

# **The Impact of High-Stakes Accountability on Teachers’ Professional Development: Evidence from the South**

## **A Final Report to the Spencer Foundation**

**Barnett Berry, Laura Turchi, & Dylan Johnson  
Southeast Center for Teaching Quality, Inc.**

**&**

**Dwight Hare & Deborah Duncan Owens, Mississippi State University  
Steve Clements, Kentucky Professional Standards Board**

**November 15, 2003**

The authors wish to thank John Williams, former program officer at the Spencer Foundation, who pushed our thinking about building capacity for educational research in and on the South. We also acknowledge the support of Diana Montgomery and Deb Ecker-Rich who assisted in field work and reviewing of initial drafts. We also wish to acknowledge the significant contributions of and partnership with Gary Henry and Darleen Opfer of Georgia State University, who directed the quantitative components of this investigation, and whose accompanying report provides a different, but complimentary angle to our own findings. Finally, we deeply appreciate the guidance and advice from a number of colleagues, Alan Tom from the University of North Carolina, Mark Smylie from the University of Illinois-Chicago, Bruce Haslam from Policy Studies Associates, Charles Thompson of East Carolina University, and Richard Ingersoll from the University of Pennsylvania. Each one offered advice and critique that helped us enormously in the design of our work and/or analysis of our data.

**Please direct correspondence to:  
Barnett Berry, President  
Southeast Center for Teaching Quality, Inc.  
976 Airport Road, Suite 250  
Chapel Hill, North Carolina 27514  
[www.teachingquality.org](http://www.teachingquality.org)  
[bberry@teachingquality.org](mailto:bberry@teachingquality.org)**

# **The Impact of High-Stakes Accountability on Teachers' Professional Development: Evidence from the South**

Barnett Berry, Laura Turchi, & Dylan Johnson  
Southeast Center for Teaching Quality, Inc.

and

Dwight Hare & Deborah Duncan Owens, Mississippi State University  
Steve Clements, Kentucky Professional Standards Board

## **Introduction**

This report presents the case-study findings from a cross-state investigation, *The Impact of High-Stakes Accountability on Teachers' Professional Development: Evidence from the South*. Our work, supported generously by the Spencer Foundation, represents collaboration between the Southeast Center for Teaching Quality (Center) and the Applied Research Center of Georgia State University (GSU). This research project, with a focus on Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee, is intended to deepen understanding around the influence of high stakes accountability on teacher learning and the capacity of districts and schools to support instructional practices that bolster student performance.

As part of the joint investigation, Center and GSU researchers sought to answer questions related to (1) the impact of different accountability and professional development policies in the six states, (2) how teachers respond to those learning opportunities in light of high stakes accountability, and (3) the relationship between high stakes accountability, teachers' professional development, and student achievement. While GSU researchers used teacher surveys and extant data bases to quantitatively respond to these issues, Center researchers drew upon case studies to focus primarily on the second issue — how teachers respond to professional development in the context of different high stakes accountability systems. In doing so, we sought to qualitatively answer three questions:

- 1. How do different state accountability systems affect teachers' responses to their own professional development?**
- 2. How do teachers from schools with different accountability ratings respond to their own professional development?**
- 3. How do schools and districts mediate high stakes accountability, professional development, and student achievement?**

We looked hard inside of 24 schools (in 12 districts), with different accountability performance ratings from the six states, which had developed or were developing different accountability systems with a range of rewards and sanctions for educators and students alike. Our unit of analysis was the school. While we were not able to analyze how instruction was changing in the schools studied, we were able to document and begin to explain how teachers, from schools with different performance ratings, responded to the accountability context in which they taught and how their school districts supported (or not) the kinds of professional development that would likely improve student achievement. We sought to understand under what conditions accountability systems would most likely promote effective professional development. We approached our work with a specific understanding of what an effective accountability system does. For us, an effective accountability system produces indicators that not only reveal which schools were doing better (or not), but also discloses why they were doing better (or not), and what educators should do to improve student learning (Darling-Hammond and Snyder, 1993).

Briefly, our site visits revealed that the high stakes accountability systems influenced teacher learning in a number of ways. Most prominently, we found that the accountability systems were helping teachers focus on academic standards, clarifying for them what needs to be taught and how. We found this to be case in both low and high performing schools. Teachers reported, rather uniformly, that their professional

development had been improving of late, with less of an emphasis on “one-shot” workshops and increased emphasis on more “hands-on” learning about how to teach to the state curriculum standards. Yet, we found very few examples, or any, of teachers’ professional development that mirrored what has been deemed effective by the research literature. This was the case even in the higher performing schools.

However, teachers and administrators in a number of states were often “baffled” by the statistical mechanics of the accountability systems, and could not easily “see” the relationship between how their students performed, how their schools were rated, what they needed to learn through professional development experiences and what they needed to do to raise student achievement. In part, teachers had had very few opportunities to learn how to use test scores and other data to unpack student learning. The lack of support for teachers to use professional knowledge in making judgments about student data limited the influence of the accountability systems. On the other hand, we found that district-level leadership and resources meant a great deal to teachers’ opportunities to learn — opportunities more likely to be found in the higher performing schools.

What follows is a brief review of some of the most germane literature that frames our work. Then, we offer a brief overview of testing, accountability, and professional development policies in each of our six states, followed by a description of our study methods. Next, we present our findings, organized by prevalent themes. We conclude with a summary of what we learned, framed by our research questions.

### **Accountability and Professional Development in Context**

*High Stakes Accountability.* According to Newman, King, and Rigdon (1997), a school accountability system should include four components: (1) information about an organization’s performance (test scores, graduation rates); (2) standards for assessing the organization’s performance; (3) consequences for success or failure (rewards and sanctions); and (4) an agency that collects information, judges whether or not standards have been met, and distributes rewards and sanctions. In any case, most school accountability systems today (as well as those from the recent past) appear to build upon several assumptions — students, teachers, and principals will work harder and achieve more when goals are clear, performance information is available and publicized, and “carrots and sticks” are placed in front of them (Hess, 2002; Moe, 2002). Sanctions and incentives for teachers, other educators, and students — based on standardized student test results — are being changed in state after state, in an effort to improve student performance and ultimately to increase the supply of educated workers in the region.

Some researchers have suggested that evidence supports the claim that states with high stakes accountability policies have achieved greater than expected success. In a recent 50-state analysis, Carnoy and Loeb (2002) have found a positive relationship between strong accountability systems (high degrees of external demands on schools to meet state standards) and improvement on the 4<sup>th</sup> and 8<sup>th</sup> grade NAEP assessments in math from 1996 and 2000). Other analyses have linked the results from the National Assessment of Educational in three states, Kentucky, North Carolina, and Texas, to their high stakes policies (Grissmer and Flanagan, 1998). In particular, a recent RAND report noted that from 1990 to 1996 North Carolina produced NAEP gains about twice as great as the national average, while also narrowing somewhat the achievement gap between minority and white students (Grissmer, Flanagan, Kawata, and Williamson, 2000). Other research has begun to support the use of teacher salary bonuses (Ladd 1999) and the importance of putting incentives for student achievement into place (Hanushek 1994). While approaches that use low-stakes accountability have been successful in other states, such as Connecticut (Baron 1999), at the time of our investigation in 2001-02, high stakes accountability had become a “way of life” in one southern state after another.

Indeed, there is a tendency to attribute student achievement gains to the features of standards-based reform that emphasize clear and ambitious student learning standards, student testing, an aligned assessment system, and a strict accountability framework. However, while these features appear to have played a role in the North Carolina’s success in closing the achievement gap during the 1990s, other analyses suggest that the pattern of student performance gains has much deeper roots. Rather, a long-term investment in the quality of the teaching force and support for teachers’ work began well before standards, assessment, and

accountability structures were fully built, and this investment appears central to the state's successes to date (Berry, Groves, and Noblit, forthcoming).

To be sure, high stakes reforms have been designed to construct new incentives and sanctions to get more of what policymakers want and expect out of public education. Their agents, those selected to respond to these incentives and increase educational performance, are teachers. While others must also change behaviors and act in ways that enable teachers to implement needed changes, teachers are the appointed agents and final arbiters of what gets taught and how. As Hawley and Valli (1998) note in their comprehensive review of research on the impact of professional development:

Teachers allocate and manage the students' time, set and communicate standards and expectations for student performance, and, in a multitude of ways, enhance or impede what students learn (p.4).

Critical to determining how teachers respond to changes brought about by high stakes accountability is how their development as professionals is shaped and molded to meet the mandates, incentives, and capacity-building opportunities presented by the system. A case can be made for looking at whether the kinds of accountability reforms that drive or influence professional development can make a difference for student achievement. Our hypotheses are that they do — or can — if the reforms are translated in ways that can support student learning. Recent research, focused on the New American Schools in the high stakes accountability environment of Maryland, suggested that teachers may not know the test sufficiently enough to create a rational change in their practices. One conclusion was that if teachers understood accountability systems and really understood what they need to know and implement to influence the results in a serious way, they would be willing to learn (Haslam, 2000).

Thus, no matter how well coordinated the high stakes accountability system may be, systems differ in ways that present a variety of objectives and incentives to the teachers charged with carrying them out. For example, rewards that go to all the teachers in a school are likely to motivate very different reactions than rewards that single out the performance of each individual teacher. In a similar vein, assessment systems that focus on reading and math provide a different incentive structure than assessments that also include science and social studies.

These issues, as well as our own findings reported herein, must be cast in light of implementing the recently enacted No Child Left Behind Act of 2001, which requires states to ensure that all students are academically proficient by 2013-14 (according to state standards) and that each school meet “annual yearly progress” (AYP) for each subgroup of students.<sup>1</sup> While each state has the leeway to select its own standardized achievement tests and set its proficiency benchmarks, schools can face serious sanctions for not meeting AYP standards. Students in schools that do not make AYP for two consecutive years are eligible for transfer to a higher performing public school, with the district required to use its federal dollars to pay for transportation. In mid 2003, several states announced the percentage of their schools *not* meeting AYP standards, ranging from 8 percent in Minnesota and 13 percent in Kansas, to 42 percent in Georgia, to 87 percent in Florida (Education Week, 2003).

Linn (2000) argues that there are few “clear-cut answers” as to whether assessment-based accountability models improve education — which strikes at the core of the validity, impact, and credibility of the system. Gains on standardized tests can be attributed to real student learning or other factors such as student selection, scale conversion errors, administration conditions, administration dates compared to norming dates, and teaching to the test (Linn, 1982). Linn's recent analysis (2000) has shown that states showing improvement in student achievement on their own measures do not always show gains on the National Assessment of Educational Progress:

A caution that might be taken from the [Minimum Competency Test] experience and results . . . Is that gains in the first few years following the introduction of a new testing requirement are generally much

---

<sup>1</sup> AYP requires improved student achievement for each subgroup of students by race, ethnicity, income, disability, and limited English fluency, and schools may be deemed failing if 95% of a particular subgroup does not take the test.

larger than those achieved after the program has been in place for several years. This tendency raises questions about the realism of some accountability systems that put in place straight-line improvement targets over extended periods (e.g., 20 years) (p. 7).

Accountability models vary considerably in terms of how data are collected, scored, and reported. Despite the fact that some states (e.g., several in the Southeast including South Carolina, North Carolina, and Tennessee) use more sophisticated hierarchical linear regression models to gauge student achievement gains, these advancements do not resolve the questions about what data the model should use. Linn and colleagues (2000, 2002) note that states may use a range of data based on any number of models, including comparisons of cross-sectional cohorts of students at different grade levels in the same year, or comparisons of cross-sectional cohorts in a fixed grade from one year to the next, longitudinal comparisons of school aggregate scores without requiring matched individual data, or longitudinal comparisons based only on matched student records. Each could have implications on what counts as schools organize instruction and professional development. This becomes even more complex, when looking at multi-year school improvement trends, such as those required for the federal Adequate Yearly Progress (AYP) measure (Linn et al, 2002).

Teachers are expected to react differently if their school has received high marks or “passing grades” than if they are in low performing or “failing” schools. Teachers in passing schools can pursue more long term or strategic teaching objectives through their staff development. Not so with teachers in low performing schools; they are much more likely to pursue short-term changes in teaching techniques to boost student performance. Similarly, teachers are expected to react differently if their students are promoted or not, and if they or their schools are labeled for such success or failure. For example, Shipps, Kahne, and Smylie (1999) found that in Chicago, teachers reported that retention did motivate *some* students to work harder and that reconstitution did motivate *some* educators to work harder.

However, it is difficult to ascertain how much of the recent test score gains can be attributed to the district’s system of sanctions and incentives, and if so, whether the system of sanctions and incentives prompts teaching to the test or professional development for teachers that can sustain long-term and more robust measures of academic achievement. The number of years that the assessments have been given is also likely to influence teachers’ reactions. The longer they have been given, the more information that teachers will have; and their responses are likely to be more strategic and rational. Pedulla and associates (2003) also found a significant relationship between the level of stakes for schools and students and the extent that teachers rely on results to drive their teaching, especially among elementary teachers.

*Highly Qualified Teachers and Professional Development.* In addition, the NCLB law specifies that all teachers are “highly qualified” by 2005-06, determined primarily through passing a subject-area test, and that states report on the percent of their teachers receiving “high-quality” professional development. While states have received “reams of federal guidance interpreting the accountability provisions of NCLB,” they have had little assistance in understanding and implementing the teacher quality provisions of the law. Although the federal government is now allocating up to \$3 billion a year to support the teacher development needs of schools and districts in light of NCLB, little evidence exists as to how much of these dollars are new and what kinds of capacity schools have to use the money to improve teacher and student learning (Education Trust, 2003).

Over the last decade, the education research community has devoted significant attention to understanding what and how teachers learn and the implications for instructional improvement and student learning. There is increasing, albeit still spotty, evidence of the kinds of professional development that leads to student achievement gains. Recently, a few studies have surfaced that the *most effective professional development focuses on the specific content that students will learn and the specific difficulties students encounter in learning the content* (Kennedy, 1998).

One such study by Cohen and Hill (1997), a survey of 1000 elementary school teachers in California, found that schools were more likely to have high student achievement when teachers took part in professional development that focused on specific curriculum issues (teaching fractions, for example). Equally important, the teachers in higher achieving schools, had opportunities to work with other teachers, to use

research methods in studying what their students did and did not know, and to improve their lessons based on what they learned. When teachers spent most of their staff development time studying general education strategies, their students did not perform nearly as well. These findings suggest that professional development needs not to focus on generic (one size fits all) teaching behaviors, but on the analysis of curriculum and student responses to it. Cohen, Raudenbush, and Ball (2000) have noted that teacher learning that is linked to student achievement may hinge on the “strategic documentation of practice” that “cultivates (their) capacities to investigate teaching and learning” (e.g., assessing student work samples and video tapes of focused lessons on teaching two-digit multiplication).

A recent large-scale study, utilizing surveys and case studies, surfaced six critical factors (underlined below) for creating effective professional development that is linked to the development of teacher knowledge and changes in teaching practices (Birman, Desimone, Porter, and Garet, 2000). The researchers concluded that:

Professional development should focus on deepening teachers’ content knowledge and knowledge of how students learn particular content, on providing opportunities for active learning, and on encouraging coherence in teachers’ professional development experiences. School and districts should pursue these goals by using activities that have greater duration and that involve collective participation. Although reform forms of professional development are more effective than traditional reforms, the advantages reform activities are explained primarily by greater duration of the activities (p. 32).

