“Safe to Say, it’s Not Like it Used to Be:”

Second-Stage Teacher Responses to External Accountability Policies and Reforms

Stacy Agee Szczesiul

Project on the Next Generation of Teachers

Harvard Graduate School of Education

This paper was prepared for the 2009 Annual Meeting of the American Educational Research Association, San Diego, CA, April 13-17.

This research was generously supported by a grant from the Spencer Foundation. The views expressed are solely those of the authors.

Introduction

Patrick Murphy was a teacher at Hill Middle School in March of 2006 when he and his colleagues learned how their school fared in a *Diagnostic Fact Finding Review* conducted by the State Department of Education. The "determination letter" they received from the Commissioner of Education read, "After consideration of the review panel's findings and the evidence on which these findings are based, it is my determination that the Hill Middle School is, at present, an under-performing school. In my judgment, the interests of Hill Middle School students will be best served by active state/local collaboration to ensure the success of efforts to improve student performance at the school." According to the report, the school failed because it was not implementing a sound plan for improvement and, thus, was not producing satisfactory gains in student performance.

After two years of state intervention, Patrick is still unsure who authored Hill's School Improvement Plan. He guesses the state dictated to the principal what to write. He's still skeptical about the content of the plan and its effect on his teaching. Patrick admits his efforts to satisfy the plan are, at best, perfunctory. He also admits there are areas of the plan that he and his colleagues quietly opt out of without fear of retribution from their principal. Patrick still does not understand why he should be subjected to state oversight since through training, research, and experience he became the expert state administrators told him to become. And, as a result, 91% of his students consistently pass the state test even though many of them have special learning needs and/or come from significantly disadvantaged households. This is particularly troubling to Patrick now. For the first time since he began teaching, he feels his control over instructional decisions has been usurped by state administrators who do not understand what teachers do and he is actually worried about his test scores. Patrick's worry seems well-founded. Since 2006 Hill Middle School students have performed progressively worse on state assessments in math and English Language Arts.

In this paper I explore how second-stage teachers, like Patrick, who work in low-performing urban districts, are responding to the current context of external accountability. Based on my analysis, most generally accept external accountability measures as part of the landscape of their work. With that said, teachers' responses to specific policies and reforms varied and appeared to be closely tied to their school-level experiences. Some teachers responded to external policy by integrating it seamlessly into their work. They described working in schools where teachers and administrators shared high expectations for teaching and learning and teachers were held to account for satisfying those expectations. The majority of teachers, however, responded as Patrick did—with questions and growing cynicism about external influence on their work. They did not adapt well to current policy and reform and they described working in schools that were disorganized and fragmented. Thus, external accountability policies had varying levels of influence on teachers' decisions about curriculum and instruction (i.e., the autonomy they experienced) as well as on the amount of pressure they felt to improve student performance (as measured by test scores). Patterns in variation by school suggest that individual schools' internal norms and systems—their internal accountability systems—served as powerful mediators of national, state, and district level accountability policies (Elmore, 2004; Spillane, 1998; Newmann, 1997), even when high-stakes were attached.

This work deepens the growing literature on teachers' experiences of external policy and school-level accountability by noting variation among low performing and high performing schools. Much of the research thus far has focused on variation between low and high performing schools. It also focuses on a specific cohort of teachers—those who came into the work force when the accountability context was either emerging or already fully in play. Like Patrick, they feel they have established a high level of expertise and they feel confident and competent in their ability to provide instruction that will produce student learning (Berg et al., 2005; Donaldson et al., 2007; Drake, 2007; Johnson et al. 2004). It seems reasonable to expect that they would also want to use their own discretion in matters related to curriculum and instruction (Huberman, 1993; Rosenholtz, 1985). This is important to consider given that external accountability

reforms tend to overlook teachers' needs for autonomy, power, and flexibility to do their job effectively (Ingersoll, 2007; Lee, 1991) and may undermine their motivation and commitment (Finnigan & Gross, 2007; Ingersoll, 2007).

Contextual Frame

The current climate of performance-based accountability can be traced back to the "horse trade" proposed by the nation's governors in the mid-1980s (Elmore, 2002; Tyack & Cuban, 1995). Made soon after the release of *A Nation at Risk* (which concluded that the inadequacies of public education in American schools were in large part due to pervasive problems in teaching), the proposal suggested that states grant more discretion about what and how to teach to districts and schools in exchange for more accountability for academic performance. Decades later, with the passage of *No Child Left Behind,* all states have created accountability systems that are characterized by curriculum standards, testing, reporting of test scores by school, and some means for identifying those not making adequate improvement (Elmore, 2004). Many see these policies as a way for school leaders and teachers to "drive reform for better learning and teaching" (Evans, 2009, p. 78; see also Fuller & Johnson, 2001; Korchoreck, 2001; Rorrer & Skrla, 2005).

Whether recent accountability reforms are actually producing better teaching is the subject of on-going debate (Diamond & Spillane, 2004; Kennedy, 2007; Stigler & Hiebert, 1999). Less debatable is the claim that teaching has *changed* with the demands and pressure of national, state, and district accountability reforms. Valli and Buese (2007) found that, under greater levels of hierarchical control, role expectations in teaching have increased, intensified, and expanded to such a degree that teachers are deprived "of the mental and temporal space needed to understand and use new curricula well" (p. 546). Furthermore, current reforms also limit the individual teacher's ability to close their doors on what has historically been perceived as unwanted intrusion (Booher-Jennings, 2005; Costigan, 2002), thereby threatening, in unprecedented ways, deeply entrenched occupational norms of privacy and autonomy (Elmore, 2000; Little, 1990; Lortie, 1975).

Also unprecedented in teachers' work may be the pressure to close race and class-based achievement gaps which leaves many feeling less competent, less responsible, and less efficacious (Evans, 2009). While teachers in low-performing schools may be initially responsive to external policies, persistent failure to satisfy policy expectations demoralizes them and deteriorates their motivation to expend more effort, try new approaches, or participate in professional development (Finnigan & Gross, 2007). This is consistent with the finding of many studies that suggest a school's academic status shapes teachers' conceptions of external accountability policy and their capacity to satisfy it (Abelman, Elmore, Even, Kenyon, & Marshall, 1999; Debray, Parson, & Woodworth, 2001; Diamond & Spillane, 2004; Evans, 2009; Haertal & Herman, 2005).

While external accountability policies and reforms may be producing changes in the work of teaching, they likely suffer from the same fatal flaw as those of previous eras insofar as they fail to consider the influence of local context. That is to say, the shared norms and expectations of those inside a given school shape, in powerful ways, how external policies are understood, valued and implemented. (Elmore, 2004; Newmann, 1998). Schools with strong internal accountability systems—those with tight alignment of individual responsibility, collective expectations, and accountability mechanisms—are more likely to respond effectively to external accountability policies, while those characterized by atomized accountability stand to react in unproductive or detrimental ways under external accountability pressure (Elmore, 2004).

In this paper I explore what role school-level influences play in second-stage teachers' experiences of external accountability policies. I argue that the threat of (further) sanctions in chronically low performing schools is of no consequence to teachers in some schools with weak internal accountability systems while it is detrimental to teachers in others. In the case of the latter, teachers in the study described mandated reforms that led to no more than tinkering around the edges of instructional improvement, meaning that teaching was reduced to sets of procedures rather than being about the meaningful engagement of students and teachers around content. Importantly, I also explore how teachers' accounts of

working in high-performing, urban schools with strong internal accountability systems illustrate the possibility of overcoming further educational stratification by race and class.

