 participated the entire four years, 6 participated for three years, 9 for two years, and 5 for only one year. Despite incentives to participate, schools dropped out of the study because of fears that participation would detract from more high-stakes priorities. Longitudinal data at the teacher level also fell far from our goal. While approximately 70 teachers participated each year, only 20 participated for the three years called for in the research design. Attrition was due to personal factors (e.g., health, maternity/family leave), job

factors (i.e. taking a position that fell outside sampling parameters), and stress factors (i.e. feeling too overwhelmed by other responsibilities to participate in the study). While personal factors affect teacher mobility at all schools, the AYP pressure in higher-poverty schools creates a high degree of mobility and a tendency to reject activities that do not directly and obviously serve that one goal. The monetary, communication, and feedback incentives we developed to recruit and retain participants had some positive effect on participation. However, they did not resolve an essential problem associated with research on teaching in the current policy environment—that principals and teachers are under mounting pressure to increase annually student performance. Participation in a study such as this is an investment with unknown returns. Such an environment is not conducive to long-term commitments. Ironically, contexts that are especially important for current policy interests in closing the achievement gap are the hardest to maintain. The very policies that give urgency to certain types of studies are the very ones that make those studies difficult to carry out.

THE QUALITY DIMENSION

The policy environment also influences what is considered to be high-quality teaching. Despite lists of “best practices,” exactly what constitutes good teaching has long been contested (Bruner, 1999; Chambliss & Graeber, 2003). Does high-quality teaching have generic or context-specific characteristics? Does it reside in the cognitive complexity of subject matter or in instructional practices? And where do competing notions of good teaching come from? How does one know what constitutes high-quality teaching? What are the sources of that knowledge?

Researchers part ways on these questions, sometimes quite dramatically. One group claims that images of good teaching should be derived from expert opinion—from standards, norms, theories, and nominations of good teachers. These experts might have different disciplinary groundings, views of learning, or social goals that shape their conceptualizations, but they all begin from a priori judgments. Shulman (1986) uses the word “correspondence” to describe this research approach in which “a given exemplar of instruction is compared to a model or conception of good teaching derived from a theory or ideology” (p. 28). A problem with this tradition is that, as the reading and math wars so vividly portray, agreement on any given set of standards is hard to achieve.

In contrast to the correspondence approach, another group of researchers looks, not to standards for teaching, but to the consequences of teaching. Teachers’ work is considered to be good if it produces desired outcomes, usually students’ academic learning. Shulman (1986) calls this a “pragmatic” or “correlative” conception of teaching effectiveness. In this conception, teaching and learning are closely connected, with students’ standardized test scores used as the measure of high-quality teaching. But, as its critics charge, the tests themselves could assess a very narrow range of learning, and achievement gains—as measured by these tests—could result from practices antithetical to most conceptions of good teaching, practices such as teaching to the test, narrowing the curriculum, or manipulating the population of test takers (Darling-Hammond, 1997; McNeil, 2000).

Partial resolutions to these dilemmas have been offered by differentiating components of quality teaching. One construction employs the term “successful” teachers to mean those who are both effective and responsible, who enact their roles in ways that

bring about “good” learning as well as human satisfaction, who attend to both the technical and moral components of the teacher-student relationship (Oser, Dick, & Patry, 1992). Another construction limits the term successful to teaching that produces learning (Oser et al.’s effective teaching), and offers the term “good” teaching as that which “comports with morally defensible and rationally sound principles of instructional practice” (Fenstermacher & Richardson, 2005, p. 189). In this framework, successful teaching would be equated with the correlative and effective definitions given above, and good teaching would be similar to the correspondence and responsible versions. While successful teaching is, what Fenstermacher and Richardson (2005) call, learning-dependent, achieved only with evidence of student learning, good teaching is learner-sensitive, evidenced by students’ substantive responses, motivation, and engagement in the lessons.