The investigation of Birman et. al. reveals how structural characteristics (form, duration, and participation) affect core features (content, active learning, coherence), that in turn predict increases in teacher knowledge and skills, and changes in teaching practices. However, their surveys and case studies revealed a mix of high and low-quality structural and core features — and that most professional development in the Eisenhower grant program was: (1) traditional in form, (2) less than a week in duration, (3) increasingly focused on content, but with very little opportunity for actively learning the content, and (4) increasingly coherent with state and district standards, but rarely cohering with other aspects of the system (e.g., teacher evaluation and building on other professional development).

The literature, such as the findings described previously, surfaced a high degree of consent regarding effective professional development. However, while built from a growing consensus in the field and from emergent research, most of the claims for effective professional development linked to student achievement are largely unsupported by rigorous empirical study (Cohen, Raudenbush, and Ball, 2000). Little (1993) found that teachers engaged productively in subject specific associations, special institutes and centers, subject specific projects, and university and school collaboratives implemented greater changes in practice than those who do not. Little (1993) also indicates that providing opportunities for teacher inquiry/research projects shows promise for teacher change. Nevertheless, several years ago, the National Partnership for Excellence & Accountability in Teaching highlighted eight research-based criteria for identifying effective professional development. These include practices that:

- are based on analyses of standards and actual student performance,
- engage teachers in identifying their learning needs,
- are primarily school-based and built into the day-to-day work of teaching,
- are organized around collective problem solving,
- involve follow-up support,
- use information on student outcomes, instruction, and the implementation of lessons;
- focus on understanding theories underlying knowledge and skills to be learned; and
- are connected to overall school change process (Hawley and Valli, 1998).

*Considering accountability and professional development in context.* There is a growing literature on how educators respond to external pressures. For example, as Abelman and Elmore (1998) have asserted:

Nested within these developing external accountability systems are real schools: schools that have their own distinctive organizational characteristics and problems; schools that have unique student populations; schools situated in diverse and particular communities; and schools with their own institutional histories. The reality of particular schools belies the pressure for uniformity behind the emerging accountability systems. External accountability systems assume a world in which all schools are held to the same expectations for student performance. The world that school administrators see, however, is bounded by their particular settings, by their own conceptions of who they are, who they serve, what they expect of students, and what they think of as good teaching and learning (p. 1).

Firestone (1989) found three main approaches used by districts in responding to state testing policies. One approach is *fragmented*, with districts, having limited capacity or “know-how,” offering little leadership for change and focus on the test. The second approach is *communicative*, with districts making some efforts to coordinate policies and programs, but still showing little focus beyond what is tested. The third approach is *coordinated*, with very deliberate efforts to align efforts and activities, and offering substantive leadership for changing instructional practices. Spillane and Thompson (1998) found greater differences in district responses within the school systems themselves. Based on their research, districts neither adopted fully nor ignored completely state policy mandates, their responses were more tempered.

Goertz and Massell (1998), in a study of 14 school districts, found significant adaptations to state accountability systems. Their work revealed that the nature of district responsiveness was determined by capacity, size, understanding, leadership, organization and governance, political climate, and the nature of state policy itself. Regarding the latter, Goertz and Massell (1998) revealed clear patterns in district responsiveness, depending on the strength or weakness of the state accountability system. And, the relative strength or weakness of the accountability system was due to prescriptiveness, rigor of measures used, alignment of system with other policies, and the stability of the system itself. Firestone and Fariman (1998) found formal sanctions, like forced reconstitution of consistently low performing schools, were more likely to promote responses to state testing policy than just “embarrassment” from grading schools and reporting results publicly.

In a review of research on the effects of accountability, Stecher (2002) outlined a long list of documented positive and negative outcomes, while warning that there is still much to know about the effects of school accountability and school improvement policies. He warned that while some components of a reform strategy may be reasonably well understood, such as the use of certain materials or the amount of time spent on math instruction, the quality of those activities may not be as clearly defined.

Pedulla et al (2003) surveyed 12,000 teachers nationwide on outcomes of state testing programs, categorizing responses according to both school type (elementary, middle, and high) and level of stakes in the state system for both districts/schools/teachers and students themselves. Their analysis showed that higher stakes for educators was related to more district alignment of curriculum and materials to the state tests, according to teachers. They found that those in states with the highest stakes felt pressure to improve test scores from both their superintendents and principals—and elementary teachers felt this pressure more often than high school peers.

Nevertheless, Cohen (1996) has shown that very little is known about how incentives affect school performance, how problematic it is to create powerful incentives absent of clear consensus on educational goals, and how strong incentives are unlikely, in and of themselves, to improve poor performing schools. Research on efforts used to increase performance of schools illustrates that the critical factor is *how the schools and teachers respond to changes in the incentives* as well as opportunities for capacity building that are presented to them (Murnane and Levy 1996; Ginsberg and Berry, 1996; Darling-Hammond, 1997; Elmore, 1997; Massell, 1998). Student standards, while apparently alike on paper from one state to the next, may be perceived differently by different teachers. We have hypothesized that teachers’ beliefs and reactions toward accountability and student standards would reflect the type of professional development they would pursue.

Finally, Newman, King, and Rigdon (1997) found that schools with strong internal accountability systems, including the means for staff to identify standards for student performance, collect data to measure their progress, and use peer review to drive goal attainment, were more likely to increase their organizational capacity. They found that internal accountability accounted for a “major source of cohesion” with the schools — and thus could be viewed as both a “building block” for and a “product” of high organizational capacity (p. 48). They also found that *organizational capacity rested on teachers’ professional knowledge and skill, effective leadership, availability of technical and financial resources, organizational autonomy to act according to demands of local context, and the effective and collective management of the human, technical, and social resources of a school*. Strong external accountability systems may help clarify educational outcomes, but they may not necessarily align themselves with “internal cultures of schools that define their operating educational missions” (p. 48). In some cases, internal and external accountability may be “hostile” to each other and may undermine a school’s organizational capacity for responding favorably to intended achievement goals. Nevertheless, they suggest strongly that, “efforts to increase external accountability should be coupled closely with efforts to enhance school organizational capacity” (p. 49).

Although one would assume that organizational capacity is endemic to the internal workings of schools and districts, external agents can help build it through new and strategic technical resources, professional development, and incentives that strengthen school autonomy (Newman and Wehlage, 1995). Clearly, the interaction between the incentives presented through state policies and local culture, goals, policies, and priorities are extremely influential in determining how schools will respond to high stakes accountability (Elmore 1996; Cohen 1996). Moreover, a number of diverse authors and analysts have been arguing for some time that many schools lack the capacity to meet higher standards, irrespective of the kinds of accountability systems and subsequent rewards and sanctions (Kearns and Doyle, 1988; Hanushek, 1988; Clune, 1993; Massell, 1998; Rothman, 1995).

We pursue these matters more deeply in understanding the effects of high stakes accountability on teacher learning. However, next we will briefly review accountability and professional development policies in the six states at the time of our study.

### **Accountability and Professional Development in the Six States.**

The review of testing, accountability, and professional development policies revealed considerable differences across the six states. Indeed, we had expected to see how the different variations in these policies would provide incentives for teachers to respond as agents of reform. In constructing the policy profiles, the GSU research team reviewed existing statutes and policy documents including accountability legislation and interviewed key informants such as state directors of accountability and professional development. The team also reviewed extant data on accountability and professional development structures and resources and obtained fine-grained information about the types of test used, how ratings were developed, the kinds of sanctions and incentives employed, how information is reported and used, and the timing of state reforms in order to establish when effects may be expected to register (See accompanying report by Henry and Opher, 2003). Importantly, the descriptions below, synthesized from data and information collected by the GSU team, were assembled in 2001, prior to the field work we conducted.

*Testing.* All states are undergoing serious changes in their testing policies. Most states had been using multiple-choice tests (some locally developed) with a primary focus on math, reading, and language arts. All states, except Kentucky, have or will have high school exit exams.

Kentucky’s original testing system included reading, math, science, social studies, and writing assessments, and initially just at grades 4, 8, and 12. The original assessments included both NAEP-like multiple-choice items, as well as open-ended questions and writing responses. (See Appendix A for sample of “NAEP-like” items in Grade 4 math and reading.) The latter required trained scorers, primarily classroom teachers. In 1996-97, the state tests, called KIRIS, began to be administered in grades 4,5,7,8, 11, and 12, with different subjects tested in different years.



In North Carolina, state testing occurs in grades 3-8, with assessments covering reading and math. End of course assessments are used in high schools and cover Algebra I & II, Biology, Chemistry, Economic, Legal, and Political Systems; English I & II, Geometry; US History; Physical Science; and Physics. Tests are multiple choice, with a prescriptive writing test given in grades 4 and 7.

In 1999, South Carolina began using the Palmetto Achievement Challenge Test (PACT), a much more rigorous, criterion referenced test that assesses math, reading, language arts, science, and social studies at nearly every grade in elementary and middle schools. The state spent more than \$13 million in launching this new assessment — whose developers relied a great deal on the NAEP framework for item development, performance categories, and cut scores. South Carolina is now adding a science assessment.

Tennessee’s current testing program, launched in 1990, is a mixed norm- and criterion-referenced basic skills test administered to grades 3-8, and assesses language arts, math, science, and social studies at each level. There is a writing component at grades 4, 7, and 11. Mississippi students have been tested in reading and math — in grades 3, 5, and 8 by a state developed criterion referenced test and grades 4 and 6 by the norm-referenced test, ITBS. Georgia’s students had been assessed with a norm referenced and a writing test in grades 3, 5, and 8.

On whole, the states are adding tests and are more frequently testing students (The trend toward increased testing was well in place before No Child Left Behind). For example, Georgia, with its 2000 legislation, is now testing kindergarten students, all students in grades 1-8 in language arts, math, and reading, all students in grades 3-8 in science and social studies, writing in grades 3,5, 8, and 11, norm-referenced tests in grades 3, 5, and 8, and a high school end of course tests in 8 different subjects, and a high school exit exam (as well as participating in NAEP).

**Table 1: Testing Policies**

Testing Policies	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Types used	NRT, CRT, WT, EE	NRT, CRT, WT	NRT, EE	CRT, WT, HS	CRT, EE	Combined CRT & NRT, EE
Tests in 1999-00	NRT 3,5,8; WT 3,5,8, 11	NRT 3,5,10 WT 4,7,12 CRT 4,5,7, 8,10,11	NRT 5,8 EE 11	CRT 3-8 WT 4,7 EE 9, 10 HS 9-12	CRT 3-8 EE 10	NRT/CRT 3-8; WT 4,7,11 EE 9
Changes planned	CRT 1-8 HS to replace EE	None	CRT 2-8 WT 4, 7,10 HS to replace EE	EE in 2004; Gateway exams in 3,5, 8 by 2005	CRT in science & SS; CRT to replace EE in 2003	HS to serve as gateways

NRT= norm referenced test, CRT= criterion referenced test, WT= writing test, EE = exit exam, HS – high school subject tests

*Accountability.* Kentucky and North Carolina have been holding schools accountable for a decade. In North Carolina, the state has developed NAEP-like proficiency standards for each of these tests. The calculation formula for labeling schools is complex, but it is advertised as expecting schools to “produce a year’s worth of growth for a year’s worth of school.” The key to the system are the “growth rates” that become constants in the formula. Schools are labeled also according to absolute performance. Based on aggregated scores, schools can be labeled in terms of “excellence” or “distinction” or “adequate” or “low performing.” In 1998-99, 81 percent of the state's public schools achieved expected or exemplary status. In 1999-2000, this number dropped to 70 percent, as absolute scores rose and potential for growth was reduced.

If schools are designated "low-performing" in North Carolina, meaning that growth targets are not met and fewer than 50 percent of students are scoring at grade level, technical assistance teams are assigned by the state. In 1999-00, only 45 schools, or less than 2 percent of all of those in the state, were labeled low performing. This number has continued to decrease of late, with only 18 low-performing schools in 2001-02. On the other hand, teachers from schools in the top categories can be paid an additional \$1500 (with teaching assistants being paid \$500). Rewards can vary depending on the state's budget allocation for the financial incentives. (Note: Nearly 75 percent of the state's schools earned high growth in 2002-03, compared with about 33 percent in 2001-02.)

In Kentucky, its reform legislation set baseline scores in 1992 and goals so that all schools would reach a "proficient" level in twenty years. The original assessments included both NAEP-like multiple-choice items, as well as open-ended questions and writing responses. The latter required trained scorers, primarily classroom teachers. However, these assessment items never figured prominently in the accountability index.

Based on assessments administered in 2-year cycles, schools were compared to original baseline scores (1992), and placed into a performance category. Schools could be eligible for monetary rewards, if they had at least a one-point gain over their threshold, and at least 10 percent of its students moved to a higher level. For example, in 1995, the state awarded more than \$26 million in bonus pay to teachers, with stipends ranging from \$1300 to \$2600. Schools could also be deemed successful, not meeting threshold, in decline, or in crisis. The latter two categories earned schools assigned "distinguished educators" (later called "highly skilled educators").

However, in the late 1990s, numerous alterations to this system began to unfold. Teachers never had much guidance in teaching to the standards (except for being issued a "phonebook-like" document); the accountability system was deemed "complex to explain"; and the testing system was expensive to administer (estimated at \$10 million per year). But, due to growing concerns about how the system actually worked, in 1998, the state switched to less complicated criterion referenced test, began drawing upon a more specified core content, and also used a normed reference test in several grades, and recalibrated the weighting formula that assigns schools to performance categories.

On the other hand, other states like Georgia and Mississippi have only passed new accountability policies of late. For example, until the passing of the A Plus Act, Georgia had a fairly minimal accountability system. Before 2000, state accountability consisted of public reporting on education data, voluntary accreditation, and self-identification as needing improvement, but very limited rewards and sanctions were in place for schools and districts.

The Mississippi Student Achievement Act of 1999 moved the state toward rewarding schools rather than districts, for student performance. The state was in the process of developing a new assessment system, similar to North Carolina's model. Priority schools were to be designated by not meeting expectations and having a certain percentage of students perform below grade level (still in planning at the time of the profile writing). There was no financial reward provision for high-performing schools.

While Tennessee has yet to establish incentives and rewards for its accountability system, the state has set standards for student achievement and publishes school-level results, using a well-publicized valued-added assessment methodology. The state does not reward schools, but does identify poor performing ones. Three levels of sanctions are applied to low performing schools: first "on notice," then "probation," and finally, "state takeover." School-level reports were issued first in 1999-2000. In 2001, the commissioner placed the 48 lowest schools (half of them in Memphis) on probationary status. The next year, 98 schools were placed "on notice," with a warning that they would be placed on probation if no improvement was achieved in one year. After two years on probation, they would be taken over. Some critics have disputed the legality of such punitive actions without school support or funding assistance to these schools. There has never been sufficient resources to reward high performers as outlined in EIA. Indeed, not every state provides funding for implementation of their accountability policies.