In the section that follows I lay out the study design and my analytic approaches. Then, I examine the distinct views these second-stage teachers held about the current climate of accountability. From there I will focus how they have responded to external policy and reform, giving particular attention to the variation across schools. I will specifically look at differences between and among teachers from four schools—two described as having strong internal systems and two as having weak systems. I conclude the paper by pointing out why these teachers' experiences should be instructive to policymakers.

The Study

Rational

In the late 1990s, as the external accountability movement was gaining strength, a new wave of teachers began replacing the ranks of retiring veterans. Today these new teachers are in the second stage of their career (years 4-10) with the acquired knowledge and skills needed to increase the capacity of schools to succeed with all students (Rockoff, 2004; Hanushek, 2004; Murnane, 1981; 121 Kane, 2006). Yet, teachers at this stage are entering the "danger zone" when it comes to decisions about staying in or leaving the career (Huberman, 1993). Despite the fact that teachers with 4-9 years of experience leave the career at nearly as high rates as those with less experience (Marvel et al., 2007), researchers and policymakers have given virtually no attention to them.

Recent research suggests that second-stage teachers have high levels of understanding of current reforms and may be more inclined to implement them than their more veteran counterparts (Drake, 2007). There is evidence that they may have expectations for autonomy that differ from those of earlier generations of teachers (Johnson & The Project on the Next Generation of Teachers, 2004; Szczesiul, 2007) and they may identify with and feel part of a larger, more coordinated effort to improve teaching and learning

(Kennedy, 2005). They appear to want standards-based curriculum that is aligned with state assessments (Costigan, et al., 2004; Grossman & Thompson, 2004; Kauffman et al., 2002; Szczesiul, 2007), time to collaborate with colleagues and coaches (Charner-Laird, 2007; Johnson, 1990; Kardos et al., 2001), and opportunities to learn new skills from professional development (Charner-Laird, 2007; Johnson & Birkeland, 2003; Szczesiul, 2007).

In my study I sought to understand the experiences of second-stage teachers who began their careers in strong external accountability contexts and have thus far chosen to stay in their urban, low-performing school districts where their competence, organizational savvy, and amenable disposition to reform is most needed (Berg et al., 2005; Donaldson et al., 2007; Drake, 2007; Johnson et al. 2004). It is important to consider whether second-stage teachers experience tension between the demands of instructional mandates and a professional need to use their own experience-based expertise to make decisions about what and how to teach. Their individual and collective expertise could be harnessed to strengthen school reform efforts at a time when vast numbers of veterans are retiring (see Drake, 2002; Hatch et al., 2005). My work was guided by the following research questions:

- Given a context of strong external accountability, how do second-stage teachers from three urban school districts describe the autonomy they experience?
  - What role do district-wide curriculum and instructional mandates play in the experiences they describe?
  - What role do school-level factors play in the experiences they describe?

<u>Design</u>

This paper is based on data from a large qualitative study of urban teachers in the second stage of the career conducted by the Project on the Next Generation of Teachers and funded by the Spencer Foundation. We interviewed 85 second-stage teachers from 14 public schools in 3 low-performing urban

districts in the Northeast (see Table 1). The larger study addressed four main areas of interest: teachers' work engagement and their experiences of autonomy, professional learning, and differentiated roles. Studying these topics together enabled us to explore important intersections among these topics.

Our sample was purposively selected to ensure variation in teacher experience, subject and/or grade level taught and gender[1]. It consisted of 3-12 second-stage teachers at each of the 14 schools represented in the study. There were 22 elementary teachers, 29 middle school teachers and 34 high school teachers. Most participants were female (72%) and white (85%), reflecting general trends in public school staffing today. First-career entrants, those entering after college, made up 51% of the sample. This meant that nearly half of the sample consisted of mid-career entrants, those entering teaching after at least 3 years in a different profession[2]. Among all in the sample, 12% of teachers held a specialized role[3].

[1] Please, see Appendix A for detailed description of methods used to select and invite participants.
[2] For the purpose of this study, a first-career entrant is defined as a teacher who had no significant professional experience prior to entering teaching. A mid-career entrant is defined as a teacher who had at least 3 years of full-time professional experience in another field prior to entering teaching.
[3] For the purpose of this study, a differentiated role is a role that a teacher might take on in for which he or she either receives extra compensation or release time. Differentiated roles can take many forms such as department head, math resource specialist, literacy coach, or curriculum developer.

Table 1. *School Demographics*

	School Pseudonym	Total Student Enrollment*	Free & Reduced Price Lunch (%)*	Students of Color (%)*	Made AYP 2006		Total number of Participants		
					Math	ELA	Elem	Mid	High
District 1	Gansford High School	1739	55%	65%	Y	N	.	.	8
District 1	Conifer School	867	62%	71%	N	N	2	4	.
District 1	Greentree School	1117	67%	70%	N	Y	3	3	.
District 2	Bingham High School	231	78%	67%	Y	Y	.	.	4
District 2	Williams High School	1121	75%	70%	N	N	.	.	10
District 2	Deer Park Middle School	584	84%	70%	N	N	.	6	.
District 2	Hill Middle School	581	65%	61%	N	N	.	5	.
District 2	Lane Elementary	345	81%	79%	N	N	5	.	.
District 2	Partridge Elementary	377	83%	72%	Y	N	3	.	.
District 3	Dryden High School	3421	60%	61%	N	N	.	.	12
District 3	Montgomery Middle School	558	82%	62%	N	N	.	6	.
District 3	Allendale Middle School	575	88%	64%	N	N	.	5	.
District 3	Alpine Elementary	482	70%	56%	N	N	3	.	.
District 3	Keller Elementary	470	66%	74%	N	N	6	.	.
		Totals					22	29	34

We focused on teachers in urban districts for several reasons. First, retaining talented teachers in high-poverty, urban schools is difficult; they "lose, on average, over one-fifth of their faculty each year" (Ingersoll, 2004, p. 2). Second, urban school districts are more likely to be funded in greater proportions by state and federal funds, making them more susceptible to external review and sanction. Indeed, teachers in urban districts are more likely than those in suburban districts to be working in schools identified as In Need of Improvement, Corrective Action, or Restructuring. It follows, then, that these districts that are compelled by external authority have devised large-scale, standards-based improvement plans aimed at directly affecting teachers' instructional practice. We

wanted to investigate what this would mean to second-stage teachers who feel they are experts in their work, are competent and confident in their abilities, and are ready to use their professional discretion in instructional matters (Huberman, 1993; Rosenholtz, 1989; Szczesiul, 2007). Finally, we were interested in urban districts because teachers are more likely to stay in their classrooms if they achieve a "sense of success" with their students (Johnson & Birkeland, 2003). This is important to consider because success in urban contexts is increasingly measured by test scores. Resnick and Glennan (2002) note that "in the national drive to raise school achievement, urban school districts pose the greatest challenges" and "gains in reducing achievement gaps have not been maintained" (p. 2).

Data Analysis

Each team member wrote a thematic summary following each interview. These summaries captured background information such as age, grades taught, route of teacher preparation taken, and number of years as a full-time teacher. They also captured relevant information about the teachers' reports of engagement and investment in her work, how she makes decisions about what and how to teach, what learning experiences she finds valuable, whether she has experienced differentiated roles, and how she thinks about her future in teaching. By completing and sharing amongst ourselves thematic summaries on each participant, we explored, early in the process, individual experiences while also noting patterns as they began to emerge across participants, within and across different contexts. This round of data analysis helped lay the foundation for the themes we pursued and the initial codes we brought to the transcripts.