As a way of synthesizing these competing conceptions, Fenstermacher and Richardson (2005) use the word quality to depict teaching that is both good and successful. But this synthesis comes with two strong cautions. The first is that quality in teaching is dependent on more than teaching practices; that it is more likely to occur “when there is willingness and effort on the part of the learner, a supportive social surround, ample opportunity to learn, and good practices employed by the teacher” (p. 191). The second caution is that quality teaching looks quite different in various approaches to teaching and programs of research on teaching (e.g., teaching as transmission, as cognition, and as facilitation).

REPRESENTING QUALITY IN TEACHING

Recognizing these cautions and conflicting approaches to research on teaching, we made four important decisions in the design of the high-quality teaching study. First, we were explicit about our own conceptions of high-quality teaching via an attribution scale that rated teachers on the prevalence of significant teaching practices (e.g. Alexander & Murphy, 1998). Second, we developed instruments and protocols that enabled us to temper our personal judgments about pedagogical practices and consider other conceptions of good teaching. Third, we asked teachers about their views of high-quality teaching rather than just rely on our own. And fourth, we collected student achievement data, but did not equate it with high-quality teaching, asking instead, “What is the correspondence between our constructs of high-quality teaching and student achievement?”

We found that, with the current policy emphasis on high-stakes accountability, the pull is toward a correlative definition of high-quality teaching, where quality is judged by its power to raise student achievement scores. Increasingly, teachers matched curriculum coverage with test content and used a variety of means to enable their students to do well on tests. Among these strategies were: focusing resources on “bubble” students, those just below the proficiency standards; providing better accommodations for special needs students during test time; aligning student assignments with assessment formats; and regularly assessing specific skill areas in order to provide targeted instruction and give students practice with test-taking skills. These changes were evident in both teacher talk and teaching practices, which we discuss below.

In focus group interviews conducted across three academic years, teachers talked about the policy influences on their practices and ways of thinking about high-quality

teaching. New mathematics and reading curricula, along with new district and state assessments, professional development around curriculum implementation and corresponding instructional practices, the monitoring of teacher performance, and block scheduling were just a few interconnected district initiatives that the teachers experienced. The sum of these initiatives created considerable dissonance in the ways teachers thought about their teaching although their responses were primarily ones of accommodation. Most teachers we spoke with felt pressure from their principals and the district administration to teach in ways that were in opposition to what they believed to be best practices, but given the scope and intensity of the policy push in the district, compliance seemed, for most, unavoidable.

The most profound policy influences that affected teachers' daily work were inarguably the implementation of the new mathematics curriculum in 2002, a new reading curriculum in 2004, and the pressure of achieving AYP. The mathematics curriculum, simultaneously implemented in grades K – 5, was structured so that curricular concepts spiraled across grade levels with the intention that students would eventually achieve mastery of skills and understanding of concepts over time. As such, the curricular topics were arranged in an order and on a schedule that the teachers did not understand or agree with. One teacher complained, “And the thing about it is it really doesn't go with...meeting the needs of all the kids because it's expected to just move on and it's not based on mastery. And, I mean, that's one of those things that I've always felt really good about, that I've made sure that my kids have mastered it” (FG-C-03).

The six unit tests that accompanied the mathematics curriculum were also given on a strict schedule. Data from the unit assessments were a means to monitor both the

success of instruction and the achievement of students under the new curriculum. The teachers resented the imposition of the strict scheduling and the rigorous testing demands on students. One teacher asked, “as a teacher...does that mean that when I teach something and some of the students don’t get it I have to move on because I’m on a timeframe?” (FG-D-03). The teachers also questioned the effects of long, topically dense unit tests on their students, “Put a 20-page test in front of a fourth grader on five different major concept areas in math that are unrelated... It doesn’t make good, logical teaching sense (FG-D-03). Ultimately, many teachers, expressed anger over being held accountable for student learning when their instructional practices were so impacted by the curricular scope and structure:

What does the data provide? The data does you no good unless it’s accurate. It does no good unless it really reflects their true abilities. It’s not reflecting their true abilities, because you’re cramming it down their throat. It’s drive-by teaching. (FG-OC-04)