**Table 2: Accountability Policies**

Accountability Policies	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Date of Legislation	2000	1990	1999	1993	1998	1992
Implementation	2004	1994	2002	1998	2001	1997
Rewards: 1999-00	No	Yes (schools & teachers)	No	Yes (teachers)	No	Not funded
Sanctions: 1999-00 (excluding labels)	No	Job loss, takeover	Job loss, loss of accreditation	Maybe	No	State takeover
Funding attached to accountability	No	Yes	No	Yes	Yes	Yes

At the time our study, professional development policies had little connection to accountability or testing policies. All states required teachers to be recertified by taking professional development units, usually the equivalent of 6 semester hours every five years. While all states except Mississippi supported professional development in some way, most of the responsibility rested on federal and local funding. Two states actually invested in professional development in ways tied to accountability. Both Kentucky and North Carolina offered assistance team models to low performing schools, and had apparently increased spending on professional development in previous years.

In Kentucky, with its 1990 reform bill, the state began spending \$23 per pupil on professional development, up from only \$1 per pupil. Not only did the state begin spending about \$15 million on teacher learning, it also required that teachers have four days of professional development a year (and encouraged districts to add 5 more days). According to researchers studying KERA over the last several years, Kentucky teachers have become more directly involved in planning their own learning, but have focused much of their professional development on traditional delivery models and more on the tests and test-taking skills than on core content knowledge and how to teach it (Corcoran and McDiramid, 2000).

In North Carolina, approximately \$12 million in state funds during 1999-00 were dedicated to professional development. However, this amount is estimated as a small fraction of what is spent by school districts (including federal flow through dollars). Each school is required to develop its own staff development plan (as part of a larger school improvement plan). The state department of education cooperates with the University of North Carolina system office, which administers a wide range of professional development offerings, including a principals executive training program, a teaching academy, a math and science network and the highly respected North Carolina Center for the Advancement of Teaching. The state has made the most significant investment in National Board Certified Teachers, with over 5100 teachers now earning that distinction.

In 2000, then Governor Roy Barnes of Georgia promoted and signed legislation that changed the way local school systems were allocated funds for professional development and raised the allocation formula from 1/2 percent to 1.5 percent of salaries of all certificated professional personnel. (However, due to economic conditions, during the middle of the fiscal year the percentage factor was changed from 1.5 percent to 1 percent.)

However, the decentralization of professional development meant that very little information could be gleaned about policies and practices at the state level. Key state informants revealed a wide variety of responses to questions about how professional development decisions were made.

**Table 3**

Professional Development Policies	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Where decisions generally made	State and district	Schools and teachers	Districts	Schools	Districts and schools	Districts & schools
Types of financing	Fed, state, & district	Fed, state, & district	Federal & district	Fed, state, & district	Fed, state, & district	Fed, state, & district
Additional funding for accountability	No	Yes & low performing	No	Low performing	Yes	No
Recertification – 1999-00 (5 year cycle)	6 sem. hrs or 120 units	6 sem. hrs or 90 units/5 years	10 CEUs in content or 3 semester hours	15 CEUs or 10 years for NBPTS	6 sem hrs.	6 sem hrs or 90 units
Require Induction 1999-00	No	Yes	No	Yes	Yes	No

### The Study

In early 2001, the Center researchers met with the GSU team to discuss the selection of states and schools/districts for the sample. Each state was categorized, for the 1999-00 school year, using decision rules for each element of the accountability system, with specific attention to types of tests used, how ratings are developed, the kinds of sanctions and incentives employed, and the use of student performance information (see above).

The six states were selected to maximize the differences in the extent to which rewards, sanctions, and interventions associated with high stakes accountability have been implemented. Careful attention was paid to the timing of state reforms, in order to establish when effects could be expected to register. Twelve districts were chosen to represent maximum difference, under the assumption that this would help to understand the district effect (within the context of the school site cases). Selected districts had low and average performing schools, without any high performing schools, or they had high and average performing schools, but no low performing schools. In total, we investigated 24 schools in these 12 districts across the six states.

	High Performing Schools	Average Performing Schools	Low Performing Schools
High Performing Districts	6 schools	6 schools	0 schools
Low Performing Districts	0 schools	6 schools	6 schools

(A more detailed summary of the school sites at the end of this section).<sup>2</sup>

In each district, human resource and professional development administrators were interviewed and testing information and professional development archives were collected and reviewed. In each school, approximately 8 teachers were interviewed for about an hour each, with careful attention to select those who were in teaching at grade levels targeted by the respective accountability systems. Principals and assistant principals were interviewed as well. Overall, we interviewed more than 250 teachers and principals.

<sup>2</sup> See Henry and Opher (2003) for information on how the research team selected schools and districts based upon available test score evidence

Both interview and other data were coded by teacher responses to *state and local accountability* (with focus on incentives, reporting or recognitions, standards, assessments, assistance toward meeting state goals), *teacher learning opportunities* (with a focus on sources, qualities, participation, content), *organizational capacity to support teacher learning* (e.g., materials, time, money, incentives, collegiality, use of data, efficacy, expectations, and political), *curriculum and instruction* (subjects taught, time allocated, intellectual depth, pedagogical approaches, and learning environment), and *perceptions of student learning* (basic skills, problem solving, and social development). For each school, the case study researchers assembled evidence from the interviews (and artifacts collected on site) demonstrating the most prevalent teachers responses. These teacher responses were examined in light of principal and district office responses, and were organized and displayed with the use of a data collection guide (See Appendix A).

Our unit of analysis was the school, and thus the team created specifications for the development of 24 case study memos (one for each school). The case memo was designed to capture a rich picture of individual schools, but also to allow for an understanding of how teachers in schools with different performance ratings may respond differently to accountability systems, with different incentives and sanctions.

Separate school information in the case memos provided history and the political context of individual sites. The researchers reported school variations on district contexts, where school leaders were more autonomous in response to accountability requirements, especially in terms of professional development opportunities and budgeting. The school memos then focused on teacher reports of their participation in learning experiences, some of which were characterized as professional development. Teachers also reported the incentives offered for professional development, their participation in an enormous range of in-house as well as external initiatives. In each interview, teachers were pressed for the impact of their professional development activities on their classroom practices, including what they taught and how their students were learning. The school memos also collected administrator and teacher reports of new or re-allocated resources, as well as other indicators of ways accountability requirements were shaping their schools.

Finally, the case memos captured the evidence of locus of control and responsibility toward student achievement. Researchers solicited descriptions of the ways teachers and administrators were motivated toward school improvement and student success, and sought to discover the degree to which motivation for improvement -- as defined by high stakes accountability -- was perceived to be part of a school culture or the result of external pressures. Through these questions, researchers could tease out the reasoning behind school and teachers' goals for student success, and find evidence of faculty agreement or common understandings of purpose.

**Table 4: Summary of School Sites**

State	Perf. Category *	FRL **	Minority **	Community **	Students (#)-**	Teachers**	Turnover Rate***
GA-1	H-H	2%	3%	Rural (8)	879	52	26%
GA-2	A-H	16%	15%	Rural (8)	938	64	12%
GA-3	A-L	87%	47%	Rural (7)	919	60	13%
GA-4	L-L	99%	85%	Rural (7)	348	25	15%
KY-5	H-H	37%	2%	Rural (7)	210	8	N/A
KY-6	A-H	39%	.2%	Rural (7)	412	14	N/A
KY-7	A-L	77%	.6%	Rural (7)	166	14	N/A
KY-8	L-L	79%	0%	Rural (7)	149	14	N/A
MS-9	H-H	51%	9%	Rural (7)	379	19	19%
MS-10	A-H	51%	9%	Rural (7)	992	59	7%
MS-11	A-L	64%	50%	Rural (8)	395	31	30%
MS-12	L-L	64%	94%	Mid-size (2)	293	27	41%
NC-13	H-H	36%	8%	Rural (7)	211	13	24%
NC-14	A-H	41%	23%	Rural (8)	355	18	24%
NC-15	A-L	95%	95%	Mid-size (2)	351	44	16%

State	Perf. Category *	FRL **	Minority **	Community **	Students (#)-**	Teachers**	Turnover Rate***
NC-16	L-L	88%	100%	Mid-size (2)	550	42	28%
SC-17	H-H	31%	16%	Suburban (4)	485	37	11%
SC-18	A-H	58%	9%	Rural (8)	1032	74	16%
SC-19	A L	87%	98%	Rural (7)	375	24	15%
SC-20	L-L	96%	89%	Rural (7)	429	35	15%
TN-21	H-H	36%	12%	Large (5)	382	N/A	N/A
TN-22	A-H	72%	27%	Large (5)	289	N/A	N/A
TN-23	A-L	61%	39%	Rural (8)	629	N/A	N/A
TN-24	L-L	95%	88%	Rural (8)	229	N/A	N/A

\*The first letter designates the performance rating of the district (high, average, or low) and second letter designates the performance level of the school (high or low). In the narrative we use HPS (high performing school) and LPS (low performing school) as abbreviated identifiers.

\*\*Calculated from the NCES Common Core of Data, 2001-2002.

\*\*\*From the Southeast Center for Teaching Quality's Teaching Quality Indicator Project, 2000-2001. Numbers represent the percentage of full-time teachers employed at the school during 2000-2001, but not in 2001-2002.

### What We Found: Results from the Case Studies

A working hypothesis for our study was that the variations in accountability, testing, and professional development policies would send different signals to teachers who then would respond as agents of school improvement. We clearly found how the current accountability systems (as of 2001-02) were helping teachers focus more clearly on what needed to be taught. In this sense, the accountability systems were sending signals that policymakers assumed needed to be sent.

However, these signals seemed to only evoke “loose connections” between high-stakes accountability and professional development. Indeed, across the six states and various low to high performing schools visited, we found little evidence of teachers receiving high quality professional development (as defined in the literature). Nevertheless, what was clear was that the accountability system in each state had the attention of teachers and administrators, who revealed considerable commitment to their students and were quite conscious of increasing demands for student performance. Not surprisingly, schools differed in leadership, resources, and capacity to act on these demands. The literature is replete with such findings, but what we found reveals considerable nuanced understandings of the relationship between accountability systems and teacher learning that can lead to increased student achievement.

To be sure, each of the states' current (or developing) accountability systems appeared to help teachers focus more *on* standards. In states with higher stakes associated with scores (established rewards and sanctions), teachers tended to focus more *on* what they knew to be on the tests, rather than the more abstract or general state standards. This was not necessarily a negative consequence when teachers — like those in Kentucky and South Carolina — viewed the state assessment as testing more important content and thinking skills. In these cases, they responded by welcoming new accountability-related activities and materials into their classrooms.

However, we found that oft-misunderstood test data and rankings can also undermine accountability. In a number of cases, we found teachers and principals having difficulty explaining their accountability scores — and thus had problems in responding to the signals sent by the system of rewards and sanctions. Indeed, teachers interviewed across the 24 schools reported that their recent professional development had improved. However, because teachers and administrators articulated relatively few direct links between accountability and professional development, it is not surprising that there was little demand for high quality professional development (except in districts led by more astute superintendents). More startling, however, were the few organizational mechanisms that linked usable student performance data to teacher learning opportunities.

The challenge then is to understand the investments of systems and schools (often of very limited human, monetary, and community resources). Even in states where accountability systems were still being implemented, where rewards and sanctions were still on the horizon, teachers knew that they needed to “do something.” Unfortunately, teachers in poor communities, serving low-performing students are beset, not as much by low expectations, but by the pervasive belief that there is only so much they can do to improve student achievement. Many of these teachers’ beliefs seemed embedded in weak district and school leadership, the lack of sound professional development, and higher turnover yielding more uncertified and/or inexperienced colleagues who limit long term school improvement. Teachers in higher performing, better resourced schools, serving wealthier students, were likely to have more control over their own professional development, draw upon informal as well as formal learning opportunities, and were more likely to be treated as professionals. While teachers in lower performing, less-resourced schools, serving poorer students, had access to additional state assistance, the quality and impact of that assistance was extremely limited. Inequities in organizational capacity to respond to high stakes accountability were evident and profound. In these schools and districts, we found teachers narrowing the curriculum and reporting more “stress and strain.” While the stress and strain has resulted from more public scrutiny, we also found that the pressure has also resulted from teachers who are overworked and exhausted from trying to cover increasing amounts of content and then having to teach to the test.

What follows are a set of themes and issues that surfaced from our site visits. We focus on differences in labeled and non-labeled states, as well as differences between higher and lower performing schools. [In defining these differences, we use short-hand designations; such as LPS for low performing school or HPD for high performing districts.]

### **Accountability and a New Focus on Standards**

For the most part, and across all 24 schools, we found how the accountability system helped educators focus more on standards and at least indirectly offered them guidance on what and how to teach. Each case offered evidence of school changes made in response to accountability systems. Educators in most cases seemed comfortable with the state standards, as framed by their respective accountability systems.

State standards have provided teachers with information about expectations for student performance, and in this way, indicated their responsibilities as educators. In a Kentucky APS, teachers recognized that the rules to be followed and the status of their school might change with each year’s testing, yet they reported that the accountability system gave them more knowledge of “where students should be,” and a clearer definition of their “job.” One teacher described herself as being “much more aware of what she should be teaching and where kids ought to be in terms of performance at given points in the year.” Kentucky teachers from a HPS described the accountability system as helping the school community design a comprehensive program (a curriculum, assessment, and professional development program) where all students can learn. Also, teachers in an APS (in a LPD) in North Carolina noted their appreciation for the state’s Standard Course of Study because it kept teachers from teaching “whatever they wanted.” As one North Carolina superintendent noted:

I remember when a lot of what went on in a classroom was determined *strictly* by the teacher. And many times, it was dictated by (their) favorite subjects — like one history teacher I had who could spend the whole semester talking about the Civil War and nothing else. He knew his subject *that well*.... but now he could not spend a whole semester talking about the Civil War — and that is a good thing.

Mississippi teachers from an APS reported confidence that the curriculum frameworks and district benchmarks were correctly aligned to the state tests (while at the same time they felt the required paperwork documenting interventions and services for students were overwhelming).

Teachers in each state responded quite consistently in their assessment of the quality of the standardized test used, irrespective of their schools’ performance level. We found that when teachers viewed the accountability test as assessing more important content and thinking skills — like in Kentucky and South

Carolina — they responded more favorably to the standards and the assessments and were more likely to use them to guide their practice. For example, teachers in Kentucky appreciated the performance-oriented nature of the writing assessment, in which students must write and reason, rather than simply select from multiple-choice items, although some expressed doubts about the accuracy of scoring tests like these. In South Carolina, teachers from all types of schools praised the new PACT tests for pushing students “to think like scientists.” Veteran teachers in South Carolina, who had been accustomed to the state’s 20-year old basic skills assessment and accountability system, reported satisfaction with the state tests because the new tests required more than simple recall of facts. The teachers implied that having good tests like these made it reasonable for them to focus heavily on test-taking strategies. As one teacher noted, “teaching to a good test was better than teaching to a bad one.”

Indeed, teachers in a Kentucky APS also described teaching to the test as a good thing, because it ensured coverage of the state’s curriculum. But, these teachers feared that the software their district was using to guide curriculum planning would push them to cram in even more activities into a school day that was becoming too crowded. On the other hand, in a LPD in Tennessee (where teachers did not think as highly of the state test), teachers believed the county curriculum guide to be useless, as a political crony of a district administrator created the curriculum guide. At the same time, several teachers in the school pointed to a teacher who had the highest test scores, but who bragged about teaching to the test and not doing much of any other kind of teaching. In this LPD, we found very little evidence of sound professional development practices in place, although teachers here and elsewhere were likely to report that their learning opportunities had been improving of late.