As a data reduction strategy, the research team generated and applied a broad list of codes grounded in both the thematic summaries as well as literature relevant to the various topics of interest. For example, in order to understand influences on teachers' decisions about practice we

coded the data with terms such as "external mandate," "school context,", "teacher competence" and "teaching description." This strategy resulted in a manageable data set that focused on specific topics while still reflecting the experiences of all teachers in the study. At this point in the analysis, individual team members focused on their own research questions and made a deeper cut at the data by applying more targeted codes, creating matrices, and profiling schools.

By using matrices I was able to see distinct patterns across the entire sample and between chronically low performing schools and high achieving schools. For example, across all second-stage teachers I noticed a general acknowledgement that external accountability policies are fundamental to the landscape of teaching today. I also noticed striking differences in how the two groups talked about what constitutes the work of teaching in their schools, who has authority over instructional decisions, and whether/how accountability plays a role in their decisions. From this, it became apparent that teachers' responses to external policies were tied to schools' internal accountability systems. Those in schools with strong internal accountability systems made their instructional decisions in vastly different ways than their counterparts in schools with weaker internal systems. Thus, I sharpened my analysis even more by bringing to it Elmore's (2004) working theory of internal accountability. Guided by this framework, I looked at three elements of internal accountability: (1) How individuals view their own participation in the school which includes what they feel responsible for; (2) How the collective views the work and practices of individuals in the school (or what the shared expectations are); and (3) What routines, practices, processes organize the work (or how individuals are held to account for what they do). In schools with strong internal accountability systems, we would expect to see tight alignment of these three elements, high levels of coherence and transparency, explicit norms and processes, focused support, and high agency. In schools with weak systems, we would expect to see individual responsibility trumping collective

expectations, weak norms around practice, low transparency and agency, and further atomization under external pressure (Elmore, 2004).

Second-stage teachers and high-stakes, standards-based accountability

Safe to say it's not like it used to be. But, I don't know what it used to be. (Maryanne, Montgomery Middle School)

…there is a lot of accountability…But, I haven't been here for 20 years so I can't compare it to what it used to be. (Anya, Lane Elementary School)

As noted earlier, the second-stage teachers in our study entered the career when strong external accountability policy and reform was emerging or already fully in play in their state and districts. Many, like Anya, had no point of comparison between what teaching is today (as they experience it) and what it might have been in the past. In discussing the influence of external policy and reform on their teaching, they revealed distinct views about state standards and district curriculum, performance-based testing, and teacher expertise and administrative oversight.

State Standards and District Curriculum

All of the teachers in the study acknowledged a need for and valued the presence of state standards to guide or serve as a foundation for determining what they teach. Although unmatched in his exuberance over standards, Oliver—a self proclaimed "standards guy"—described how he, like others in the study, used state standards to teach kids what they agree is important in their respective areas:

> I am a standards person. So, that part of the job and the DOE stuff has been fine for me. In fact, I've accommodated to it because I really like it. I think it's focused. I believe the standards are decent. You could have other standards, but these are perfectly good ones. I have come to believe if my kids can do well at the standards…then they will be better junior science citizens. (Oliver, Deer Park Middle School)

Many of the teachers in the study described how they appreciated the way state standards "focused" their attention to what is important content. They also saw state standards as serving an even greater

purpose in so far as they are a way to make sure all students, whether urban or suburban, are being exposed to high level content. Laura, a sixth grade teacher, put it like this:

> I think it's a good baseline just to keep everyone on track, it brings the equality to our educational system… I guess it just kind of levels the playing field and makes sure we're all covering the same materials. (Laura, Lane Elementary School)

In a similar vein, teachers viewed the use of standards, district curriculum, and curriculum maps as serving a practical purpose as well. Danielle, a middle school English Language Arts teacher, did not like the idea of getting "a kid from another school [who] has already done all of the things that we haven't gotten to yet because they've chosen a different sequence." Almost all of the teachers in our study, particularly those working in elementary and middle schools, echoed the sentiment that teachers need to be more coordinated across schools in the content they teach because of "highly mobile" urban students. It made sense to them that, if a student left one middle school in their district to go to another in their district, he should not miss or have to repeat any content. While none of the teachers wanted overly prescriptive approaches to make such coordination occur, they did view the specified curriculum and curriculum maps provided by the district as a way to make it possible. Laura draws an interesting distinction between herself as a "newer teacher" and her more veteran colleagues in this regard:

> As a newer teacher, I'm more comfortable with change when it's necessary but [for] a lot of teachers who have been in the system, any change is really hard and you see that. You see the resistance to it. I guess as a younger teacher you're kind of used to kind of being thrown in a different grade or a different school…I'm a little more flexible or open to these things…but not everyone feels the same way about these curricula and they miss that creativity and that flexibility. (Laura, Lane Elementary)

The call for using standards is not new in teaching; however, "first wave" reforms of the 1980s produced only "routine compliance" among teachers and certainly did not create academic excellence in U.S. schools (Tyack & Cuban, 1995, p. 79-80). The second-stage teachers in our study described state standards that they valued and that figured prominently into their instruction. These findings are consistent with those of a 2003 Public Agenda report in which 80% teachers said

"having guidelines for what students should learn helps improve academic performance" and 53% said they wanted "local standards initiatives to proceed as planned" (p. 12).

Performance-based testing

Teachers in the study felt students' progress toward a set of performance outcomes should be expected and monitored in a broad, systematic way, but most had problems with the current test-based approach. They couched their opinions about the current system in what they perceived as "fair" practice. Because Oliver felt the state test was "ultimately a fair exam," he made the decision to not only become the "clean-up hitter" who prepares students in the 8th grade for the science test, he also made it the formal criteria by which he will "measure [his] success as a teacher." Other teachers, such as Shirley, were of the opinion that while "making teachers generally accountable" for student outcomes was "a good thing," test scores were not a fair measure because "it's just impossible to make the gains" needed with "special education and English Language Learner students." That performance-based accountability was a reality in these teachers' work lives was not a point of contention. The means by which students were being assessed was the issue, and teachers in the study devoted time and thought to better, more valid alternatives to testing:

> It would be probably like a reform that looks at the individual student…the MAP testing because it really identifies where the child is today and if we set a goal with them for the year and they reflect on that goal and they set up how they're going to achieve that goal in each subject area and you see those gains and they get excited. It may be like a five point gain or a ten point gain but you can say to them, like you show them the chart. This is how far you've gone in one year and it really just focuses on the individual. (Laura)

Laura, like many in the study, is frustrated by "politicians" who do not see the "reality of the situation" and the challenge of bringing all kids up to a "93 or 97% proficiency" level despite their wide range of cognitive abilities. Similar to other second-stage teachers, her inclination was to bring policymakers into her classroom so that they could experience, first hand, the challenges of satisfying the expectations of this particular performance-based measure.

It was clear that many of these teachers disagreed with a test-based performance policy, yet the way in which they talked about standardized testing arguably marks a departure from that of previous generations. In his seminal work, Lortie (1975) found that teachers talked about such tests in ways that suggest a disconnection between what they do and what students produce on tests. He noted that "it was as if they are uncertain of the tangibility of measured gains or the rightfulness of their claiming credit for them" (p. 128). Regardless of their opinions of the state test, the second-stage teachers in our study saw clear connections between it and what they do in their classrooms.