The new reading/language arts curriculum was implemented in grades 4 and 5 in 2004. Unlike the math curriculum, it was implemented up to grade three the previous year. Therefore, teachers did not expect the gaps in student knowledge and skills that seemed to accompany the concurrent, all-grades implementation of the mathematics curriculum. Like the math curriculum, teachers were expected to adhere to a schedule and the curriculum guidelines provided detailed lessons for each of the curricular topics. However, the new curriculum presented a greater teaching change than most anticipated. Many 4th and 5th grade teachers were not accustomed to using guided-reading groups like those that were commonplace in the primary grades. The new curriculum guidelines mandated this practice and many teachers felt ill prepared to accomplish it. Teaching reading also meant teaching to curriculum indicators (e.g., “identify and use text features

to facilitate understanding of informational texts”) rather than focusing primarily on the content of the text. The reading specialist at one of our schools who coached the teachers in enacting the new curriculum described this change in the approach to teaching reading:

You teach with the end in mind so you really need to know what that indicator is. And you really need to know...what...accomplishing that indicator look[s] like. So you need to know that instead of – you’re not just teaching that book. So you’re not just teaching that novel but what do I need to teach out of that novel... to teach this indicator, and this one, and this one? (McNeil, 5/17/05)

By the 2004-2005 school year, “curriculum indicators” became a key term in the language of instruction. Curriculum indicators were tested by the state assessment and the state explicitly specified which indicators from the state curricula would be tested. With the focus on AYP, covering the required indicators and teaching testing formats became instructional foci across all of our study’s schools. Exercises featuring short written responses that mirrored test formats and matched tested indicators were practiced repeatedly in math and reading classes. In one of our case study schools, teachers began to pride themselves on the high quality of their mandated test practice sessions:

We made [students] into analysts. Instead of just taking the test, going over the test, taking a test, going over – they had to analyze themselves. And we spent a lot of time in the reading block doing their [written] responses so in the afternoon (during the reading intervention period), whatever the focus was, poetry or context clues or whatever it was, then...they would have text to read and they would do a multiple choice on their own. And then we would put them in groups and give them three choices for the [short written response] that would match that reading and they had to score them. (BF-4th grade team, 05)

For these teachers, not only had high-quality instruction come to include good test preparation practices, but time for test preparation leached the reading block.

Faced with the normal implementation challenges, the district attempted to be methodical in its curriculum roll-out and uniform in relaying expectations to teachers.

However, unintended policy messages may have been sent. By promoting adherence to new curricula that required teaching concepts in an organization and at a pace that was unfamiliar (and often unacceptable) to teachers, using professional development that encouraged teachers to employ specific teaching methods and materials, and imposing uniform ways of monitoring teacher and student performance, the district's notion of high-quality teaching came to be seen as rigid and standardized. Teachers missed the autonomy of enacting curriculum in ways they believed best met the needs and matched the backgrounds of their students. In fact, the very notion of the teacher coming to know his or her students became formalized, based on increasingly detailed data from numerous district-wide achievement and diagnostic tests which were added on from year to year.

Parallel to their reflections on teaching, teachers' instructional practices increasingly privileged conceptions of teaching that were learner-driven, not learner-sensitive. For example, even when curriculum indicators were broad (e.g., "*Develop and apply comprehension skills through exposure to a variety of texts, including traditional print and electronic texts*"), teachers were told to have students practice specific reading comprehension skills. Soon, reading comprehension was reduced to the "The 12 Reading Strategies" (e.g., identifying main idea/details, cause/effect, problem/solution, and compare/contrast) and the new curriculum (four volumes long and more than four years in the making) was frequently supplanted by short workbook selections.

We suspect that a learner-driven orientation is also reflected in Table 3, where the cognitive demand of lessons significantly decreases over time. It is also seen in Table 1, which similarly depicts differences between instruction in regular reading classes and

reading intervention classes, which were specifically designed to help students score in the proficient range on state tests. To illustrate the power a test-preparation context can have on teachers, we use the example of Mrs. Gabriel, one of the most experienced teachers in the study and widely regarded as one of the best reading teachers.