#### **A More Focused, But Narrow Curriculum**

An almost inevitable result of high-stakes accountability is a more focused, but also more narrow curriculum. In Mississippi, North Carolina, and South Carolina, teachers discussed the changing emphasis of their curriculum as influenced by the expectations of their respective state accountability systems. For example, teachers in Mississippi schools reported a primary emphasis on reading and language arts, with a secondary focus on math. Science and social studies lessons were rarely discussed, and apparently very little instructional time was allocated for either. North Carolina teachers spoke of the need to try to teach science and social studies lessons “through reading” even though they had very little training to do so. In South Carolina, science was scheduled for upcoming separate testing in the elementary grades, and so it was taught separately and social studies was being squeezed out. At best, the social studies textbook was used sporadically for content-area reading practice. Teachers in South Carolina schools described a loss of integrative and creative curriculum units such as world geography and weather. Because elementary school teachers had such limited content knowledge of science, one HPS began rotating students to a different teacher who would specialize in the solar system, or in rocks and volcanoes. In the low performing district, students moved to different teachers — almost like in high school — for math, reading, and science beginning in the third grade.

North Carolina teachers described social studies and science as lessons that were “a break.” They noted that the district had developed pacing guides only for tested areas (math and reading, somewhat in writing). The North Carolina Standard Course of Study was not similarly organized for science or social studies or health. The principal at one school said that it was her obligation to help teachers “eliminate what might be enjoyable but is not the best use of time.” This school also alternated a single time slot for social studies and science, and the teachers reported this as evidence that such studies were marginalized in importance. They heard the message from the administration as being that they “can’t worry about that until everyone is on grade level” in reading and math.

Teachers in LPS felt more pressure to adopt particular methods of teaching as preparation for doing well on the tests. This included helping students to internalize strategies for identifying main ideas in reading passages and recognizing the math work implicit in a story problem. In a North Carolina LPD, teachers described every classroom as having the same activities for literacy – word attack, vocabulary building, etc. The teachers said teaching these strategies was required “to appease the gods,” but criticized the work for not addressing “broad concepts” that students should be able to understand and use. In a South Carolina LPD, the principal was very clear: Integrated curriculum units and “centers” for instruction were fine, but



that there had to be a time “when they sit down and work.” For her, this seated, concentrated work was the best evidence of a good elementary school. With high teacher turnover, this school was in the process of adopting the highly prescriptive Saxon math and phonics in order to have more tools for improving fundamental reading and math abilities. In a Georgia LPD, teachers said that test results drove decisions for instruction – skills, procedures, and instructional time allotment. But they indicated that these were issues of classroom management rather than deep curricular considerations.

One 25-year teacher veteran from Tennessee said she was expected to teach more and more facts as the state expands the curriculum, but that no one ever removes any items from the curricular frameworks. She reported feeling like she was on a “treadmill, pushing students harder and harder.” Her school and others have begun to eliminate recess (even for young elementary age children), cut back on art and music, and cancel field trips all in the name of spending more time on preparing students for the tests. Teachers in the LPS in Georgia, which was not officially labeled as low-performing, reported improved information sharing, but stress coming from implementing so many new things at once time. The North Carolina teachers in a LPS echoed this by describing a bewildering array of new programs and initiatives.

In Mississippi, teachers were concerned about raised expectations that “pushed down” requirements to younger grades as part of desire for higher scores. One teacher from a HPS said:

I wonder, do we push more and more down on them? Are we expecting too much? I wonder sometimes, does the school system need to slow down? Let’s just sit down and read a book, do some math, instead of trying to cram all this other stuff in. They [the students] get to where they want to give up, ‘I can’t do anymore.’ You know, that’s not the way we used to teach second grade.

Although teachers did not name administrators as being specifically at fault for these pressures and changes, the cases reveal the new and problematic working conditions that teachers encountered as a result of the high-stakes accountability system in their state. For example, a number of teachers in Georgia described the many things they had to implement, and said that it was almost impossible to do it all or do it well. One teacher from a LPD noted that, “It is only going to help if we can figure out what will be taken away.” Teachers also reported that the stress and strain caused additional possible harm to working relationships within the school because there was “so much to meet about all the time.” So many good things had to be done and demonstrated that there was “not time to have a good time with the children.” The teachers were not opposed to the new mandates and initiatives, but they believed that earlier programs to be taken “off their plates.” Teachers in a North Carolina LPD reported that they were seeing some great things happening for the kids, but they wondered about their ability to “stick with teaching, even making it to Thanksgiving because of exhaustion.”

In the schools at the very lowest levels of achievement, there was also stress and strain arising from the work required for increasing student achievement. But it was difficult to untangle what of the reported stress arose from the overwhelming amount of work that needed to be done to improve student achievement, and what pressure came from having to carry out those efforts under increased scrutiny and their own lack of knowledge of how to teach differently.

### **Professional Development Improves, Albeit Unevenly**

Our interviews with teachers revealed their general satisfaction with the professional development they have begun to receive of late. From our high, average, and low performing schools we found examples of teachers reporting that their learning opportunities have improved. In Kentucky, teachers said before KERA professional development “used to be a joke,” but now they believed that standards and accountability have “kept us on our toes” and there seemed to be more positive professional development opportunities as well as attitudes among the younger teachers.

We consistently heard from teachers that their voluntary professional development opportunities were better than the mandatory activities. In addition, teachers from higher performing schools had far more autonomy in choosing their learning activities. For example, in a Mississippi HPS, teachers reported they

enjoyed new-found professional development autonomy: their weekly lesson plans were only checked for teaching state frameworks. As a result, teachers summarily skimmed over the professional development offered by the district and focused on activities they initiated or participated in on their own as their most important learning opportunities. For example, we heard from teachers that the process of becoming “Board Certified” with the National Board for Professional Teaching Standards was a powerful professional development experience.

In a South Carolina HPS, teachers were merely required to create individual PD plans — which were only generally monitored. In a Kentucky HPS, teachers spoke about the opportunities to review assessment results and determine steps to take next with their students. Teachers said they were given considerable flexibility and support for professional development through an active PTA and town council support for initiatives for improved student learning. However, they recognized how little time they had for such efforts.

Still, teachers in the HPS were more likely to report that they had a broader range of opportunities to learn, and they were sometimes quick to pass judgment on whether different offerings had been appropriate or worth their time. We consistently heard from teachers that they still had to attend large group professional development activities that were intended to “motivate” them to perform at higher levels, but not necessarily give them the skills to do so. These learning opportunities were not what many of these teachers wanted or needed. Instead, they were seeking more opportunities to problem solve how to teach certain concepts or analyze the work of specific students in order to build a more adaptive curriculum. We found little evidence of this type of professional development — even in the higher performing schools we studied.

On the other hand, we found virtually non-existent or grossly inadequate professional development in several of our lower performing schools from under-resourced, rural areas. For example, in one LPD in South Carolina, professional development often included prescriptive formula for teaching to the test, but also “games” and “fun activities” to encourage teacher to be more comfortable in working together. Here we found teachers spending a considerable amount of their professional development time engaged in learning how to use technology, including how to use PowerPoint presentations without a clear connection to improved instruction. This type of professional development also counted toward teachers accumulating graduate credit for advanced degrees.

In Mississippi, one LP district designs the professional development calendar for each of the schools, based on a needs-assessment survey of teachers. During the year of our visit, the professional development calendar was filled with eleven 3-hour modules, including sessions on classroom management, test analysis, technology (including PowerPoint), CPR, crisis management (for new teachers), internet training, general curriculum development, instructional management, and something called “Grab Bags and Home Style Learning.” During the year, teachers had short sessions on working with bilingual learners as well as creating “effective” bulletin boards.

Professional development resources in the school were quite dispersed — although teachers there were not without access to promising modes of support. For example, teachers there now spend 90 minutes each week in peer coaching study team meetings. During these meetings, grade groups of teachers work together to develop targeted plans for interventions with individual children, based on their reading scores. One 2<sup>nd</sup> grade teacher reported that this approach was just what they needed:

Now there is provided time to work with (each other)...to work together as a group. We’re able to discuss different issues in relation to the students and to discuss what we need to do as a team.

In one Kentucky LPS, we learned of teachers pursuing professional development to improve their computer literacy skills. We also heard from a teacher seeking new ideas and materials to teach science, since materials she had been using in her classes for years were “old and stale.” One teacher specifically mentioned how much she benefited from sharing ideas with other teachers in the building, and from serving as a supervising teacher for a first-year teacher intern—neither of which, it should be noted, can count as formal professional development for this individual.

Even in Kentucky, with a more mature accountability system and deeper investments in professional development, teachers from rural, LPS rarely saw their learning as a linchpin for increasing student achievement. Here, we found evidence that the school had made strong academic progress with a small group of rural, lower income children, and teachers saw professional development as potentially useful. Here, teachers discussed professional development with us in fairly individualistic terms. They work with the principal each year to develop and refine a “Professional Growth Plan,” and to find staff development opportunities that match their areas of need, their interests, and their schedules. Granted, their professional development planning takes place within the context of the school’s consolidated plan, which must be based in large part on Kentucky assessment results. Unfortunately, the small number of students at this LPS means that the school has only a modest professional development budget, about \$1,500 in most years. This is enough money to send teachers individually or in small groups to nearby professional development sessions, but not enough to offer more intensive, more embedded professional development known to improve student achievement. All in all, we found a very decentralized approach to professional development that relies heavily on teachers’ own needs assessments and conscientiousness. This amount of flexibility does seem to be appreciated by teachers, who at least feel that they are not roped into pursuing professional development they consider irrelevant to them.

In North Carolina, we found a low performing district that was making considerable headway in improving teacher learning — in part due the leadership committed to and resources allocated for professional development. In the APS we studied, we discovered more intensive approaches to literacy and math (with the latter focused on the more ambitious curriculum). For example, several teachers in the APS mentioned the principal as an important force in their learning. One teacher noted the principal’s respectful, long-term vision for improving teaching and learning.

She does not jump into the “*have a problem? get a program*” approach, but rather makes a commitment to “a long term goal that keeps us focused” and to letting us know we are going to “go beyond a year” before they again start something new.

The school’s assistant principal (who recently hails from Los Angeles public schools) attested that professional development activities designed by the principal and the teachers are “exceptional” — with a growing focus on using data and working together on adapting curriculum. For example, teachers in the APS described three main parts of their math professional development program: intensive summer workshops, regular study groups by grade level, and regular in-class coaching and support by teachers released to perform this function full-time for 3 years.

In the district’s LPS we visited teachers who reported:

[The district’s mathematics initiative] and Investigations helps you to do a lot of hands-on activities, versus sticking to the textbook. I am a person who strictly sticks to the textbook if I have one, so this has given me a new way of doing things that I’d never thought about.

[The district’s mathematics initiative]—the whole math piece, period [has been the experience that has had the most impact]...because I wasn’t a math person, and now all of a sudden I have become a math person.

However, this district has had difficulty recruiting and retaining teachers, and has hired a large numbers of uncertified teachers, often referred to as, “lateral entry” teachers across the state. The need for teachers’ professional development is paramount, given that many newly hired teachers have little preparation for teaching. In this case, the district’s mathematics initiative has been a big help to some of these teachers. One teacher reported that her “coach” comes to her classroom twice each month, teaching lessons and giving her strategies and ideas to implement. For example, she brought a “hundreds board,” and showed how to add and subtract using the hundreds board, which the teacher noted, “was new to me.”

Indeed, the district was said to offer a “wealth of resources” for teacher learning, as well. Teachers reported feeling “exposed to everything,” and said that the district “has really put a lot of work into

educating its teachers.” However, we found a teacher turnover rate of more than 25 percent that, in some cases, limited the impact of the more comprehensive and effective professional development that was unfolding in the district. In addition, the district now needs to educate growing numbers of second language learners for whom their teachers are vastly under prepared to teach.

### **The Press for Better Professional Development: Greater Teacher Collaboration**

We found only a few examples of effective professional development as defined in the literature. For example, across the six states, we found only a few instances of teacher study groups, and rarely would teachers have a chance to observe and critique another teacher teaching a lesson. Even when teachers were favorable toward their professional development (and many were, compared to previous satisfaction levels), criticisms of follow-up and support were reported. As one teacher from a HPS noted:

They got you excited and gave you great materials – but the only thing that was difficult was that they had wonderful materials, but there is no lab here to really use the ideas. It is difficult to do experiments without a sink and with carpet on the floor, you need to be able to make a mess to learn science, and they really do not want us making messes in our room. They had wonderful ideas, but (the proved) hard to implement.

Occasionally, teachers talked about collaboration as a key component of professional development, which may be one of the most important effects that the accountability systems have had on teacher learning. For example, one South Carolina teacher from a HPS tried to describe the “best professional development” she had experienced:

My best professional development? This is so hard to (identify). The district has offered us a lot of workshops on how to use manipulatives, and that has been good. Also, we meet on our grade level once a week where we have a chance to plan together as a group and share ideas. We also meet once a month in different committees. I am on the “high expectations committee” and we talk about research. Then we have cross grade level meetings that meet once a month.

However, we were most likely to hear reports of increased collaboration in the HPSs that we visited. Principals in these schools described working for a collegial approach to decision-making and changing professional development. In HPSs, we also documented increased likelihood of teachers learning together to introduce math concepts in new ways, working with different levels of readers, and utilizing technology in non-regimented ways. Grade level meetings in the average schools were described as discussing standards, benchmark testing, and scores, and principals talked about getting their teachers to see where “great things are happening” by having teachers observe their peers in the school.

Still, in some places, collaboration was lacking. In one Mississippi HPS, there were no formal means for teachers to learn from each — although some did so on their own (for example, when teachers decided to work together on their National Board Certification or discuss informally their test score results). In a South Carolina LPS, we learned that teachers wanted more help from other teachers in teaching hard-to-reach students. One teacher noted that she “needed someone to work one-on-one with struggling students while others move on.” It seemed that the assistance that she was receiving, from an outside consultant, just was not helpful enough to make a difference in her teaching and student learning. In a Tennessee LPS, teachers reported that they had no opportunities to experiment or refine teaching methods. In this school, while the principal frequently observes and evaluates teachers, the process is not sync with the accountability system or the teachers’ own perceived needs.

One teacher, when asked about her evaluations, wasn’t sure what the principal was looking for when he observed her. She stated:

I’m not exactly sure what she’s doing, but she’s checking off something. I have been observed twice in the past two weeks, and I have not gotten any feedback about these observations. I’m hoping I will get some.

In the APSs and LPSs, if we found collaboration, we were more likely to find it with a specific focus on test score improvement. We found numerous examples of teacher teams working together in “planning weekly lessons, unit tests, and homework together so that (their teaching) is pretty consistent.” However, collaborative time in the lower performing schools focused more on standardizing certain practices and gaining much needed consistency, as opposed to building a collaborative teaching culture whereby teachers critiqued each other’s lessons.

Nevertheless, all teachers felt stretched to find time to collaborate; but in the high performing districts, teachers were more likely to find ways to do so informally, and in doing so, emphasize teaching students more than to the test. In the lower performing, more impoverished school districts, the blatant lack of substitute teachers seemed to inhibit administrators from freeing up teachers to attend workshops or visit other classrooms. In any case, it was common for teachers to have an hour planning time each day while the students go to physical education, art, music, and/or library. In one Mississippi school, one day a week is designated as a 60 percent day to allow the teachers additional time for planning activities. Beyond this example, we found few instances of teachers and administrators using time differently so teachers can learn from each other.