Teacher expertise and administrative oversight

Teachers acknowledged the need for student performance measures and most believed that administrators should be generally monitoring their day-to-day practice. This may seem counterintuitive given the fact that the vast majority of these teachers felt they had achieved a high level of expertise, were confident in their practice, and were skeptical about the competence of administrators in matters of teaching and learning. Interestingly, it was *because* they felt they had reached this level that they wanted state, district, and school level administrators to monitor their performance. Patrick laid out this seeming contradiction by first pointing to his own demonstrated competence and then describing a site visit by state administrators in which he felt he was being arbitrarily evaluated. Ultimately, he wants state administrators to be more consistent about their own demands while also being more present and less evaluative so that they can understand what it is they see, rather than what they think they need to see:

> I'm the expert. You told me to become the expert. You must have some level of trust for me to do this job correctly. And…I have these two inclusion groups and I still have a 91 percent pass rate…[State administrators] literally flooded in my room with their clipboards. Eight adults in this room with 24 children and me. Clipboards. I said it felt - do you ever read *Harry Potter*? It felt like the dementors just entered the room. All joy was sucked out...they want to talk about research, and they can't show me anywhere where it says if I put that number on that, and I use this, my students are going to do better. I'm harping on it because… I used to always include that, and they told me to stop, and now they're telling me to do it again. So I want the state to understand better what we really do. (Patrick, Hill Middle)

There was a 'seeing-is-believing' mentality among teachers: If administrators do not believe or "trust" that teachers are doing their jobs, they should come and see for themselves. Many of the teachers in the study offered another reason for increased oversight: they had themselves observed poor teacher quality in their own colleagues. Teachers like Colin, a high school teacher, echoed Patrick's sentiments, but took it one step further. They want administrators to be less concerned with teachers who demonstrate competence and more concerned with teachers who really need support:

> …And they come in and they take a snapshot of your room; and they feel the need to tell you to do certain things. And they don't understand that I was doing them before they came up with the code name for them…when you make decisions for the school and you just kind of treat everyone the same, I think that, I don't really do that for my students either. I assess their needs. You need to find the teachers that need help and give it to them; and you need to…kind of leave the experts alone and let them be creative and do their thing. (Colin GHS)

Others, like Tia, a 7th grade English Language Arts teacher, wanted administrators to hold all teachers accountable for satisfying a simple professional code. She spoke plainly about the need for oversight and accountability when asked if she thought teachers would benefit from more time to work collaboratively on instruction:

> I think on some level it would be again, good in theory. Unfortunately, I think too many people would blow it off and just sit in their rooms and watch CNN or do personal stuff. All the time. Does it piss me off to no end? Yes it does… They don't show up to professional days and when they leave, they leave early. They might get scolded, but ultimately nothing happens to them. That is the most annoying, that somebody can come to work a third of the time and do 1/8 of the work I do and get paid more because they have been here longer. (Tia, Conifer ES)

It is interesting that these second-stage teachers seemed to depart, in some ways, from earlier generations of teachers with regard to the norms of egalitarianism and autonomy (Lortie, 1974). They seemed ready to acknowledge differentiation among the ranks with regard to teachers' expertise and skill and they did not reject out of hand administrators who exercise supervision. These teachers suggest that the ideal relationship is between teachers who are confident enough in

their own practice to make it public and administrators who are both astute enough to distinguish accomplished teachers from those in need of support and firm enough to hold teachers accountable for professional behavior.

While second-stage teachers across all three districts described working in similar accountability contexts and they appeared to hold common views about standards and tests, there were striking differences in how they experienced external polices and reforms at the school level. They described "particularities" about their schools that influenced how they interpreted and carried out external policy (Elmore, 2004, p. 135; see also Spillane, 1998).

"I" v. "We" in teacher response to external accountability policy

We teach kids to read and write (Paul, Bingham High School)

I understood it was a test we would all take in the end and I hoped that they would do well. (Dianna, Deer Park Middle School)

Across districts teachers in the study described working in similarly strong external accountability environments. Yet, external policy and reform had varying levels of influence on teachers' decisions about curriculum and instruction (i.e., the autonomy they experienced) as well as on the amount of pressure they felt to improve student performance (as measured by test scores). This is consistent with Spillane's (1998) findings that there is wide within district variability in response to state and national policies. Patterns emerged at the school level in how teachers talked about: what constitutes the work of teaching in their schools, who has authority over instructional decisions, and how accountability factors into their work. These patterns suggest that teacher response to external accountability policies are related to their school's internal accountability systems. In the section that follows, I will profile the experiences of teachers from four schools—two with strong and two with weak accountability systems—to illustrate the similarities and differences within and between both categories. Their accountability status is shown in Table 2.

Table 2 Accountability Status for Schools in 2006 and 2007

School	Internal Accountability	Accountability Status 2006 ELA	Accountability Status 2006 Math	Accountability Status 2007 ELA	Accountability Status 2007 Math
Bingham HS	Strong	None	None	None	None
Lane ES	Strong	None	Improvement Yr 1	Improvement Yr 1	Improvement Yr 2
Deer Park MS*	Weak	Restructuring Yr. 1	Restructuring Yr 1	Restructuring Yr 2	Restructuring Yr. 2
Montgomery MS	Weak	Corrective Action	Corrective Action	Restructuring Yr. 1	Restructuring Yr. 1

*Deer Park Middle School came under state control in 2005

Second-stage teachers in schools with strong internal accountability systems

The only teachers in our study who described working in schools characterized by strong internal accountability systems were from Bingham High School and Lane Elementary School. That is to say, teachers from these two schools were in the minority when it came to talking about working in an environment where there was a shared commitment to a clear purpose for student learning and means by which teachers were held to account for demonstrating what teachers and administrators agreed was high quality of teaching (Elmore, 2004; Newmann, 1998).

Bingham High School

At Bingham HS the work of teaching was summed up by Paul's simple comment, "We teach kids to read and write. We very rarely even bring the state test up. But if we are always reading books and talking about them and writing about them, we are prepared for it." This sentiment was echoed by all of the Bingham teachers we interviewed. State standards and high-stakes tests were part of the larger gestalt. Testing didn't produce anxiety or a sense of urgency and teachers described high expectations for student learning that were observable in ways other than performance the state test.

Importantly, teachers described being confident in their ability to produce the level of student learning that was expected in the building. They also made it clear that power over decisions about curriculum and instruction was claimed by teachers who were recognized as experts. Fred, who doubles as a teacher and building administrator, made this clear:

> [O]ne of the things that I always liked about being a teacher here, and still do, is that the administration in this building allows you to be a professional enough to say, we know you know your content, we know you're a good practitioner, go forth and educate. I'm not going to dictatorially determine that you have to be on this page of this curriculum at this time of year kind of thing. It never works that way here. We've never done it that way, and I think that's part of why people enjoy teaching here. (Fred, Bingham HS)

What the teachers at Bingham described was a simple (and familiar) agreement: administrators (who are perceived as instructional leaders) trust teachers to use their own professional discretion regarding decisions about what and how to teach and teachers produce high levels of student achievement in return. Teachers' descriptions of their instruction made it clear that the work at Bingham was about meaningfully engaging students in the content and teachers constantly learning about their own practice. Lydia, who incorporated case studies in her classes was constantly asking herself if the work was "authentic," if her students "were bringing [what they learned] back with them" and "making connections" to other content. Fred also illustrated the high expectations members of the school have for teaching and learning when he talked about "rounds." Partnering with an area university, Bingham High has always practiced something of a medical model to refine and improve instructional methods. Unlike many of the teachers in other schools we spoke with, those at Bingham, like Fred, have a nuanced understanding of effective practice. Early on in his experience at the school, he began to see the difference between effective practices and "activities:"

> [We] had the round, and the kids lost sight of the forest through the trees. They had so much fun jumping into the boxes and digging, the connections to what it was I was doing, the reason why I was doing it, the historical stuff got lost. And so in the post round, you know, in a tactful way, I think it was my professor who said, you know, what do you think they got out of this? And what do you think they know now about archeology? Just walking