Expressing pride in what the school was accomplishing in reading intervention, Mrs. Gabriel told us that to understand how their students were able to do so well on the state test, we needed to observe those supplementary classes as well as the regular reading classes. At her encouragement, we audio-taped one of her reading intervention classes as well as one of her reading classes. Unlike many intervention classes, where teachers were given a script to follow, Mrs. Gabriel had full authority to develop the supplementary reading program she used. Thus, we expected to record two high-quality lessons, believing that, in the hands of an exceptional teacher, even test preparation could be of high-quality. But that was not the case.

As analyzed by the reading expert on the research team, the reading intervention class failed to exemplify many of the characteristics of high-quality teaching found in Mrs. Gabriel's regular reading class. Structured around guided-reading groups on historical fiction, Mrs. Gabriel's regular reading class promoted principled learning of content, encouraged the development of students' metacognitive skills, created a successful learning environment, and prompted dialogue through teacher questioning and feedback. As summarized by our reading expert:

The teacher builds on children's experiences, explicitly expanding their knowledge, and encouraging them to be analytical and metacognitive....The lesson is coherent...[it] encourages them...to connect the story to their own lives, to draw connections to other stories...to consider genres and themes, and to develop nuanced meanings for vocabulary words...The teacher is enthusiastic

about the children's work...[they] complain when she has them return to their seats.

In contrast the reading expert found none of these characteristics of high-quality teaching in Mrs. Gabriel's reading intervention class, which employed whole group instruction to analyze short expository text on Jane Goodall:

The instruction is not coherent and is guided by the items on the worksheets.... [The teacher] attempts to focus their attention on how to write in the test-required format by reminding them of similar tasks they've done before. Students seem quite confused....both the teacher and the children seemed frustrated by and discouraged with some of the children's in-class answers.

As stakes increased for schools to make AYP, less time was spent on the more cognitively-demanding instruction exemplified by Mrs. Gabriel's first lesson and more time spent on the test-preparation type of teaching depicted in this Intervention class.

CHALLENGES FOR RESEARCH

This policy influence on conceptions and enactments of teaching quality poses two key challenges for researchers: one pragmatic, the other normative. The pragmatic challenge is the potential restriction in the range of teaching practices and access to teachers. With high-stakes assessments focused on isolated skills and short written responses, how often will teachers make high-cognitive demands of their students or connections to their real-world context? We saw little and, as indicated in the previous section, increasingly less time spent on what we considered to be important indicators of high-quality teaching. How interested will teachers and other stakeholders be in aspects of teaching that do not have direct and obvious associations with improved test scores? With helping schools achieve AYP?

The normative challenge for researchers is maintaining a critical, outsider perspective in relation to both practitioners and policymakers. One of the important

lessons we learned about the influence of policy on a study of teaching is how high-stakes assessments ratchet up the pressure to respond to practitioners' needs for useful feedback. Although interactions between researchers and practitioners frequently have some strain associated with differences in professional standards and motivation, a policy environment that requires dramatic improvements in student performance heightens these strains. Neither administrators nor teachers have much patience for generalized results or equivocal findings; they desire answers and assistance in meeting immediate demands for accountability. In this environment, a research dilemma is to present work with integrity: to resist pressure to report on work mid-stream, to protect confidences, and to guard against unintended interpretations or uses of data and research findings by policy makers (Valli et al., 2004).

We have increasingly come to appreciate the importance of finding ways of providing feedback without overstating or misleading practitioners about what we know and what we don't know regarding high-quality teaching. We thought to diminish the tension between practitioner and research needs by making teachers "co-researchers" in our research design, but for many teachers this was interpreted less as an opportunity to examine their own practice and more as a burden—especially because we could not give them immediate feedback from the data collection. More successful have been feedback sessions, in which we discuss broad patterns in the log data collected by teachers, and annual debriefing meetings, in which we discuss with principal and teachers data collection activities, answer questions about the study, discuss broad patterns in the data, and express our appreciation for support. Nonetheless, even these events place an

additional burden on time and fail to provide direct responses to questions about improving student achievement (Valli et al., 2004).