### **The Complexities of Teacher Learning & Autonomy**

We found that not all professional development assistance is welcome, nor is collaboration an automatic result of high-stakes accountability. For example, one district curriculum coordinator in South Carolina, assigned to a low performing school, found she could not make an impact through direct coaching in the classroom. Instead, she needed to work through other teachers, who were “insiders” and thus were more likely to be trusted. In several of the LPSs we studied, teachers felt very threatened, and were seemingly less responsive to outside intervention.

A state assistance team assigned to a North Carolina school was “devastating,” because it was described by the teachers as promoting a one size fits all “fix it” kit. The LPSs experienced a high level of tension, with teachers needing to be reassured daily that they weren’t about to be fired. In one LPS in North Carolina, teachers complained of a constant “snoopervisor” who, according to them, was visiting their classrooms not to help, but to uncover evidence to be used against them. In a LPS in Tennessee, the principal reported that his school had “more problems than Van Camps has pork-n-beans” and that he has needed to be diligent in setting up guidelines for his faculty because of all the (teaching) “messes he has needed to clean up.” These local conditions may have nothing to do with the kind of state accountability system in place, but they do affect how teachers learn.

In any case, state accountability systems imply a more standardized curriculum, at least in terms of all teachers teaching to achieve, and all students expecting to reach, the same standards. In LPSs, where the press for higher performance is understandably greater, teachers seemed to expect less autonomy than their counterparts in HPSs.

In one LPS in South Carolina, teachers are required to teach Saxon math – where they are expected to teach “according to the book,” and where they “review and review” so there is no chance that students will “have an opportunity to forget what will be tested.” In a Mississippi LPS, administrators maintain a strong control over what is taught, requiring all teachers to follow a common scope and sequence of the curriculum, maintaining checklists for compliance with this district mandate. We actually found very few teachers in these schools who resisted these more prescriptive teaching and learning practices.

On the other hand, one teacher in a South Carolina HPS noted:

I try not to look at the test. I pay attention to the standards. I think some teachers pay more attention to just what is on the test. I think all learning is important. I do not think that all teachers know how to teach differently.

Teachers respond to their respective accountability systems in different ways, especially if they are highly confident of their practice – whether or not that confidence is justified by student results. We also found

that higher performing schools were staffed by teachers who tended to have more experience and more advanced degrees, and believed that they had the skills and “know-how” to teach to more than the test. In the more impoverished, rural, low performing schools, the opposite was likely to be found. Indeed, in a LPS in Tennessee, the principal lamented the large numbers of teacher who were on “interim certificates.” He noted that, “some of these teachers have been on the faculty for 7 years with interim certificates because they haven’t passed the NTE.” While he claimed that he could not allow these teachers to continue teaching, he also recognized the fact that there were few teachers to replace them. With this in mind, he felt compelled to tell his teachers what to teach.

In most of the HPSs we visited, teachers talked about the inherent tensions reported between curriculum alignment, teaching to the test, and teacher autonomy. Because of their students’ success to date, these teachers appeared not to be overly concerned about potential negative reports on their teaching, and so teachers in these schools were more confident that they could “make it work,” that is, teach in much the same way as they had been. It was reasonable for them to expect the students to be able to adapt what they had learned to whatever the tests would ask.

At the same time, the teachers we interviewed in the higher performing schools tended to be pleased with the curriculum programs in place. In part, teachers there had access to more materials and seemingly greater quality and diversity. In one North Carolina HPD, teachers were enthusiastic about two school-wide approaches, Write From The Beginning (WFB) in preparation for the 4<sup>th</sup> grade writing test, and Accelerated Reader, a curriculum that provided additional assessments for grade level reading achievement. These programs offered the kind of curriculum these teachers were prepared to teach, and tended to fit the pedagogical approaches they preferred.

Overall, in lower performing schools, we were less likely to find highly confident teachers. Also, in the low performing districts, we heard far less concern about teacher autonomy, and more about missing resources and resignation that someone was going to tell them what to teach. In some cases, teachers in LPSs reported few ideas of how they would or could teach differently. There, the expectation was that teaching to the test, *which was how teachers characterized more prescribed and formulaic teaching*, could be a good thing, as long as they had the tools needed to do the job. However, a number of teachers just did not have all the tools they needed to make the kind of difference for student achievement.

### **A Need for More Tools: Using Data for Teacher Learning**

We found ample evidence of the desire and motivation for teachers and administrators to make a difference in the lives of students. Although a substantial proportion of teachers we interviewed revealed doubts that had about all students being able to reach high academic standards, we also learned that many, especially in the LPSs, did not appear to have the tools needed to make a difference. One needed tool that teachers were definitely lacking was the effective use of data to influence both teacher and student learning.

However, we found that across the board teachers and administrators are eager for what they perceive as missing tools – especially for diagnosis of student learning needs. Although they did not always use this terminology, administrators and teachers across the different performance levels judged professional development opportunities *and other resources, programs, materials, and score-raising initiatives* by the extent to which they were provided methods for making clear predictions of student performance and, most importantly, for providing targeted interventions for improving those performances

It was not that teachers and administrators were not using test data, in fact, they were. For example, in one Mississippi LPS, all of the teachers interviewed spoke of test analysis they conducted, required by the district. All teachers are “required to analyze the test scores of all of their students” and “write a plan for remediation.” This plan is then given to the principal, where it is presumably checked and given to the student’s current teacher (this is because test scores are generally not available until the summer or fall after the child has completed the grade in which he/she is tested). However, teachers are not systematically prepared to unpack the test data (other than, for example, noting that a student scored poorly on “identifying a main idea”) and subsequently write a remediation strategy. Also, we found no systematic

follow-up of whether or not the proposed remediation activities were actually used by the student's next teacher, and if so, whether or not the proposed strategies were making a difference.

In North Carolina, we found teachers using a more sophisticated "K-5 portfolio assessment" that outlines the kinds of student work samples and profiles required to track students through elementary school, and presumably becomes important for gateways. However, administrators reported that it "was not clear what the teachers were really doing with this information. As one district official noted, "We just kind of talk about it, about their emphasis for the year – then the school improvement teams talk about their focus, etc. that they want to take."

In Mississippi's HPD, teachers reported that they were often discussing the importance of scores and the need to raise them, but that they were not carrying out test score analysis. For this high scoring school, it may have been enough to be aware of the test itself. As one teacher stated:

Our staff development has mostly been about working and being sure that we're teaching benchmarks in our schedule and being aware of testing.

As part of their test analysis strategy, the teachers in a South Carolina HPS looked at trends from aggregate PACT scores (by grade level) and some at benchmark testing provided by the district. But, teachers were unsure of how to make the match between the two and with the norm-referenced test (Terra Nova) used by the district. Teachers in South Carolina recognized that the PACT tests provided no diagnostic information.

In Kentucky, where data analysis was most supported in schools, the most critical missing resource was time. In Kentucky's HPD, new mechanisms were put in place for using data to improve professional development, yet teachers had only four professional development days a year to work on such matters. If one of these day was dedicated for district training (this is true across the state), and a second day involved faculty working together to analyze school test scores, that left only two days of teacher-selected professional development to build upon the analysis of the test scores. It is inconceivable that such a small amount of time could have a major impact on one's teaching work.

In a HPS in North Carolina, teachers accepted that there would be many tests in the course of the school year, but these tests were not always linked to instructional decisions, because the curriculum programs (Write From the Beginning, Accelerated Reader) were already in place. Instead, students were being tested in order to ascertain their current reading levels and to predict the outcomes likely on the final (end of grade) tests. There was no clear connection between test results, teacher learning, and new instruction. In this sense, test analysis was not diagnostic for teacher learning at all. Teachers said they wanted knowledge about what to do for kids to help them succeed on tests, and they did not see professional development (insofar as they had experienced it) as the solution.

Results on tests played a much different role in the HPS in Tennessee, which posted scores like trophies on the school walls near the main office. The district did some score analysis, but there was no sense of urgency to understand the numbers in terms of individual student performance in a given set of classrooms. The expectation from the principal was that teachers would adjust teaching according to testing data (as if this were the obvious and easy response to test results) as necessary.

Teachers in the APS in the same district also lacked test data analysis tools. Because these teachers felt like they were "under the gun," they were more likely to be concerned with documentation of student progress than with interventions, and rather than test analysis they were more likely to receive exhortations to do better from their administrators. In Georgia (also in an APS), teachers talked about an extensive burden in "documenting," which seemed to encompass charting student progress and recording special services received. The teachers said they had good district data management and were able to look up student scores on different tests and "structure more appropriate lessons based upon strengths and weaknesses for individual students." There was a clear expectation from the system that teachers would provide appropriate lessons, but there was no mechanism for doing this except individually, if a teacher had the capacity to find the resources, the time, and the appropriate materials for supporting student growth.

On the other hand, responses to using student data in the lower performing districts and schools were even more problematic. For example, in a Tennessee LPS, tests scores were reported as a driving force in curriculum and instruction, with teachers working to determine how to ask questions in their lessons and assignments in the exact same way as found in the state (TCAP) test. At the same time, the teachers described being led through quasi-test analysis, without gaining deep knowledge of implications of scores, or making plans for ongoing use of data as a tool for decisions. In fact, teachers in the Tennessee LPS saw test score analysis as a source of stress. Teachers knew they were monitored and that test scores needed to go up, but the question was “how.” Where an instructional response to test scores was unfocused, there was no capacity for the accountability system to make a useful difference.

The teachers in a Mississippi LPD were supposed to document progress and show analysis of tests scores as information that is turned into the principal (who monitored compliance). Teachers did not describe a connection between this documentation and changes in instruction or work with other teachers or administrators that would use data to guide efforts to improve achievement.

In a Kentucky LPS, a principal reported looking at student scores and doing considerable analyses. The principal said that he met with teachers in the fall to review scores and discuss “curriculum implications.” He described the talks as demonstrating that results mattered, but he could not name specific actions take by teachers in response to test results.

Teachers in the Georgia LPS were forming committees to study the state tests, but reported no instructional impact from their work together. At this point, the teachers believed that test score analysis was less important than understanding the relationship between the tests and the curriculum.

On the one hand, the rather well-resourced North Carolina LPS (with the reform minded superintendent) offered some encouraging examples of tests scores serving as the basis for understanding student needs. One administrator described teacher meetings where, student-by-student, name-by-name, the teachers reviewed rosters and talked about the services each child was getting to meet the needs determined by scores. Test results were discussed weekly at grade level meetings. Teachers were supposed to “utilize everything available to them.” The principal, who was well supported by the district office and had numerous learning opportunities herself, said teachers were motivated because they have had a “taste of scores going up.”

As each of these examples suggests, high-stakes accountability systems would be well served by teachers prepared and aided in more comprehensive data analysis. In the course of district-level interviews, we found some (but not all) administrators to have a deep understanding of the testing system. In some instances, district officials thus had the knowledge of what the assessments required, knew how these aligned with curriculum, recognized what individual student scores could or could not reveal, and could likely evaluate teaching materials and organizational decisions in light of this expertise. Indeed, we found administrators who prepared elaborate reports, disaggregating scores and marking trends. But there was little indication, outside of a few instances, that these materials moved out of the hands of administrators and into classrooms, or that there was clear thinking about how such information might inform teacher practice and school decision-making. There was little evidence that score analysis would be used to help administrators or teachers understand what professional learning opportunities could contribute to the improvement of schools. All in all, we found that this critical lack of teacher preparation and support in using test and other student data inhibited school improvement and drove a wedge between accountability and professional development. This factor was one of several loose connections in the system.

### **Loose Connections: What Accountability Means in Different Schools**

The 150-plus teachers we interviewed in all six states — irrespective of whether or not their state’s accountability system included rewards and sanctions — were responsive to the state standards and assessments. Indeed, from our interviews, we learned that most of teachers were committed to standards-based reform and the use of standardized assessments to improve student achievement.



However, the message of high stakes accountability is often perceived as political, *with the intent of* exhorting educators to work harder, demanding improved performance, and decrying putative wasted resources. In this study, high-stakes accountability systems were not seen as providing a vision of, or a clear set of directions for, improvements in the delivery of education. While schools were aware of the possibilities for bad press and more substantial penalties for extremely low performance, teachers tended to understand the “teeth” of accountability as coming from increased demands for compliance through documentation. Even though we interviewed teachers from six states that had very different accountability components (e.g., rewards and sanctions) in place, we found virtually no variation from state to state. In large part, the *loose connections* between accountability and teacher responses had a great deal to do with beliefs about what some teachers could accomplish with at-risk students, and what the high-stakes tests actually measured and reported.

Despite their commitment to accountability as a concept, teachers in all schools were skeptical of the fairness of the state tests (an issue we address later). No teacher expressed confidence that a school would be adequately understood by “outsiders” as a result of the state accountability system, and none believed that the intention of the state mandates was to celebrate the hard work teachers were contributing to their schools. (Indeed, a good example of this phenomena was revealed in a recent newspaper article in North Carolina whereby education officials expressed concerns about state testing standards being too low in light of increasingly higher passing rates among schools.<sup>3</sup>)

For example, in a Mississippi HPS, a number of teachers discussed how their role in tests score results — high or low — was limited by factors out of their control. One teacher noted:

Accountability puts a lot of responsibility on you because you can do a good job, but it all depends on the kids that you have. Sometimes you have better students than you have at other times... I realize there has to be some kind of accountability, but a lot depends on the students.

Another Mississippi teacher from a HPS noted that her students always did well on tests and, therefore, she doesn't really think about accountability all that much. She added:

If you're teaching in your classroom, it always works out. They're going to do well anyway.

Educators from these schools felt compassion for their counterparts in lower performing schools. And, many administrators felt the same way. In one HPS in Tennessee, a principal discussed how he felt that his school's successful test scores were due to the relatively high socio-economic level of its students. He believed that Title I schools in his district should not be held accountable for the test scores in the same way as his students. He asserted:

Many (of those in the Title I schools) are doing an excellent job and they're teaching to the level of their class. ... They're doing exactly what they've been asked to do — take the child from where he is and move him forward. But their schools will never be where some schools are and that's sad. Then they get marketed as a poor school.

In a lower performing school district in North Carolina, teachers and administrators felt overwhelmed by the challenges their students bring from their homes and family backgrounds — including drug use in the neighborhood, developmental learning delays, a dire lack of parental involvement, and growing numbers of children whose primary language is not English. Indeed, few teachers, if any, reported having specific teaching skills in working with second language learners.

Some teachers asked themselves what they were doing wrong, but most have few (or know of few) resources to address problems that result from “parents who neglect their kids” or “kids who act out because they are angry.” One teacher spoke sadly about a class where “not one child in a family has a bedroom — they sleep on a couch or in the hall.” She claimed:

---

<sup>3</sup> See T.Silberman (2003). Success Breeds Concerns for ABCs. News and Observer. October 13. 1b, 4b.

I don't care (so much) about the state tests – because I know that I am doing the best that I can do. And the people who make the rules don't always understand – what I consider a success, they don't always consider a success. So if I've got kids who are moving up 4 or 5 points, instead of 8 points, and that's good, because I know where they were at the beginning.

Instead, teachers saw that efforts needed to be publicized, and that it was critical for their communities to see their school as a place that was doing the best for its children.