> around the room, it was apparent, not much. And so that became a really informative position for me. I still use the archeological dig as kind of a, I always ask myself the question, at the conclusion of this activity, is it more the activity, or is it more what is meant to come out of the activity. (Fred, Bingham HS)

Although individuals talked about decisions they had the authority to make on their own, they also described explicit norms and expectations about what constitutes effective instruction. These were powerful influences over teachers' decisions and Paul made it clear that there were consequences for not meeting shared expectations:

> …someone not pulling their weight here would just not fit in…students tell you straight up if class isn't very good or if they think that you are having them do busy work. They know what that means and they know that they don't do that here. So that would just be out of place. Not to mention that you would hear about it probably from the administrators and we are in classes with each other all day, too, seeing each other teach and it would just be embarrassing. (Paul, Bingham HS)

According to teachers, the consensus in the building was to engage students at high levels in reading and writing and the rest would take care of itself. Importantly, teachers said there were shared expectations about what high level instruction looked like, that they felt responsible for meeting them, and that there were mechanisms to hold teachers to account for their work. It appears this is what influenced Bingham teachers' work the most. They did not describe having to respond to external policy and reform in a way that was different from how they would respond to any other day-to-day classroom demand. Thus, external accountability policy was of little consequence to these teachers and a reputation of high test scores kept them buffered from state and district level "minutiae."

Lane Elementary School

Like at Bingham, the sentiment among teachers at Lane Elementary School was "…If I do what I'm supposed to be doing, I'm teaching the curriculum, if I'm looking at the standards, then they should be fine. Then I don't need to spend an hour a day on test prep books." Lane teachers

described high expectations for student learning and they were confident in their ability to produce it. They characterized the work of teaching as being driven by multiple forms of data. These data were used to inform curriculum and instruction decisions, particularly in how teachers targeted their expertise to make the curriculum accessible to students with a wide range of needs. Testing, standards, and district curricula were seen as an integral part of the work, not as something in addition to what they were already doing. According to teachers, state testing was not the ideal way to measure student performance, but it did not produce undue anxiety in them. In fact, Lane teachers reported that assessment data (including the state test) kept them sharp, focused, and competitive. Rebecca, below, described how the use of regular assessment data provides her with diagnostic information about where kids need support:

> I know how well my kids did or how poorly they did from last year's [state test]. But the breakdown -- it really means nothing to me to be completely honest with you. But we do the MAPS testing. And that gives you a better breakdown of where your kids -- is it -- you know they don't understand vowel sounds -- or their comprehension skills are weak -- or maybe they're sentence attack skills are great but their vocabulary is poor. That gives a better breakdown. (Rebecca, Lane ES)

At Lane, teachers like Rebecca described how they make data-driven decisions that allow them to tailor their instruction more precisely to individual students' needs.

Although teachers at Lane described making targeted changes to curriculum and instruction, the principal at Lane had the authority to make broad decisions that limited what teachers had control over. Each teacher we spoke with described being entirely on board with the principal's plan of center-based learning which is informed by district curricula and data-based decision making. It was clear that there were coherent, explicit, and shared norms and expectations around practice and individuals felt responsible for satisfying them. Anya represented the sentiment of all the Lane teachers we spoke with when she described the requirement of center-based learning:

> It's just about trying to move everybody forward and feel confident. So do I have to? Yes, I do what she [the principal] tells me but she wouldn't have to. It's just the way we do it here. (Anya, Lane ES)

It seemed that at Lane issues of individual autonomy were relatively moot because the teachers consent to the plan the principal has laid out. It defined how they were to carry out their teaching and they said they could not see themselves doing it any other way. According to teachers, they were committed to the principal's plan and there were understood consequences for not meeting shared expectations. Rebecca described the presence of the principal in the building and in her own classroom:

> [She] is very good at just doing quick walkthroughs through the building. When you're teaching, she'll come over, and she might kneel next to a kid and be like -- what are you working on now? What's the objective here, so? It makes me feel good when the kids are telling her what the objective is -- and they really know what they're doing…This is her building, and she wants what's best for the kids, and I do too, so. And if she's going to keep me on my toes, I think that's great. (Rebecca, Lane ES)

Although teaching was more directed at Lane than at Bingham, teachers made it clear that the principal recognized them for the expertise and skill they demonstrate in differentiating curriculum to meet the diverse needs of their students. Most of the teachers we spoke with also described being tapped by the principal to take on a new class or grade level or to lead a workshop for their colleagues. This was validating to them, even though in some cases the challenge they anticipated experiencing initially reduced them to tears. There was a general sense among teachers that, as Anya put it, "you have to be able to move around where it's best for the school."

As illustrated by both Bingham and Lane, even schools with strong internal accountability systems had vastly divergent ways of responding to test-based accountability. At Bingham consensus around mission, expectations, and purpose created internal accountability mechanisms for how the work of teaching was carried out. At Lane how the work of teaching was carried out created internal accountability mechanisms which led to consensus and commitment among teachers. At both schools, teachers described having a high level of agency and purpose, even though at Lane students

had not performed well enough on the state assessment to keep the school out of the Improvement category.

Second-stage teachers in schools without strong internal accountability systems

It was not uncommon to find teachers in our study working in schools that were not organized around a coherent set of expectations and practices for teaching and learning. It was also not uncommon to find that these schools were chronically low-performing. In the section that follows I focus on teachers from two such schools to illustrate the different ways in which they adapt to external policies.

Deer Park Middle School

At Deer Park Middle School state assessments and DOE pressure defined the work of teaching. Teachers talked about the weight and urgency they felt to produce high scores. Unlike Bingham and Lane, Deer Park is a chronically underperforming school and has been under state intervention since 2005. Eileen, a reading teacher at the school, was like others who "did not understand" the gravity of student performance on the state test. "I just didn't understand how big it was. I understood it was a test we would all take in the end and I hoped that they would do well." She described how much of her time is spent pulling questions from the DOE website to read so that she would "know where to place my emphasis - I think, I hope, I pray - I know where to place the emphasis."

At Deer Park, teachers' expectations for student learning were inconsistent and they often referred to the troubling reputation of DP students. They also frequently questioned their capacity to produce student performance outcomes. For example, Danielle commented:

> "It's too bad that teachers struggle so completely because of the state test. And we struggle every day mightily to ensure that our students can pass that test; and it's very, very difficult.

> How do you overcome children that have absolutely no vocabulary by the time they've reached seventh grade? I mean you've got to be kidding me. (Danielle, Deer Park MS)

Unlike the teachers at Bingham and Lane, some teachers at Deer Park located problems associated with student learning squarely in the domain of the student. In the rare case that Deer Park teachers described having an influence on student learning, it was most likely couched in discussion about using old test questions or test data to effectively prepare for the next assessment. Oliver's account of using data is strikingly different in nature than that of Rebecca's at Lane:

> I analyzed the raw scores on what it means to get this open response…you would need a 50 percent of all multiple choice. I've done the whole thing -- what it takes to get advanced because I need to have targets. Advanced is hard, by the way. But I've identified how many people can pull off this 13 out of 20…Yes, I know they can do it…to get proficient. (Oliver, Deer Park MS)

Oliver used the state test data to figure out where he could most likely make some gains in the next round of testing. Unlike the practice of using assessment data as a diagnostic tool, this practice does little, if anything, to address student learning or to improve his odds of actually effectively teaching the content of the questions he has determined his students can "pull off."