In addition to these challenges associated with relations to practitioners, researchers have a similar set of challenges in relating to funders and policymakers. In an era of high-stakes accountability, funding incentives are likely to promote the adoption of a restricted range of research agendas. In such a climate, it is important for researchers to resist pressure to relinquish characterizations of good teaching that do not directly and immediately translate into student achievement gains. With increased encouragement and funding incentives to develop partnerships with policymakers and practitioners comes pressure to adopt their agendas. How much will funders, especially government agencies, be willing to consider studies of teaching that do not posit student achievement as the main—or sole—criterion of high-quality teaching?

Yet without independent and alternative agendas, researchers would lose the space needed to make a full contribution to scholarship. Rather than contributing to theory-building and the development of alternative perspectives, their contributions would be reduced to evaluating goals, programs and practices set by others. Research on teaching would be even more pressed to study the impact of teaching on learning as measured by state-mandated tests. What would happen to the ex-centric research perspectives reviewed in numerous chapters in the most recent *Handbook of Research on Teaching* (Richardson, 2001a)? To research that looks at teaching as communicative action, caring, dialogue, moral activity, or liberation? As Bruner (1999) reminds us:

Education research should never have been conceived as principally dedicated to evaluating the efficacy or impact of “present practices.” The major research question that it faces is *not* simply how well our schools and systems of pedagogy teach spelling or mathematics or literacy. Rather, the master question from which

the mission of education research is derived is: *What should be taught to whom, and with what pedagogical object in mind?* (p. 408)

The current policy environment already encourages restricted, singular notions of scientifically-based research. Further constrictions on notions of teaching, and what aspects of teaching are worth studying, is an additional, troubling concern.