In the LPSs involved in our study, we were more likely to find teachers under an increased demand for paperwork trails and documentation of remediation efforts for less successful students; and this focus on documentation seemed to loosen the connections between accountability and teacher learning. We heard numerous examples of teachers being under siege by accountability systems and embattled by what they called “overwhelming demands.” Teachers described invasive scrutiny of their classrooms, increased intra-district competition, and specific pressures on teachers to make changes whether or not there were professional reasons to do so.

Teachers in these schools described their increased workload related to the accountability system. One Georgia teacher said, “You have to document how you run your classroom and what you are doing for enrichment and remediation. The majority of teachers are documenting more in order to know where their kids are scoring on a regular basis.” In a Tennessee LPS, the principal requires monthly reports from teachers showing compliance with each item in the school improvement plan. They may submit a captioned photograph, a copy of a lesson plan, or an actual product created in the classroom. These artifacts are filed monthly and routinely checked to ensure, according to the principal, that all teachers are trying to “work with this system and stay involved.” These compliance issues dominated our interviews with teachers.

Given what we saw and learned, schools did need better ways to capture what was being taught and to what extent new teaching practices were successful. However, what we heard from teachers was that accountability was more of a burden and that the documentation required was not helping them become more effective as teachers.

In a South Carolina LPS, teachers raised concerns about the impending state report cards, and anticipated a need for further proof of their efforts to succeed with students. Other teachers were beginning to worry about professional development — would it count for re-certification or count as an unexcused absence. They talked little about the relationship between what their students knew or did not know and what they needed to know. More importantly, many teachers were feeling hemmed in by the accountability system, and one teacher in this school asserted:

It's almost a luxury to think about children as readers and writers and mathematicians and scientists and artists and musicians, because the accountability on that one measured tool is made so public that until you deal with that and allay fears of the parents and the community about how healthy your school is, you can't really pay attention to all the pieces of the child.

In the lower performing schools and districts, teachers and administrators were apt to describe a *further* competition between the school and “the world,” because of the belief that most of the public has negative views of schools and educators. Principals and teachers felt they were under attack – that the new accountability system seemed geared to put them on the defensive. We heard many times that “the public's confidence in teachers is not what it was twenty years ago.” We also heard many comments about the frustrations of not knowing how to best communicate what students know and what teachers and administrators are able to do.

Several low performing schools and districts we visited face enormous economic and social challenges. For example, several educators in a Tennessee LPD reported they were facing extreme economic and social challenges that included the lingering effects of slavery, the continuation of tenant farming and extreme poverty, and the community's power structure that did not value education for its poor black children. For teachers and administrators facing these challenges, the state's accountability system may not seem as

important as it does in school communities where their counterparts believe they have a chance in succeeding. Many teachers in these communities did not believe they could succeed with the students they were teaching, so they identified other accomplishments or values for education beyond what is recognized by tests scores. These teachers and administrators had little understanding of effective professional development practices or where they existed. If they did know of these practices it would be difficult for them to believe that using them would improve student learning — or raise student test scores.

In fact, in every LPS we visited there was little optimism. Indeed, some teachers felt high levels of responsibility for student learning while others felt only that they were blamed for perhaps inevitable poor performance. While some teachers and administrators reported that they were given additional help and resources, most of what they said centered on “punishment for low scores” and which teachers were “bringing down the district.” One South Carolina teacher warily described a “sketchy” process her superintendent uses to compare individual teachers across grade levels where “you don’t want to be that bottom teacher” who then gets assigned a “Teacher Support Team.” Teachers there did not fully understand the relationship between what they taught, what students scored, and how the superintendent was assessing them.

Given their students’ past achievement levels, South Carolina teachers did not believe they would earn bonuses or rewards, but “there might be praise here and there if (they) were able to raise scores.” Another teacher said they just needed to do the very best that they could do, because “we’re not in control of what’s printed on the report card.” A 4<sup>th</sup> grade math teacher reported concern that her work was reduced “not to the lessons you taught, but the scores.” A teacher in a LPS located in a very impoverished community described a need to “run for cover” because the scrutiny of the state would expose teachers themselves as poor and uneducated.

Where school performances were lower, the educators were more likely to acquiesce that there was only so much they could do for “some” students. “How much can we **do** for this population?” was the question asked by a South Carolina teacher. At her school, teachers felt good about their efforts in their individual classrooms, but were concerned that despite their hard work, they don’t know that they can do enough for their disadvantaged students. These educators did not expect that the accountability system would provide a mechanism for improving the chances for their students’ success.

To be sure, we found some counter examples in an urban, LPD with more resources, and led by a reform-minded, activist superintendent. There, we found teachers engaged in more powerful professional development and more effective teaching practices — and as a consequence there was a greater tendency for teachers to believe that all students could meet the higher academic standards. But we also found high turnover rates and high proportions of inexperienced and not fully licensed teachers in these schools (although not as high as in other sites where experienced, certified teachers were very difficult to recruit and retain). While these factors seemed to limit the capacity of the schools in these districts to respond to the accountability system, we also believe that the resources provided served as a catalyst for teacher learning and much less of the defeatist attitudes we found elsewhere.

### **Further Disconnections: Teachers and Administrators Baffled By Results**

With student performance seen as inevitable by some, we found many teachers complacent. Thus, there was only so much they could do, and thereby loosening the connections between high stakes accountability and teacher responses to the system of rewards and sanctions. However, we also found teachers “baffled by results” generated by their respective state accountability systems. This finding seemed to hold irrespective of the complexity of the state system for calculating and reporting scores.

We found much confusion among educators in the schools and districts as to how these ratings are figured, with many claiming that even knowing a school’s raw scores would not allow them to predict their school’s status. One rural superintendent in North Carolina went so far as to call the state growth formula “voodoo statistics.” One teacher, in a high performing North Carolina school of a high performing district, believed that the state was planning to save money on bonuses because so many schools were improving: as a result, *she reasoned*, fewer would be able to show growth, because all were growing. *No teacher or administrator*

*reported that the accountability system was easy to understand or straightforward in its measurements of one school system against another. None of these educators offered to explain how the state was deciding how many schools would or should be low-performing, or how a school could be sure to attain an excellent or high-performing rating.*

Teachers in North Carolina, which has a rather mature accountability system, expressed frustration with a number of rule changes within the rating system. One *average performing school* in a high performing district had been recently hit by “re-leveling” of state expectations: as a result, teachers realized they could not predict results *from one year to the next*, and so they were not expecting any rewards for their hard work. The teachers had believed (based on the raw scores) that they had qualified for exemplary (bonus pay) status, and instead they received “no recognition” which was, as far as they were concerned, a “negative label” for their students’ performance. One teacher noted, “You push so hard – you know the kids did a good job – and then the state changes the standards after the fact.” The principal in this school reported a “tremendous impact (negative) on the faculty following the re-leveling of scores.” In the state’s low-performing district, teachers believed that the North Carolina tests had no logic or predictability – especially because they believed there to be no connection between teacher effort and student skills. As another North Carolina teacher in this LPD reported, “The year that the scores were really bad, we had all killed ourselves.” Suspicions about the quality of the tests were only made worse, according to teachers, when the state acknowledged a math test score screw up in 2001 when 98% of the state’s 5<sup>th</sup> graders passed the test. Another teacher remarked on that situation as being ridiculous and yet frustrating at a time when the system “has a gun to our heads.”

There was some general confusion about which schools qualified for rewards and sanctions according to state policy, especially in states where the accountability systems are only now developing, such as Georgia and South Carolina. In those states especially, teachers generally feared for the worst, such as reporting that principals and teachers in low-scoring schools might be fired — an expectation not shared by their district officials. At the same time, educators in the states with longer-standing systems, such as Kentucky, were more sanguine about the process and somewhat more articulate on the subject of positive and negative consequences of testing and how they are assigned scores and ratings.

As most psychometricians will claim, small schools can experience extreme annual variance in test scores due to statistical probabilities, and thus policymakers need to be careful using results for accountability and other purposes. Depending on a cohort size in a particular grade level, and student mobility, test score reliability can fluctuate (Linn, et.al., 2002). In Kentucky where we studied a number of small schools (about 200 students), teachers and administrators expressed frustration over their testing system — which to them “compared apples to oranges” by examining the score of one group of students one year, with another cohort’s scores in another. Although teachers and administrators here were more satisfied about the validity of the tests their students took, they did not fully trust the test score results and were eager for their state to implement a system that was more focused on “value-added” learning (in contrast to NCLB which is promoting cohort analyses). We found this to be the case in both high and low performing schools in the state.

Additionally, teachers and administrators expressed fears of the power of a single score, and related concerns about the quality and purposes of the state assessments. We heard deep concerns about the limitations of using just one test on one given day as a sound measure of student learning. Teachers offered a persistent criticism of the state assessments, the familiar problem of students who are having a “bad day” on the day they are tested. Across the spectrum of achievement levels in all six states, teachers worried aloud about how much might be riding on students’ motivation, concentration, and preparation to take a particular test on a given day

Indeed, the most powerful complaint about the fairness of the accountability system was that the assessment results were indicators of the socioeconomic class of the students. In many cases, teachers and administrators were not looking to eschew accountability for their performance, but “questioned the fairness of the assessments given their “deep understanding” of their students knowledge and skills and the inadequacies of the tests to measure them. We heard countless examples of students facing dire life circumstances, which limited their ability to show what they knew on a given test day.

In addition, the tests themselves, and the state assistance toward success, were also critiqued by a number of teachers and administrators across the spectrum of schools. Teachers reported concerns about what was on the state tests, and the kinds of requirements students were expected to meet. Several teachers in Georgia reported that they did not know “what to expect or how to prepare students.” Similarly, in Mississippi teachers said that state testing “throws in” skills not in curriculum – for example, three digit problems to solve without regrouping. In South Carolina, where teachers generally were pleased with the content and format of the test, some educators had real concerns. As one teacher noted:

I am afraid that the test tries to trick the children...the way the questions are asked and the kinds of different answers from which they choose. I am into testing students — do not get me wrong — but when a child has to read something exactly one way to get an answer correct then something is wrong. These kids are so different from those who write the test questions. There are questions on the test that if you read just one word wrong you will not get the answer right – that is tricky and not a test of educational value. I am not the only one who feels this way.

Indeed, a testing official in a LPD in North Carolina told us plainly that the state’s End of Grade tests for grades three through eight were not valid for decisions about individual students (such as the promotion gateways they determine). He reported that the test was created in three forms that together make one validated test. The forms are not necessarily of the same difficulty. All teachers must hand out equal number of all three tests. Each form may not cover the same topics. There are only 80 items on math, so the standard error is high. According to the state, there are 240 items to draw from to form a valid exam.

It was never designed to be used as an individual measure. It was designed to be an accountability tool at the *school level*, for grade three....At the individual level, it was never designed to make a high-stakes decision. But that’s what they’re using it for, including AG identification, promotion/retention.

What is important to understand is that the experts on assessment in a district – unfortunately often far removed from the teachers in the classroom – were well aware of the information that state tests could not provide, and the frequent bad timing of useful results. Teachers were likely less aware of the technical complexities of the assessments, but they recognized that what they would be judge by were results that could be reflective of far more circumstances than student performance.

As we will discuss further below, most of the distinctions found between teacher responses to accountability and professional development had more to do with leadership, resources, and capacity available than with the signals sent by the accountability system itself. Perhaps there is so much farther to go in the low-performing schools, where one administrator said (in a non-labeled state) that “we may teach to the standards, but we may not assess the standards we think we are assessing.”

In sum, these disconnections between the accountability systems and teacher learning may help to explain why teachers question the high stakes testing in their respective states. What could be characterized as complacency or excuse-making on the part of teachers may very well reveal the missing links in these state accountability systems. Across the states, we found teachers and administrators knowing and caring more or less about the incentives or rewards of the accountability system, depending on whether or not their own schools were subject to them. Schools that received rewards due to test scores were proud, although they were also often not sure how they achieved their status and concerned about maintaining their reputations. Schools that were sanctioned, usually in the form of a low performing label from their states, had a motivation for improvement, although their plans varied as far as how to get to the next level. While we found some links *in these plans*, where teachers’ professional development *was expected to contribute to meeting* the demands of the accountability system, these links were tenuous.

### **Teacher Learning and Leadership**

Our findings demonstrate that it is important to understand variations in leadership in the face of disconnections between the intentions of policy makers and how high-stakes accountability systems are

experienced in schools. Central office administrators, principals, and assistant principals advocated, supported, or demanded various visions for equity, opportunity, and improvement. Leadership strategies for the deployment of resources are clearly influenced by, or shaped by accountability systems, and these responses are indicators of a system's organizational capacity. Leadership also helped to determine how teachers interpreted accountability requirements (standards and curriculum as indicated by assessments) in their classrooms. Whereas leadership varied, and different school climates may have arisen from different interpretations of what "getting serious" about student performance should mean, there were clear curricular consequences arising from the state accountability systems.

In the higher performing districts, we found, not surprisingly, much stronger superintendents, who they themselves were better educated, better connected, and more likely to find and secure professional development opportunities for their administrators and teachers. They also had considerably more resources at their disposal. These facts of life cannot be underestimated in trying to understand how school leadership plays out in the various teacher responses to accountability and professional development.

In HPSs, central office administrators and principals were more empathetic toward teachers regarding the accountability pressures they were facing. In turn, when the central office was not perceived as being helpful to teachers, principals seemed to be more likely to serve in roles as champions and brokers to help them get what they needed to help students achieve. In HPSs, we were likely to find more transformative leadership, while in lower performing schools, authoritarian leadership was more the norm than the exception. As the following illustrations from Tennessee demonstrate, such differences in leadership style have a significant impact on how teachers experience high-stakes accountability pressures.

### **Transformative Leadership in a HPD**

In the Tennessee high performing district, within moments of speaking with Dr. Pete Walker, the district director (Tennessee's version of a superintendent), it becomes clear that he intends to be successful not only in improving achievement in the district, but also in becoming one of the top ten districts in the state. Total Quality Management (TQM), strategic planning and continuous growth are the cornerstones of Walker's leadership style and belief system. As a director, Walker operates as an entrepreneur, with an extremely high degree of commitment and sense of ownership.

Dr. Walker envisions himself as not simply the director of the public schools, but as a key leader of the community, responsible for providing the quality of schools that will attract business and industry to the area. He is not simply accountable to the state for academic education, he is accountable to the community he serves. In his message from the director of schools in the Annual Report, Dr. Walker states:

In the business world, an annual report lets stockholders know the economic status of the company and draws comparisons to previous years. In education, an annual report gives our stakeholders (i.e., our parents, students, and community) statistics about students, staff, and other quantifiable data. In addition to a review of last year's performance, we want to communicate to you, our stakeholders, our system's direction in areas such as accountability, student achievement, communication, curriculum, staff development, technological advances, student safety, and facilities. (p. 2)

Walker claims that he has tried to stay ahead of the state in issues of accountability. For example, the Grover City School District developed its own curriculum before the state developed its curriculum. The superintendent works as an adjunct instructor at a branch of a state university. He claims that he taught for two years for free in exchange for the services of a team of professors who audited the newly developed curriculum in the district. The school district has proactively worked at developing community support for schools and has built a partnership between area businesses through the Grover Education Foundation. This foundation's mission statement, as published in the Annual Report, is "to support the Grover City School System by providing resources for enhancing education in our community and fostering the bond between education and the community." Among its accomplishments, the Grover Education Foundation claims to have secured sponsorship of thirty Teacher Mini-Grants from 27 different businesses, organizations, and individuals. The district superintendent actively promotes the schools in the district and solicits support and donations.