Interestingly, these sorts of practices seemed to be supported by the principal because, according to teachers at Deer Park, it was her role to make sure they were preparing students for the test. Thus, she enforced that teachers follow the district and state policies, but her enforcement varied, depending on subject area. Teachers described how those teaching outside of ELA and Math were given the authority to make decisions on their own or in their departments, but only if their decisions did not conflict with the larger goal of meeting DOE and district demands. This meant that teachers were left to make decisions on their own or in pockets creating a school that is fragmented and incoherent. There appeared to be no consensus regarding what effective teaching looked like or how teachers would be held to account for carrying it out. Based on teachers' accounts, if there was any shared purpose among teachers it was to prepare students for the test so

that the school would longer have to be under state "policing." As noted by Eileen, teachers and administrators wanted to escape the pressure because they had already been "slapped for having too much standing and lecturing on some of our DOE visits." It might, therefore, not come as a surprise to find there were serious internal consequences for not at least giving the appearance of going along with this purpose. As Pam described, after conflicting with the principal about retaining students who had not reached academic expectations, she was essentially told, "shut up and go back to your classroom." Pam realized that her principal's unilateral decision to promote students was based on a need to produce satisfactory numbers to state and district authorities—to give the appearance of progress. Indeed, fear of further sanction appeared to drive the actions of both teachers and administrators at Deer Park. Teachers such as Pam whose values were not aligned with the 'compliance-at-all-costs' culture of the school, did not feel safe or secure enough to challenge it. Perhaps not surprisingly, at Deer Park, teachers described being recognized for compliance with mandates, not as experts in matters of teaching and learning.

Montgomery Elementary School

Like Deer Park, teachers at Montgomery Middle School described the work of teaching as being predicated on the question: Are you a tested subject area? But, there appeared to be decidedly less pressure from administration to produce improved student test scores. In fact, whether the teachers we spoke with felt pressure or felt responsible for producing better scores seemed to be entirely about the individual. If they felt pressure, the reasons for it may not be the ones intended by policymakers or administrators. For example, Natalie noted that "because our scores are so low and we could be in jeopardy of the schools close," scores have to be brought up so that teachers can "keep all their jobs." Even though the school was under DOE review in 2000 and is currently in restructuring, teachers described how they are left to make sense of state and district mandates

virtually on their own with little oversight. Many of their comments suggested they don't really understand the way that external accountability plays out in schools. Again Natalie spoke plainly:

> I don't really know how the process works, but our principal has said that AYP numbers and all that, the academics for that, if the test scores don't go up the DOE comes in and then they will say, I don't even know what it is called but they come in and go into every single classroom and they watch to make sure teaching is done right and all that. (Natalie, Montgomery MS)

Couple not understanding the logistics of the process with an uncertainty about how to make sense of testing in their daily work lives, and teachers at the Montgomery were left with the authority to, as Ashley put it, "try to forget it's even there…because my kids never do good anyway…"

Discussion

In my study I found that the uniform expectations of external accountability policy can be interpreted in a variety of ways by teachers and have starkly different implications for their work. How teachers responded to the intentions of state and district level policies and reforms had less to do with the amount of external oversight, technical support, materials, and funding provided, and more to do with the shared norms and expectations that existed among those within the school. These internal norms and values proved, in powerful ways, to shape how teachers understood, valued, and implemented external policy (Elmore, 2004; Evans, 2009; Newmann, 1998). In schools such as Bingham and Lane where there was reportedly tight alignment of individual responsibility, collective expectations, and accountability mechanisms, external policies were secondary to what was standard operating procedure at the school. At both schools there was the general sentiment that "we teach kids how to read and write" and the rest follows. Still, there were also notable differences between teachers' experiences of external accountability policy at the two schools. External policies were integrated seamlessly into the agreed upon practices at Lane. This was important because it was

the practices that drove shared expectations and accountability for student learning at the school. The work being done at Lane was already highly aligned with the expectations of external policy. At Bingham, external policies were not a defining part of the discourse around teaching and learning. Their founding mission and commitment to meeting agreed upon expectations created the means for effective internal accountability and, thus, influenced teacher practice which included preparing students to perform well on all types of tests.

It is important to highlight the work of teachers and the effect of external policy in these two schools. The research literature is replete with examples of how urban schools serving racially and ethnically diverse student populations from low socioeconomic backgrounds fail to meet external accountability expectations. According to the teachers we spoke with, students, teachers, and administrators in these schools thrive, external expectations are almost always realized and, students and teachers feel a sense of purpose and accomplishment. They described the important ways in which leadership made this possible. For example, building administrators provided opportunities for teacher learning through practices like rounds or they organized schedules so that teachers could collaborate. While supporting teacher learning, they also validated teacher expertise and created buffers between the school and the district. In both schools teachers described how administrators built consensus around mission and practice and nurtured teacher commitment to it. It is worth noting that building principals at Lane and Bingham were strategic about their recruitment and hiring of teachers. Such selectivity improved their odds of sustaining a high level of commitment to shared expectations around teaching and learning.

The experiences of teachers from schools with weak internal accountability systems are also instructive. They clearly teach us that sanctions are not enough incentive for improvement. For teachers in schools like Deer Park where state officials were actively involved in the day-to-day

running of the school, the goal was to avoid further sanction. That translated into producing better test scores, which meant teachers felt they had to teach to the test. While in high performing schools external incentives structured around rewards such as recognition (rather than sanctions) may prove to be a motivator for teachers and administrators, the incentive structure for low performing schools may be less influential, if not detrimental (Diamond & Spillane, 2004). This certainly appeared to be the case among teachers at Deer Park where (according to their accounts) sanctions led to a fragmented, frenetic compliance that did not logically connect teaching and learning. Teachers did not implement initiatives such that they pertained to the quality of students' thinking and learning (Stigler & Hiebert, 1999). In low-performing schools under state pressure, the test became the curriculum and the center of teacher "learning." And, pressure in ELA and Math led to further segmentalism (Kanter, 1983) making separate areas of the organization responsible for different parts or pieces of the problem. Not surprisingly, teachers described how this created very different responses to the external policy and inhibited shared understanding or commitment to the solution within schools (Spillane, 1998) and led to some to believe that serious accountability sanctions were "something other people suffer from."

At Montgomery Middle, the incentive structure failed for other reasons. Teachers were virtually nonplussed by their accountability status, if they knew it at all. Therefore, consequences were not taken seriously or were unknown. This feeling that external accountability threats are simply empty was expressed by many teachers who described their schools as being fragmented and atomized—that is to say, they described their schools as having weak internal accountability structures. The experiences described by teachers in these types of schools lead to a fundamental question: Does external accountability (as it is currently conceived) really even matter to teachers in these schools?

Conclusion

The second-stage teachers in this study had distinct views about their work in the current context of strong, external accountability. Most differentiated themselves from their veteran counterparts and teachers of earlier generations in areas of standardized curriculum, performance-based accountability, and administrative oversight. Despite their common views, their responses to external policy and reform varied and appeared to be related to their school level experiences. Teachers in higher performing schools with strong internal accountability systems responded seamlessly to external policy while those in chronically low-performing schools did not. Those at Bingham and Lane described shared expectations about teaching and learning as well as practices that made them accountable to each other. Moreover, according to their accounts, consensus around practice did not mean that they did not challenge each other to improve (Achinstein, 2002). Such school-level "particularities" appeared to play a greater influence on their work than any external demands or pressure. According to these teachers, because they consistently met their adequate yearly progress goals or met other, internal student performance indicators their administrators buffered them from district "nonsense," which allowed them to spend time deepening their content knowledge or refining their skill at differentiating curriculum. This was on-going and tied to shared expectations around practice.