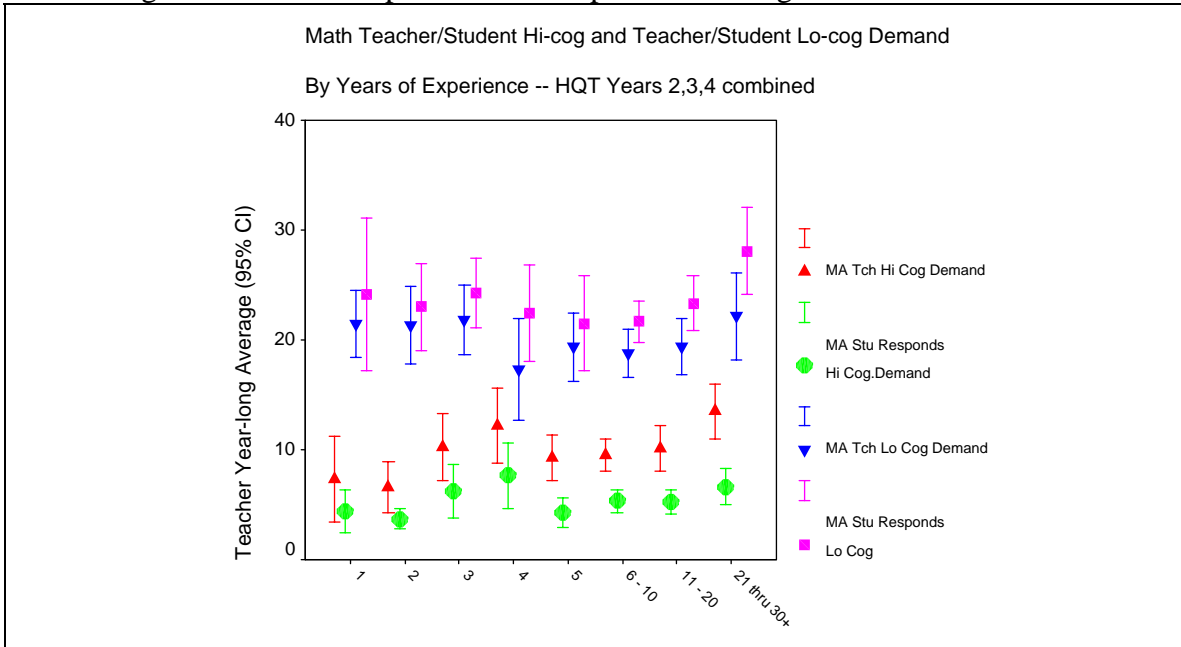
References

- Alexander, P. A., & Murphy, P. K. (1998). The research base for APA's learner-centered principles. In N. M. Lambert & B. L. McCombs (Eds.), *Issues in school reform: A sampler of psychological perspectives on learner-centered schools*. (pp. 25-60). Washington DC: The American Psychological Association.
- Ballou, D. (2002). Sizing up test scores. *Education Next*, 2 (2),10-15.
- Boser, U. (2001, January 11). Pressure without support. *Education Week*, XX (17), 68-71.
- Bradley, A. (1999, January 11). Zeroing in on teachers. *Education Week*, XVIII (17), 46-47, 49-52.
- Bruner, J. (1999). Postscript: Some reflections on education research. In E. C. Lagemann & L. S. Shulman (Eds.), *Issues in education research*. San Francisco: Jossey-Bass.
- Chambliss, M., & Graeber, A. (2003, April). Does subject matter *matter*? Paper presented at the meeting of the American Educational Research Association, Chicago.
- Cohen, D. K. (1988). Teaching practice: Plus que ca change... In P. Jackson (Ed.), *Contributions to educational change: Perspectives on research and practice* (pp. 27-84). Berkeley: McCutcheon.
- Cohen, D. K. (1995). What is the system in systemic reform? *Educational Researcher*, 24 (9), 11-17.
- Cohen, D. K., & Spillane, J. P. (1992). Policy and practice: The relations between governance and instruction. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 18, pp. 3-49). Washington DC: American Educational Research Association.
- Croninger, R., Valli, L., & Price, J. (2003, April). Mapping the policy environment for high-quality teaching: Can we get there from here? Paper presented to the annual meeting of the American Educational Research Association. Chicago, IL.
- Cuban, L. (1988). A fundamental puzzle of school reform. *Phi Delta Kappan*, 69 (87/88), 341-44.
- Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms 1880-1990, 2nd Ed.* New York: Teachers College Press.
- Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.

- Epstein, N. (2004). *Who's in charge here? The tangled web of school governance and policy*. Washington, DC: Brookings.
- Fenstermacher, G.D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, 107(1), 186-213.
- 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 edition (pp. 3-16). Washington, DC: AERA.
- 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.
- Goldhaber, D., & Anthony, E. (2003). *Teacher quality and student achievement*. New York: Teachers College Press.
- Goldrick, L. (2002). Improving teacher evaluation to improve teaching quality. National Governors' Association: NGA Center for Best Practices Issue Brief (Dec. 9). http://www.nga.org/center/divisions/1,1188,C_ISSUE_BRIEF^D_4732,00.html
- 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.
- Little, J. W. (1999). Organizing schools for teacher learning. In L. Darling-Hammond & G. Sykes, (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 233-262). San Francisco: Jossey-Bass
- Lortie, D. (1975). *Schoolteacher: A sociological study*. Chicago: The University of Chicago Press.
- McCaffrey, D., Lockwood, J., Koretz, D., & Hamilton, L. (2003) *Evaluating value-added models for teacher accountability*. Santa Monica, CA: Rand Corporation.
- McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York & London: Routledge.
- No Child Left Behind Act, (2001). Public Law 107-110.
- Oser, F., Dick, A., & Patry, J. (Eds.) (1992). *Effectiveness and responsible teaching*. San Francisco: Jossey-Bass.
- Raudenbush, S. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29 (1), 121-129.