The football field and track for the new middle school were provided through a \$500,000 donation to the district by a private donor. An annual report benchmarks central office progress toward goals.

## **Transformative Leadership in a HPD**

A “no-nonsense atmosphere” describes Drummond, an APS in this high-performing district. At the time of her interview, Diane Hemingway had been on the job for two years. Amidst some local controversy, she replaced her predecessor, who was removed for low test scores in a Title I school. Hemingway approached her mandate to improve test scores strategically, knowing that she would encounter resistance among her faculty. The removal of the previous principal created a great deal of resentment among teachers – particularly those who had been with the school for a long time. As a matter of fact, some of the teachers worked concurrently with a district parent organization protesting the removal of the principal, demanding her reinstatement. Hemingway recognized the difficulty of her position and developed a long-term plan for instituting the necessary changes she would have to make.

Hemingway realized that her first task would be to lead her faculty in understanding the magnitude of the problems at Drummond Elementary School. She knew that, although the teachers knew the reason for the former principal’s removal – low test scores — they did not thoroughly understand the implications of their poor performance and what could be done about it.

In order to develop an understanding of the task ahead of them, Hemingway employed the assistance of a consultant. Together they selected a group of 12 key faculty members and analyzed test score data over several previous years. This group of teachers was able to see first hand the lack of improvement in achievement in the school. Three of these teachers subsequently conducted a faculty meeting in which they explained their findings based on the analysis of testing data. They also described the perils for the school that were associated with the new federal ESEA mandates and the failure to make annual improvement annually in math and reading. Once the faculty as a whole had been educated about the challenges they faced, then they were lead through the process of analyzing their individual test scores and those of their grade groups.

Through this process of data analysis, the teachers reached some conclusions about specific problems they needed to address. The decision about the first problem they needed to tackle arose from the fact that they had fallen into the trap of teaching to the lower end of their student population. Hemingway suggested to the teachers that by teaching to the middle and upper level students they could raise test scores more quickly. What was needed, according to Hemingway, was a “quick fix” while they got “about the business of really solving the problem.” Teachers were led through a review of Bloom’s taxonomy and higher order thinking skills that they incorporated into lesson plans.

This, Hemingway stressed, was the only change she made during her first year as principal, hoping to mitigate any difficulties relating to the circumstances surrounding her hiring as principal. At the conclusion of the first school year several teachers resigned or retired. Hemingway replaced these teachers with younger teachers, recently graduated from teacher education programs and in some expectations for her faculty and more concretely about the changes that would be implemented over the next year. Teachers who were still resistant to change by that point, she realized, probably would never change and she encouraged them to “move on” so that Drummond could continue to progress.

During the second year of her principalship, Hemingway began talking to her faculty about the process of change. By January, she began to talk about her expectations for her faculty and more concretely about the changes that would be implemented over the next year. Teachers who were still resistant to change by that point, she realized, probably would never change and she encouraged them to “move on” so that Drummond could continue to progress.



Beginning in January of her second year as principal, the focus shifted to the need to implement a balanced literacy model for reading instruction. In order to lead the faculty to the realization that this change was needed, Hemingway sent every teacher in the school to some kind of staff development that emphasized guided reading and a balanced literacy approach. These staff development activities were not local, but elsewhere in the state. Because the teachers were learning about the virtues of balanced literacy from external sources, they were more open to begin implementing the model in their classrooms. Once Hemingway was sure that the majority of her faculty was “on board,” she began to bring consultants to the school to help develop their understanding of the methodologies associated with balanced literacy – guided reading, whole language comprehension strategies, phonological awareness, making and building words, and writing.

Thus, what we found was a principal who knew that change could not come over night but with the right resources and support from the central office and recruitment of some new, well prepared teachers, she was able to tilt the balance of school change in the right direction.

#### **Authoritarian Leadership in a LPS**

A contrasting leadership illustration is found in a Tennessee low performing district. There, Cal Dennis was appointed by the county’s Board of Commissioners to the position of Director of the Glorietta School District only months earlier, replacing the former director who had recently retired. Dennis was formerly the principal of the county’s only public high school and also serves as a county commissioner. The Glorietta School District is plagued by many problems, not the least of which are financial difficulties. Administrators revealed that, although it was already mid-October and two months into the school year, the County Board of Commissioners had only approved the school district’s budget days earlier. Additionally, these administrators seemed overwhelmed by the state and federal expectations and issues of accountability. They expressed a sense of futility in attempting to improve achievement in light of the gravity of the district’s financial woes and the threats of sanctions from the state for employing uncertified teachers and failing to make improvements in achievement. Three of the district’s elementary schools are on the state’s failing list and, therefore, have exemplary educators assigned to them to facilitate improvement. There is no evidence of an articulated plan for professional development, district goals, or policies at the district level. While one of the principals we spoke with mentioned a calendar from the district office that delineated the monthly staff development offered at the schools by the school district, we were never shown a calendar and teachers seemed unaware that there was a calendar, much less a coherent plan for their own learning.

### **Authoritarian Leadership in a LPS**

The principal of the low performing school (Hillsborough), Dr. Ben Washington, is new to Glorietta County, and he was hired recently to turn the problem-laden school around. Washington's primary vocation is as the minister of a large church in another city. In an interview, Washington appeared confident that he will be successful in leading the school toward improvement in academic achievement, and approaches his task with a missionary zeal of one who has come to free the downtrodden, or, rather as mercenary appointed to the task of reclamation.

Speaking with Washington, one is aware that he is never far removed from the pulpit. His manner is that of one who is accustomed to public speaking and preaching, interweaving his speech with humor and catchy phrases. When he assumed his principalship, Washington stated, "Hillsborough Elementary School had more problems than Van Camps has pork-n-beans." While discussing the school, the conversation ultimately turned to the community which was the state's largest slave holding county in the state, many of his students' parents are tenant farmers and few have high school diplomas, and many are on welfare and live in public housing. The problems are further complicated by the difficulty the district has in recruiting certified teachers to work at a failing school. As a result, Washington reported that the school has been forced to retain quite a few interim, uncertified, teachers.

A major part of his job, according to Washington, revolves around the need to involve parents in the educational experience and helping them to understand their responsibility in educational accountability. He envisions a day in the future when Hillsborough Elementary School will become a community center, educating adults at night and children during the daytime. Washington acknowledges that there is an almost insurmountable amount of work to be done in the community. He also recognizes that there is a lack of financial capacity in the district and speaks of writing grants.

He believes his major problems are student discipline and poor teachers not teaching the basics. Once discipline problems are under control, Washington believes he will be able to tackle achievement problems more intently. Elaborating on his campaign to raise test scores, he spoke of his intention to have an assembly with his students, the purpose of which is to "declare war on Terra Nova," encouraging his students to "defeat the enemy," and beginning the countdown to the big showdown."

Washington stated that he would inform teachers at the end of the first school semester about who would be asked to renew their contracts the following year. His classroom observations, therefore, are a part of his evaluation process. According to Washington, he is in the process of "pencil whipping" poor teachers to get them out of Hillsborough Elementary School — although he has few teachers in what might be called a hiring pool. He lamented that he had so many teachers on emergency and alternative licenses, but he also was somewhat reconciled to the fact that he could do nothing about attracting more highly prepared teachers to his school.

The principal spoke at length about the need to concentrate on teaching the basics – reading, writing, and arithmetic. He believed that a return to teaching the basics combined with a positive self-image for his students will be the salvation of his school. Teachers, however, did not speak in terms of teaching the basics – possibly because they perceive that they have been teaching the basics all along. The school had recently added a computer lab equipped with software to reinforce basic skills. In addition, the school has received a grant for electronic "mini-computers" that students can check out to use at home to help reinforce skills.

Washington maintained a strong grip over all aspects of the management of Hillsborough Elementary School, even sitting in on one of the interviews with a teacher. The effect of his presence on the teacher interview, as would be expected, ended the conversation we were hoping to have with this teacher, who was fearful of talking to us frankly in his presence.

Regarding leadership, the two mini-cases displayed above surfaced several issues. First, strong superintendents can more readily recruit and develop strong principals who can, in turn, assist teachers. In contrast to Hillsborough where the principal was focused on “pencil-whipping” under-prepared and out-of-field teachers, in Drummond, the principal was creating what might be called “developmentally appropriate” professional development for her teachers. Quite simply, we found weaker principals and administrators in our lower performing and more impoverished schools. In fact, many of the LPS had experienced significant principal turnover over the last several years, only some of which was voluntary.

Second, more transformative leadership practices seemed to lead to more coherent and productive responses to accountability. For example, teachers in a Kentucky APS believed they were making big strides in the state’s accountability system — largely attributed to a beloved principal there who enjoyed high teacher respect, put a big emphasis on hiring good people, and was known for being fair, flexible, and innovative. This principal was a mid-career professional recruited into the job by the superintendent. She was known for being impatient with some forms of regulation, but a serious advocate for teachers and children. In turn, the teachers reported considerable satisfaction with their professional development and were hopeful with their school’s efforts to improve student achievement.

Third, in labeled states there was a tendency for principals — especially those *not* in the higher performing schools — to bring clear focus to their schools like “fixing the fluff” and “tightening up” many procedures and routines. In North Carolina, a principal described raising expectations in terms of teacher attendance, time on task in classrooms, and strategies to ensure that students were able to articulate not only what they were studying but also how their work fit into an overall class plan or objective. In some ways, this may be a natural response to fact that these schools have higher turnover and less prepared and experienced faculty. However, we found that stronger principals were hard to find and keep for these LPS.

Indeed, leadership surfaced as an indicator of the organizational capacity of a school and district to respond to high-stakes accountability. The cases illustrate the complex decisions to be made and relationships to be forged in order for school leaders to be successful. However, we found little evidence of systematic means to develop kind of teacher leadership that could drive more internal sources of accountability.

## **Conclusions**

With qualitative research into a few schools, one expects to surface nuanced, understandings, as opposed to broad generalizations. We have done so here, with our case studies of 24 schools (and 12 districts) in 6 states, and our effort to understand how certain types of accountability systems *influence* certain teacher responses to their own learning and efforts to improve student achievement. Our findings are expected to be coupled with a set of quantitative findings, where our partners have used teacher surveys and extant databases to quantitatively respond to questions about high stakes accountability, professional development, and student achievement.

In any case, our efforts to answer a set of questions proved to be more complex than we anticipated. The links between accountability and teacher learning were even more nuanced than we expected — or at least as described in previous research into matters of accountability, testing, and professional development. Indeed, much has been written of late about how schools need to develop their own internal accountability systems that in turn can drive improvement and close the achievement gap. As Abelman and Elmore

(1998) noted, the “the reality of particular schools belies the pressure for uniformity behind the emerging accountability systems” (p. 1). In addition, much has been written of late about the difficulties of establishing, assessing, and sustaining effective professional development programs that make a difference for student achievement (The Finance Project, forthcoming). There is nothing that we found that would counter any of these research-base assertions.

### **How do different state accountability systems affect teachers’ responses to their own professional development?**

First of all, we found very little differences, based on state accountability systems, in teachers’ responses to their professional development. In part, teachers and administrators did not fully understand the accountability systems in which they worked; they did not necessarily trust the test results; and sometimes they just did not believe that they could make the kind of difference in student achievement gains posited for their most under-performing and most disadvantaged students. Teachers and administrators in a number of states were often “baffled” by the statistical mechanics of the accountability systems, and could not easily make the connections between student performance, school ratings, and what they needed to learn in order to raise student achievement. Still, the variability of responses may be more striking within states than across them.

In addition, because teachers did not necessarily believe that all students could reach the academic standards posed by the accountability tests, they also believe there was only so much for them to learn. In other words, why seek professional development that is better than what has been experienced in the past, but perhaps only marginally so. Also, why seek professional development that cannot be expected to help one make a difference in producing higher achievement gains on a single test that may not be as statistically reliable as suggested to policy makers and the public.

Nevertheless, teachers took their respective accountability systems seriously, and very few, if any, of the 150 teachers interviewed eschewed responsibility for helping their students achieve. Virtually all of them appeared to believe in the importance of accountability, and recognized how the systems in place helped them focus more intently on what needed to be taught. However, these same teachers expressed concern about how the accountability systems in place, no matter how configured with different sanctions and rewards, were narrowing the curriculum, and forcing them to abandon other important content in social studies and science and/or more creative means to teaching and learning. However, an important caveat is in order. Teachers from the higher performing schools, were far more concerned about such matters — primarily because they were feeling less pressure to teach to the test, and perhaps because of their experience and credentials, they had more confidence in their capacity to teach beyond the test.

### **How do teachers from schools with different accountability ratings respond to their own professional development?**

In the face of growing accountability demand, both school systems and practitioners are taking professional development more seriously. And, teachers reported uniformly that the professional development they were experiencing was getting better, albeit unevenly. By this, teachers revealed how their own learning was less driven by the proverbial, one-shot workshops and more influenced by standards-based or curriculum-based programs that helped them teach differently. Teachers also reported more follow-up from coaches than previously experienced.

Resources and leadership mattered a great deal. Higher performing schools and those with more financial wherewithal were able to offer their teachers much more sound professional development. Still, while teachers from the HPSs were more likely to report that they had a broader range of opportunities to learn, they also were likely to reveal the limits of their professional development. We saw little evidence of systematic professional development that allowed teachers opportunities to problem solve how to teach certain concepts or analyze the work of specific students in order to build a more adaptive curriculum. While all teachers were being asked to look at student test score results as a means to improve achievement, we saw little evidence of schools helping teachers use data more comprehensively and even “lesson study” as a means to drive teacher development and student achievement. Across the six states, we found just a

few instances of teacher study groups, and rarely would teachers have a chance to observe and critique another teacher teach a lesson.

However, we were most likely to hear reports of increased teacher collaboration — especially in the higher performing schools that we visited. There we saw more evidence of teachers learning together to introduce math concepts in new ways, work with different levels of readers, and utilize technology in non-regimented ways. All schools were using grade level meetings in new and seemingly more productive ways — although few schools have built much more than 50 minutes a day for teachers to engage in collaborative planning and professional development.

Teacher learning could easily be confounded by school leadership. In some of our lower-performing, more rural, under-resourced schools, we found striking examples of teachers feeling threatened by administrators who were known as “snoopers” who, according to them, were visiting their classrooms not to help, but to uncover evidence to be used against them. For these teachers, these feelings and beliefs seemed to run counter to the kind of openness and professional behavior needed to dramatically change teaching practices. At the same time, in these lower performing schools, where the pressure for higher performance was understandably greater, teachers seemed to expect less autonomy than their counterparts in higher performing schools. Here more teaching to test is prevalent and accepted.

### **How do schools and districts mediate high stakes accountability, professional development, and student achievement?**

As one would expect, how schools and districts mediate the demands of accountability, professional development, and student achievement depends on its organizational capacity, which has a great deal to do with its expertise and internal workings, as well as strategic responses to state policies and local culture. To a large extent, we found in several of our lower performing, under-resourced schools, the blatant lack of capacity to meet higher standards, irrespective of the kinds of accountability systems and subsequent rewards and sanctions. In any case, we found a continuum of responses along two broad approaches.