Teachers in schools that suffered from chronic failure and were under state control responded with cynicism about district and state level administrators who were making decisions about practice and with an urgency to improve test scores. Many described a need to comply with external reforms, if for no other reason than to give the appearance of improvement activity. This is not a new finding (see Diamond & Spillane, 2004), but disturbing nonetheless. These teachers described being constantly bombarded by new initiatives to implement in their classrooms with little time to think about their effects. Teachers frequently referred to new mandated procedures (such as posting objectives or turning in lesson plans) that they felt would have no direct effect on student

learning. Interestingly, teachers in schools that were low-performing, but were not under direct state control described the same sort of experiences and expressed similar cynicism. They, however, were not as likely to comply with the mandates because they were not experiencing the intense level of oversight nor did they really believe the threats of state intervention would ever become their reality.

These second-stage teachers' experiences have important implications for policymakers. First, in describing their work many suggested that there is an opportunity cost associated with constantly implementing discrete, disconnected instructional mandates that reduce teaching to sets of procedures and patchwork strategies. Teachers' obligation to the policy churn takes the place of what could be their contextualized, gradual, continuous, long term efforts to improve instructional methods (Kennedy, 2005; Stigler & Hiebert, 1999). This is important to consider with regard to teachers in the second-stage who feel competent and ready to meaningfully deepen their practice. Second, policymakers should attend to these teachers' responses to external accountability reforms because their responses inform us about *where* and *how* reforms matter. Among the teachers in our study, those in highly functional, organized schools were virtually unaffected by external policy. These schools had a history of making AYP goals and/or meeting internal student performance benchmarks. Those most affected were in the chronic low performing schools under state control. Decisions about these schools which influence teachers' work arguably did not create optimal conditions for improved teaching and learning. In fact, many teachers suggested that mandates further exacerbated problems with their schools. Policymakers might consider directing their efforts and resources toward the development of internal accountability systems if they want to see improvements in teaching and learning at low performing schools.

Works Cited

Abelman, C. & Elmore, R. F. (1999). *When accountability knocks, will anyone answer?* Research support series RR-42 Consortium for Policy Research in Education Philadelphia University of Pennsylvania Graduate School of Education

Achinstein, B. (2002). Conflict amid community: The micro-politics of teacher collaboration. *Teachers College Record, 104* (3), 421-455.

Booher-Jennings, J. (2005). "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal, Vol. 42*(2), 231-268.

Charner-Laird, M. (2007). "Ready and willing: Second-stage teachers and professional collaboration," Paper presented at American Educational Research Association Meeting, Chicago.

Costigan, A., & Crocco, M. (2004). *Learning to teach in an age of accountability.* New Jersey: Lawrence Erlbaum Associates.

Debray, E., Parson, G., & Woodworth, K. (2001). Patterns of response in four high schools under state accountability policies in Vermont and New York. In S. Fuhrman (Ed.), *From the capital to the classroom: Standards-based reform inthe states. The one hundredth yearbook of the National Society for the Study of Education* (pp. 170–192). Chicago: National Society for the Study of Education.

Diamond, J. & Spillane, J. (2004). "High Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality?" A working paper for the *Institute for Policy Research,* Northwestern University.

Donaldson, M. L., Kirkpatrick, C., Marinell, W., Steele, J., Szczesiul, S., Johnson, S. M. (2008). "Angling for access, bartering for change: How second-stage teachers experience differentiated roles." *Teachers College Record, Vol. 110*(5).

Drake, C. (2007). "Experience counts: Career stages and teachers' responses to mathematics education reform." *Educational Policy, Vol. 16*(2), 311-337.

Elmore, R. F. (2002). "The Testing Trap." *Harvard Magazine. September-October.*

Elmore, R. F. (2004). *School reform from the inside out: Policy, practice, and performance.* Cambridge: Harvard Education Press.

Evans, A. (2009). "No Child Left Behind and the Quest for Educational Equity: The Role of Teachers' Collective Sense of Efficacy." *Leadership and Policy in Schools (*8)64–91.

Finnigan, K. & Gross, P. (2007). "Do accountability policy sanctions influence teacher motivation?" American Educational Research Journal, Vol. 44, No. 3, 594-630 *American Educational Research Journal, Vol. 44,* No. 3, 594-630

Grossman, P. & Thompson, C. (2004). *Curriculum Materials:Scaffolds for New Teacher Learning?* Seattle, WA: Center for the Study of Teaching and Policy.

Haertel, E. H., & Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In E. H. Haertel & J. L. Herman (Eds.), *Uses andmisuses of data for educational accountability and improvement. The one hundredand fourth yearbook of the National Society for the Study of Education*(part 2, pp. 1–34). Chicago: National Society for the Study of Education.

Hatch, T., White, M., Faigenbaum, D. (2005). "Expertise, credibility, and influence: How teachers can influence policy, advance research, and improve performance." *Teachers College Record, Vol. 107*(5), 1004-1035.

Huberman, M. (1993). "The model of the independent artisan in teachers' professional relations, in Little, J. W. & McLauglin, M. W. (Eds.) *Teachers' work: Individuals, colleagues, and contexts,* New York: Teachers College Press.

Ingersoll, R. M. (2004). Why do high-poverty schools have difficulty staffing their classrooms with qualified teachers?Center for American Progress and the Institute for America's Future.

Ingersoll, R. (2007). "Short on Power, Long on Responsibility." *Educational Leadership, September.*

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. Future of Children, 17(1), 129-153. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ795883&site=ehost-live&scope=site

Johnson, S. M., & The Project on the Next Generation of Teachers. (2004). *Finders and keepers: Helping new teachers survive and thrive in our schools.* San Francisco: Jossey-Bass.

Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools.* New York: Basic Books, Inc.

Johnson, S. M. and S. E. Birkeland. (2003). "Pursuing a 'Sense of Success': New Teachers Explain their Career Decisions." *American Educational Research Journal* 40(3): 581–617.

Kanter, R. (1983), *The Change Masters.* NY: Simon & Schuster.

Kardos, S. M., Johnson, S. M., Peske, H. G., Kauffman, D. and Liu, E. (2001)."Counting on Colleagues: Teachers Encounter the Professional Cultures of Their Schools." *Educational Administration Quarterly* 37(2): 250–90.

Kauffman, D., Johnson, S.M., Kardos, S., Lui, E., & Peske, H. (2002). "Lost at Sea": New teachers' experiences with curriculum and assessment. *Teachers College Record, 104* (2), 2730300.

Korchoreck, J. W. (2001). Accountability and educational equity in the transformation of an urban district. *Education and Urban Society*, *33*(3), 284–304.

Kvale, S. (1996). *InterViews: An Introduction to Qualitative Research Interviewing.* Thousand Oaks, CA: Sage Publications, Inc.

Kennedy, M. (2005). *Inside Teaching: How classroom life undermines reform.* Cambridge, MA: Harvard University Press.

Lee, V. & Smith, J. (1996). Collective responsibility for learning and its effects on gains in achievement for early secondary school students. *American Journal of Education, 102*(4), 103- 147.

Little, J. W. (1990). "The persistence of privacy: Autonomy and initiative in teachers' professional relations." *Teachers College Record, Vol. 91*(4), 509-536.

Lortie, D. (1975). *Schoolteacher. A Sociological Study* (Second ed.). Chicago: University of Chicago Press.

Marvel, J., Lyter, D. M., Peltola, P., Strizek, G. A., Morton, B. & Rowland, R. (2007). "Teacher attrition and mobility: Results from the 2004-2005 teacher follow-up survey." Washington,D.C.: National Center for Education Statistics Institute for Education Sciences.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., and L. S. Hamilton. (2003). *Evaluating Value-Added Models for Teacher Accountability.* Santa Monica, CA: The RAND Corporation.