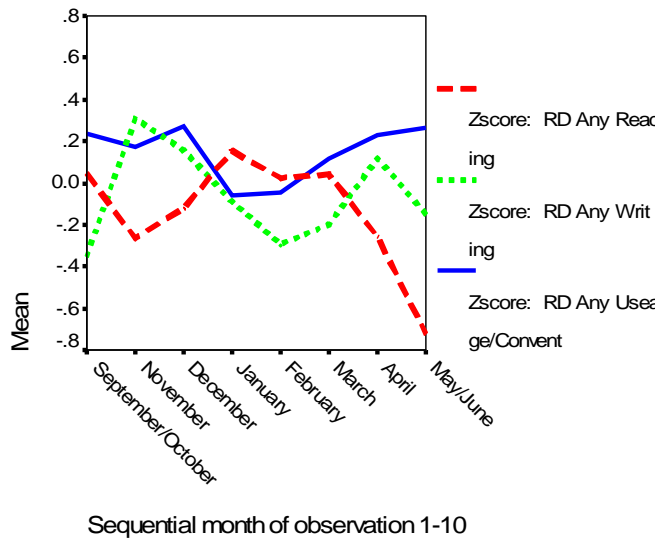
- Rice, J.K. (2003). *Teacher-quality. Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.
- Richardson, V. (Ed.) (2001a). *Handbook of research on teaching*, 4th Edition. Washington, DC: AERA.
- Richardson, V. (2001b). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching*, 4th edition (pp. 905-47). Washington, DC: AERA.
- 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.
- Schemo, D. J. (2004). When students' gains help teachers' bottom line." *The New York Times* (May 9).
- Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. Wittrock, *Handbook of research on teaching*, 3rd edition (pp. 3 -36). New York: Macmillan.
- Smylie, M. (1994). Redesigning teachers' work: Connections to the classroom. In L. Darling-Hammond (Ed.), *Review of research in education* (pp. 129-177). Washington, DC: American Educational Research Association.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of school reform*. Cambridge, MA: Harvard University Press.
- Valli, L., Croninger, R., Alexander, P., Chambliss, M., Graeber, A., & Price, J. (2004, April). A study of high-quality teaching: Mathematics and reading. Symposium paper presented at the annual meeting of the American Educational Research Association. San Diego, CA.
- Valli, L., Croninger, R., & Walters, K. (2004, April). Who is the teacher: Cautionary notes on teacher accountability systems. Paper presented at the annual meeting of the American Educational Research Association. San Diego, CA.
- Wainer, H. (2004). Introduction to special issue on value-added assessment. *Journal of Educational and Behavioral Statistics*, 29 (1), 1-3.

Figure 1: Relationship of Teacher Experience to Cognitive Demand*

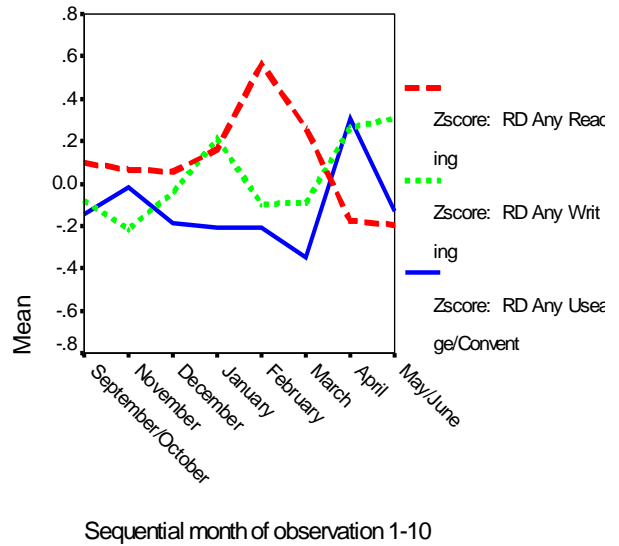


*Data combined from HQT Years 2, 3, and 4

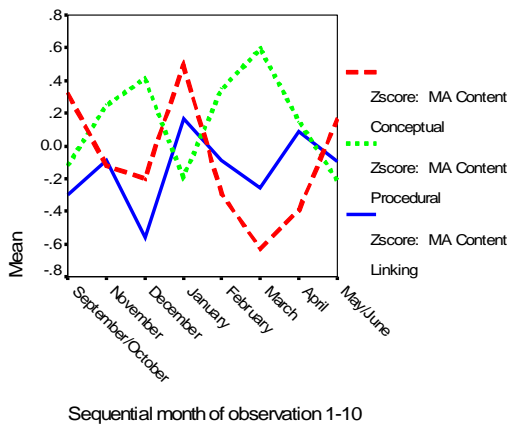
Figures 2: 4th grade Reading Content (03-04)



Figures 3: 5th grade Reading Content (03-04)



Figures 4: 4th grade Math Content (03-04)



Figures 5: 5th grade Math Content (03-04)