On the one hand, we found schools and districts — primarily higher performing ones — that would align standards, curriculum, and new professional development to give relatively well-prepared and more experienced teachers new focus and skills to reach more of its students and to improve scores on the high stakes tests. In this case, teachers would make adjustments (sometimes marginal, sometimes substantial) to their teaching, reach a few more students, although they may not have sufficient skills to help their most challenging students. However, at least at the time of our visits, these schools and districts seemed to score well enough on their respective accountability report card that they did not have to worry too much about low performance or poor public image. Despite the fact that these schools and districts benefited from stronger (and more sophisticated) administrators who believed in and could allocate new resources for “better” and more focused professional development, teacher learning still does not resemble much of what is classified as “effective” by the literature.

On the other hand, we also found schools and districts — primarily lower performing ones — that would align standards, curriculum, and new professional development, but would do so for a relatively under-prepared and less experienced group of teachers. In this case, the alignment would take the form of a prescriptive curriculum to teach and a more narrowly construed set of professional development activities to follow. However, here we found teachers still engaged in a wide variety of teacher learning opportunities, ranging from learning CPR to using PowerPoint, that had little to do with helping them improve student achievement. Teachers, and in some cases, principals in these lower performing schools believed that their habitually poor ratings were inevitable, and they were somewhat resigned to the fact that they would suffer a poor public image. Teachers and administrators, in a number of these schools and districts, did not know what effective professional development would look like and how they could create teacher learning opportunities that could dramatically change the way students are taught and what they are expected to learn. These schools and districts had no perceived means of recruiting, preparing, and retaining the kinds of teachers needed to do the job that could substantially ratchet up student achievement.

We sought to understand under what conditions accountability systems would most likely promote effective professional development. We went about our work with the belief and assumption that for an accountability system to be effective it needs to produce indicators that not only reveal which schools were doing better (or not), but also discloses why they were doing better (or not), and what educators should do to improve student learning. None of the accountability systems in which our 24 schools operated could produce a more comprehensive set of indicators based on such an assumption. Our case study findings suggest that to do so would require dramatic new ways to measure and report on school leadership, professional development opportunities, and teacher working conditions. These issues are, for the most part, totally off the radar screen of those policy makers focused on using accountability to fuel significant increases in student achievement — especially now in light of No Child Left Behind and new mandates to report on annual yearly progress for each subgroup in every school. Without such information and public reporting on issues of professional development, school leadership, and working conditions, we suspect that many of these schools will continue to do as they have done, their teachers will continue to learn as they have learned, and their students will continue to perform generally as they have in the past.

## References

- Abelmann, C. and Elmore, R. (1998). When accountability knocks, will anyone answer? Unpublished draft. Paper prepared for Consortium for Policy Research in Education.
- Baron, J. B. (1999). *Exploring high and improving reading achievement in Connecticut*. Washington: National Educational Goals Panel.
- Berry, B., Groves, P., and Noblit, G. (forthcoming). The Effects of Long-Term Investment in the Improvement of Teaching: North Carolina, 1980-2000. Seattle, WA: Center for the Study of Teaching Policy.
- Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership*, 57, (8), 28-33.
- Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24, 305-331.
- Clune, W. E. (1993). The best path to systemic educational policy: Standard/centralized or differentiated/decentralized? *Educational Evaluation and Policy Analysis*, 15, 233-254.
- Cohen, D. K. (1996). Rewarding teachers for student performance. In S. Fuhrman & J. O'Day (Eds.). Rewards and reform: Creating educational incentives that work (pp. 60-112). San Francisco: Jossey-Bass.
- Cohen, D. K., & Hill, H. (1997). Teaching and learning mathematics in California. Paper presented at the annual meeting of the American Educational Research Association, Chicago. April.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. (2000). Education resources, institution, and research. Paper prepared for Center for Study of Teaching and Policy. Seattle: University of Washington.
- Darling-Hammond, L. and Snyder, J. (1993) "Reframing Accountability: Creating Learner-Centered Schools." In Ann Lieberman (ed.), The Changing Context of Teaching, Ninety-first Yearbook of the National Society for the Study of Education, pp. 11-36. Chicago: University of Chicago Press, 1992. Reprinted in Creating Learner-Centered Accountability, pp. 1-20. New York: National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University.
- Education Trust (2003). In Need of Improvement: Ten Ways the U.S. Department of Education Has Failed to Live Up to Its Teacher Quality Commitments. Washington DC: Author.
- Education Week (2003). *Quality Counts 2003: If I Can't Learn From You*. Washington, DC: Author.
- Elmore, R. F. (1997). Accountability in local school districts: Learning to do the right things. *Advances in Educational Administration*, 5, 59-82.
- Firestone, W. A. (Summer, 1989). Using reform: Conceptualizing district initiative. *Educational Evaluation and Policy Analysis*, 11, (2), 151-164.
- Firestone, W. A., & Fairman, J. C. (1998) The district role in state assessment policy: An exploratory study. Rutgers Graduate School of Education. Paper presented at the American Educational Research Association in San Diego, CA. April.
- Ginsberg, R., & Berry, B. (1998). Expanding responsibility to enhance accountability. In R. J. S. MacPherson (ed.) The Politics of Accountability: Educative and International Perspectives. Thousand Oaks, CA: Corwin Press.

Goertz, M., Massell, D., & Chun, T. (1998). District response to state accountability systems. CPRE, University of Pennsylvania and University of Michigan; RAND. Paper presented at Annual Meeting of the Association for Public Policy Analysis and Management. March.

Grissmer, David, and Flanagan, Ann (1998). Exploring Rapid Achievement Gains in North Carolina and Texas. National Education Goals Panel.

Grissmer, D.W, Flanagan, A, Kawata, J., and Williamson, S. (2000). Improving Student Achievement: What State NAEP Test Scores Tell Us. Santa Monica, CA: RAND.

Hanushek, E. A., et al. (1994). *Making Schools Work: Improving Performance and Controlling Costs*. Washington, DC: The Brookings Institution.

Haslam, B. (2000). Personal email communication. June 5.

Hawley, W. D., & Valli, L. (1998). The essentials of effective professional development: A new consensus. In G. S. L. Darling-Hammond (Ed.), *The heart of the matter*. San Francisco: Jossey-Bass.

Henry, G.T. and Opfer, V. D. (2003). Responses to High Stakes Accountability in the South: Final Report to the Spencer Foundation. Atlanta, GA: Georgia State University, Andrew Young School of Policy Studies.

Hess, F. M. (2002). *I say "refining," you say "retreating": The politics of high-stakes accountability*. Paper presented at "Taking Account of Accountability: Assessing Politics and Policy," Harvard Program on Education Policy and Governance, Cambridge, MA.

Kearns, D. T., & Doyle, D. P. (1988). Winning the brain race: A bold plan to make our schools competitive. San Francisco: Institute for Contemporary Studies.

Kennedy, M. (1998, December). Form and substance in inservice teacher education. Madison WI: National Institute for Science Education.

Ladd, H. (1996). Holding schools accountable: Performance-Based reform in Education. Washington DC: Brookings.

Little, J. W. (1993, Summer). Teachers' professional development in a climate of education reform. Educational Evaluation and Policy Analysis. 15, (2), 129-151.

Linn, R. L., Dunbar, S. B., Harnisch, D. L., & Hastings, C. N. (1982). The validity of the Title I evaluation and reporting system. In E. R. House, S. Mathison, J. Pearsol, & H. Preskill (Eds.), Evaluation studies review annual 7, 427-442. Beverly Hills, CA: Sage Publications.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher* 29(2): 4-16.

Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31, 3-16.

Massell, D. (1998). *State Strategies for Building Capacity in Education: Progress and Continuing Challenges* (RR-41). Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania, Graduate School of Education.

Moe, T. M. (2002). *Politics, control, and the future of school accountability*. Paper presented at "Taking Account of Accountability: Assessing Politics and Policy," Harvard Program on Education Policy and Governance, Cambridge, MA.



Murnane, R. J., & Levy, F. (1996). Teaching to new standards. In Susan H. Fuhrman & Jennifer A. O'Day (Eds.), Rewards and Reform: Creating Educational Incentives that Work, (pp. 257-293). San Francisco: Jossey-Bass.

Newman, F. M., & King, M. B., and Rigdon, M. (1997). Accountability and school performance: Implications from restructuring schools. Harvard Educational Review 67, (1) Spring 41

Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers*. Boston, MA: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College.

Rothman, R. (1995). Measuring up: Standards, assessment and school reform. San Francisco: Jossey-Bass.

Shippo, D., Kahne, J., & Smylie, M. A. (1999). The Politics of urban school reform: Legitimacy, city growth, and school improvement in Chicago. Educational Policy, 13, (4), 518-545.

Spillane, J. P., & Thompson, C. L. (1998). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. Education Evaluation and Policy Analysis.

Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In Hamilton, L. S., Stecher, B. M., and Klein, S. P. (eds.) *Making Sense of Test-Based Accountability in Education*. Washington, DC: RAND.

## Appendix A: Sample NAEP Grade 4 Items – Math and Reading

Note: The two NAEP questions were downloaded from the NAEP website and are presented as examples of what NAEP asks students at 4th, 8th, and 12th grades. The released questions in this tool do not represent complete coverage of the content, cognitive skills, and range of difficulty in the NAEP assessment for a particular subject area. (<http://nces.ed.gov/nationsreportcard/itmrls/>)

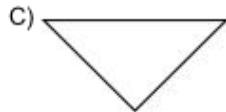
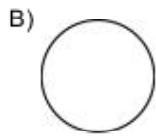
Subject: Math

Grade: 4

Block: 2003-4M6

No.: 07

7. Alan says that if a figure has four sides, it must be a rectangle. Gina does not agree. Which of the following figures shows that Gina is correct?



Subject: Reading  
Grade: 4  
Block: 2003-4R9  
No.: 04

4. Tell how Elisa and Cory are alike and different. Use examples from the story to explain your opinion. (Story below)

### **The River**

*Based on a true story*  
**By Yetti Frenkel**

"Sh," whispered Elisa. "I think she's coming!"

Elisa and Cory stifled their giggles and crouched behind the pine tree. Peeping out through the snow-covered branches, the children held their breath and listened for the tinkle of Minnie's collar as the old dog tried to find their hiding place. It was usually the hound's favorite game, but today the only sounds the children heard were the wind whistling softly across the frozen snow and ice cracking on the river.

Cory shivered with cold. "I wonder where she is," he said. "I hope she isn't off chasing a deer."

Elisa snorted. "Minnie's too lame for that. I bet she went home to wait where it's nice and warm."

Cory looked doubtful. "She wouldn't go home without us," he said. "Maybe she got ahead, and we didn't notice. Let's go to the bridge and see if she's there."

They started down the trail at a quick pace, glad to be moving again. The bare branches of the trees rattled forlornly as they tramped through the frozen snow.

Elisa struggled hard to keep up with her older brother. "Wouldn't it be easier to walk on the ice on the river?" she called to him.

Cory slowed his pace and waited for her to catch up. "It's too dangerous," he said. "The water is still flowing underneath, and the ice is thin. We might fall through." He held out a mittened hand. "I'll help you."

"No, thanks," said Elisa stubbornly. "I can keep up." But she was secretly glad when Cory walked beside her until they reached the bridge.

The old wooden bridge spanned the widest part of the river. In summer they often came here to fish or lie in the sun, but now it was a desolate, wind-swept place. They could hear the water gurgling softly beneath the ice as they looked out over the railing, hoping to glimpse Minnie walking along the bank.

Cory cupped his hands to his mouth and called, "Minnie, Min-nie!" His voice echoed back to him from the lonely woods. "I don't see her, Elisa. Do you?" he asked.

Just then Elisa gave a startled cry, and Cory turned sharply to see Minnie ten feet from shore. The old dog had fallen through the ice and was paddling in desperate circles.

"Hang on, Minnie, I'm coming!" Cory cried, racing toward the river. Elisa was already ahead of him, pulling off her coat, scarf, and mittens, ready to plunge in and save her dog. Blinded by tears, she stumbled out onto the ice.

Cory caught up with her and pulled her back. "Do you want to drown yourself?" he shouted. His face was white as he held out the warm clothes she'd dropped. "Put these back on and let me think of something." He looked grimly at the river.

Elisa sobbed as she struggled into her coat. "You can save her, can't you, Cory? She won't die, will she?"

"Of course not," he said, wishing he felt as confident as he was trying to sound.

The sight of her masters had given Minnie new hope, and she managed to get her front paws up on the ice. She scratched and clawed frantically at the slippery surface, but her hind legs were too arthritic to be of much help. For a moment her frightened brown eyes met Cory's, then she slipped back into the icy water and began wearily swimming once more.

Cory searched the bank until he found a long, twisted branch. Holding it firmly, he maneuvered the end until he had it hooked under Minnie's collar. "C'mon, girl," he said to the tired dog. She heaved her front paws onto the ice and struggled desperately while he tried to help her by pulling on the branch. But frost and moisture had made the wood brittle, and it snapped almost immediately. Once more Minnie struck out swimming, but now her head was barely above the surface of the water.

A terrible thought crossed Cory's mind - Minnie was going to drown before their eyes. It's not fair, he thought. Why doesn't someone come along to help us? He scanned the woods for a game warden or hunter, but saw no one. The woods were dark and silent, waiting. "I don't know what to do," he said, frightened.

"I know what to do," cried Elisa. "I'm going to help her!"

Once again Cory grabbed his sister's arm to prevent her from going out onto the ice. She bit and kicked at him like a small fury as tears of frustration ran down her cheeks.

"Listen to me!" yelled Cory. "I thought of something, but I need your help." Elisa wiped the tears from her face. "I'm going to lie down on the ice and try to crawl to Minnie. You lie down behind me and hold my ankles. Don't let go, no matter what, and don't stand up. Understand?" Elisa nodded, sniffing.

Cory lay on the ice so that his weight would be distributed more evenly and there would be less chance of breaking through. He felt Elisa's hands close around his ankles. As he inched his way forward, he could hear the water rushing beneath the ice. A few feet in front of him was the deep green hole where the dog had broken through. Cory's heart pounded with fear, but he bit his lip and kept going. At last he reached the edge of the hole and threw his arms around Minnie's neck. It felt reassuring to have a hold on her, but he soon realized that there was little else he could do. The ice was slippery, and every time he tried to pull her out, he began to slide forward himself.

"Have you got her?" called Elisa anxiously.

"Yes," Cory yelled over his shoulder, "but I can't" - Before he could explain, he found himself being pulled back across the ice with Minnie in his arms. He looked around in amazement, expecting to see a big man with a broad grin standing behind him, but there was only his sturdy little sister, laughing and crawling over the ice to throw her arms around the shivering dog. "How did you ever do that?" cried Cory. "You're not that strong!" Then as Minnie, tail wagging wildly, began to lick his face, he saw what had happened.

Elisa had put her wool coat down on the ice to protect her from the cold. The warmth of her body lying on the top of it had made the wool fibers stick firmly to the ice so that when she pulled on Cory's legs, he slipped across the surface to her as easily as a cork popping from a bottle.

Cory grinned in admiration. "You sure are one smart little sister!" he said, tousling her hair. He took off his plaid shirt and dried Minnie with it. "It's a good thing we were all together today," he said to the old dog softly as he rubbed her lopsided ears. She wagged her tail in agreement, and the three hurried toward the warmth of home without looking back.

Copyright © 1992 Yetti Frenkel.