Murnane, R. J., & Phillips, B. R. (1981). "What do effective teachers of inner city children have in common?" *Social Science Research, 10*, 83-100.

Newmann, F.M, King, M.B, Rigdon, M (1997), "Accountability and school performance: implications from restructuring schools", *Harvard Educational Review*, Vol. 67 No.1, pp.41-74. Nyberg, D. (1981). *Power over Power: What Power Means in Ordinary Life, How It Is Related to Acting Freely, and What It Can Contribute to a Renovated Ethics of Education*. Ithica: Cornell University Press.

Nye, B, Konstantopoulos, S. & Hedges, L. (2004). "How large are teacher effects?" *Educational Evaluation and Policy Analysis, Vol. 26*(3), 237-257.

Resnick, L. & Glennan, T. (2002). "Leadership for Learning: A Theory of Action for Urban School Districts. In A.T. Hightower, M.S. Knapp, J.A. Marsh, and M.W. McLaughlin (Eds.) *School Districts and Instructional Renewal,* NY: Teachers College Press
.

Rockoff, J. (2003). *The impact of individual teachers on student achievement: Evidence from panel data*. Cambridge, MA: National Bureau of Economic Research.

Rorrer, A. K., & Sklra, L. (2005). Leaders as policy mediators: The reconceptualization of accountability. *Theory into Practice*, *44*(1), 53–62.

Rosenholtz, S., & Simpson, C. (1990). "Workplace conditions and the rise and fall of teachers' commitment." *Sociology of Education,* 63 (4), 241-257.

Spillane, J. "The Progress of Standards-Based Reforms and the Non-Monolithic Nature of the Local School District: Organizational andProfessional Considerations." *American Educational Research Journal,35(1).*

Stigler, J. & Hiebert, J. (1999). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom.* NY: The Free Press.

Szczesiul, S. A. (2007). "Initiatives and Initiative: Second-stage Teachers' Assessment of Autonomy," Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Timperley, H. & Alton-Lee, A. (2008) "Reframing teacher learning: An alternative policy approach to strengthening valued outcomes for diverse learners." *Review of Research in Education, Vol. 32,*(1), 328-369.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform.* Cambridge: Harvard University Press.

Valli, L. & Buese, D. (2007). "The changing roles of teachers in an age of high-stakes accountability." *American Educational Research Journal(3),* 519-558.

Appendix A – Detailed Sample Selection

District Selection

We drew our sample from 3 urban districts in the same northeastern state because research indicates that urban districts face greater challenges maintaining a qualified teaching force {{216 Jacob,Brian A. 2007; 229 Ingersoll, Richard M. 2004}. Selecting three districts enabled us to compare teachers' experience across districts. Because urban schools from the same state operate in the same state-level context with the same accountability requirements, selecting districts within the same state facilitated comparisons between districts and schools by eliminating state-level variation.

In each district, we obtained the district's permission for study and requested a list of all teachers that had been employed in the district from 1-10 years. Although we were only interested in speaking with teachers who had taught for 4-10 years, our decision to request district lists of teachers who had been employed in the district for 1-10 years, we found that because the lists did not reflect teachers' experience prior to entering the district, that it was necessary to invite some teachers who may have taught in the district for fewer than 4 years. Additionally, because some teachers with 1-10 years of experience in a district actually had more than 10 years of total teaching experience, some of the teachers on these lists were not able to participate in the study.

Despite this problem, the lists indicted roughly how many potential second-stage teachers worked at each school in the district. Thus we could target schools with more than one or two

second-stage teachers. These lists also enabled us to contact teachers without involving school personnel and, thus, we could protect the confidentiality of participants.

School Selection

Within each district we selected three to six schools from which we invited teachers to participate. In part, the schools were selected because each of them seemed to have ample second-stage teachers. Also, we aimed to include teachers who taught at the high school, middle school and elementary levels in our sample. Thus, we selected schools in each district to achieve this distribution (see Table 2). In District 1, we selected the one and only high school in the district and, two of the districts' K-8 schools. In District 2, we selected two high schools; one a mid-sized comprehensive high school, and the other was a small, alternative high school. We also selected two middle schools and two elementary schools. In District 3 we selected the only high school in the district as well as two middle schools and two elementary schools.

Except when there was only on high school, our decisions were based on a number of factors. First, we wanted schools with sufficient numbers of potential second-stage teachers. Second, we selected schools that were not extremes in their districts in terms of student demographics. That is to say, we selected schools that neither had the lowest nor the highest percentage of students with free & reduced priced lunch, percentage of students of color, or percentages of dropouts. Size was a factor when it came to the selection of the high schools in District 2, and we selected 2 schools that were very different in size. In total 14 schools were selected from these 3 districts, and principals were sent a letter informing them that we had received permission from the district to interview teachers in their schools.

Teacher selection

Once schools were chosen, we selected teachers from the lists provided by the district. Since these lists had teacher subjects, grade taught and years of experience, we were able to use them to ensure variation in all of these variables. We also tried to invite both men and women to participate, but had to make these determinations using only participant names as gender was not a field on these lists.

We sent invitations by mail to participants' schools. In each school we over-sampled anticipating that some teachers would refuse or not respond to our invitation[4]. Follow-up emails were sent to those teachers who did not initially respond. After 2-3 follow-up emails, we attempted to contact teaches by phone. Table 2 describes the number of invitations and response rate at each school.

We confirmed the teachers' years of experience prior to scheduling an interview, since many teachers had more experience than the lists indicated. In some cases, additional invitations had to be issued in an attempt to meet our school targets. Targets varied depending on the size of the schools[5].

We conducted individual interviews at the convenience of participants (See protocol in Appendix A). Most interviews occurred after school in the teachers' classrooms. On average, interviews lasted approximately 90 minutes. Each interview was recorded for the purpose of transcription. Participants' background information including age, education history, number of

[4] Since we obtained permission from districts at different times, teacher invitations in each district were sent out at different times. As such our over-sampling strategy evolved. In District 1, we minimally over-sampled teachers and spent a great deal of time contacting invitees. However, when it came to inviting participants form District 3, we had less time to contact participants before the end of the school year, and therefore, over-sample much more.

[5] For smaller schools we hoped for between 3-5 participants. We interviewed many more participants in the larger high schools.

students, and employment history, was collected prior to each interview. Later, each interview was transcribed, and names and identifiers changed in order to protect participant confidentiality.

Table 2. Obtaining the Sample of Second-Stage Teachers (SSTs)

	School Pseudonym	Total possible SSTs	Teachers invited	Invited teachers who were not SSTs	No Reply or Declined	Interviewed*
District 1	Gansford High School	52	14	2	4	8 (67%)
	Conifer School	39	16	1	9	6 (40%)
	Greentree School	39	20	3	11	6 (33%)
District 2	Bingham High School	9	9	2	3	4 (57%)
	Williams High School	36	25	1	12	10 (42%)
	Deer Park Middle School	10	9	0	2	6 (67%)
	Hill Middle School	36	16	1	10	5 (33%)
	Lane Elementary	8	8	2	1	5 (83%)
	Partridge Elementary	10	9	3	3	3 (50%)
District 3	Dryden High School	141	27	3	9	12 (50%)
	Montgomery Middle School	31	16	3	7	6 (46%)
	Allendale Middle School	25	10	2	3	5 (63%)
	Alpine Elementary	14	9	2	4	3 (43%)
	Keller Elementary	22	14	1	7	6 (46%)

*Percentage of teachers interviewed is calculated by dividing the number of teachers interviewed at each school by the number of teachers invited minus the number of teachers who were invited but were not STTs