
The District Effect: Systemic Responses to High Stakes Accountability Policies in Six Southern States

V. DARLEEN OPFER

University of Cambridge

GARY T. HENRY

University of North Carolina at Chapel Hill

ANDREW J. MASHBURN

University of Virginia

High stakes accountability (HSA) reforms were enacted in state after state and federally through the No Child Left Behind law, based on the belief that incentives that have consequences attached are effective ways to motivate educators to improve student performance. Our focus for this article is on school district level responses to HSA reforms that could produce positive changes in teaching and learning. We set out to determine whether a district effect was present in the implementation of HSA systems in six southern states and whether that effect was accompanied by the types of activities previously identified in the research literature as being associated with changes in teaching and learning and student achievement. We tested a theory of action that assumed that HSA would cause school districts to develop coherent instructional strategies that would be evidenced by the provision of coherent, high-quality professional development and the alignment of district policy and resources in support of school improvement. These activities on the part of districts would then improve student achievement as measured by state tests.

Introduction

No Child Left Behind (NCLB) legislation “nationalized” a set of reforms that had been championed by those advocating for standards-based education in many states across the United States. Prior to this federal legislation, reforms that included high stakes accountability (HSA) had largely occurred at the

Electronically published October 10, 2007

American Journal of Education 114 (February 2008)

© 2008 by The University of Chicago. All rights reserved.

0195-6744/2008/11402-0006\$10.00

Impact of High Stakes Accountability

state level, with several of the most visible reforms taking place in the South, notably, Texas, Kentucky, and North Carolina. The South, where improvements in student performance have not matched the pace of economic change and where less unionization has occurred (Cooke 1985; Hirschman and Blankenship 1981; Stamas 1981), was especially ripe for these educational reforms. The documented need for these reforms in the South intensified after 1990 when the National Assessment of Educational Progress began to report state-by-state scores. For the six southern states included in this study, percentages of children judged proficient on these assessments were often in the single digits when initially tested, and, after 10 years, scores in the South still lagged behind the rest of the nation. Only North Carolina students have scored above the national average on any of the last eight assessments, and, by 2000 when we began this study, only 12 out of 48 times have any of the states performed within the range to be judged statistically “at the national average.”

High stakes accountability reforms were enacted in state after state, and federally through NCLB, based on the belief that incentives that have consequences attached are effective ways to motivate educators to improve student performance. Consider a quote from an early proponent of HSA:

Educators and students, at all levels of the school system, should be rewarded for actions that improve student performance. Education is too complex an endeavor to manage by rote, or, as is often attempted today, by curricula and rigid rules handed down from state and local boards of education. These agencies fail to recognize that teachers and other local decision makers inevitably have great leeway to improve or reduce school efficiency. Performance incentives that reward them for progress toward goals of the schools—while recognizing their freedom to determine how that progress is best achieved—are the best way to focus teachers, principals, and other school personnel on improving education. (Hanushek 1994, xvi)

V. DARLEEN OPFER is a senior lecturer in research methods, school leadership, and school improvement at the Cambridge University. Her research interests include the impact of educational policies and politics on schools, teachers, and students. GARY T. HENRY is the William Neal Reynolds Distinguished University Professor at the University of North Carolina at Chapel Hill. His research interests include the effects of education policy reforms on student outcomes and teachers. ANDREW J. MASHBURN is a senior research scientist at the Center for Advanced Study of Teaching and Learning at the University of Virginia. His research focuses on educational measurement, evaluation, and policy.

Amid the slowly growing body of evidence that supports the positive effects of HSA systems on student achievement (Carnoy and Loeb 2002; Grissmer and Flannagan 1998; Raymond and Hanushek 2003), questions have begun to arise concerning responses to HSA that may be responsible for improvements in student achievement. While the evidence is not yet conclusive that HSA positively influences students' test scores, in the wake of No Child Left Behind legislation, it is timely to investigate the responses to HSA that may explain the effects.

Responses to accountability mandates are generally believed to occur at three levels: (1) changes in teachers' behaviors directed toward improving their teaching, (2) changes in school level support for improving student outcomes, or (3) changes in school district support for improving student outcomes. Our focus for this article is on school district level responses to HSA reforms that could produce positive changes in teaching and learning. However, we implemented a research strategy and obtained data that enabled us to assess the relative strength of the three alternative ways that HSA could affect student achievement.

After identifying the theory of action that could be undertaken by school districts, we describe the HSA policies that were in place in each of the six states in our study when data for the analysis were collected. Then the sample, measures, and methods of analysis that we use to test the district response are presented. Our findings follow: first describing teachers' responses to surveys that asked about district support and then results from hierarchical linear models (HLM) comparing teachers' attitudes and behaviors across these southern states. Finally, we offer some conclusions about the theories that seem to hold the most promise for explaining how school districts respond to HSA reforms.

Theory of Action: How Do Districts Respond to High Stakes Accountability?

In identifying a theory of action for how districts respond to HSA policies we follow a straightforward line of reasoning: districts are compelled to develop coherent instructional policies, including professional development opportunities; these policies focus teacher attention on instructional improvement; and this instructional focus results in improved teaching and learning. The research evidence on district effects and the role of the district in the delivery of coherent instructional policies is helpful in detailing how this theory of action may occur.

Impact of High Stakes Accountability

District Effect Research

Until recently, little attention has been given to the actions districts take to support instructional improvement. Many have assumed that school districts lack the organizational conditions necessary for success—mainly specific goals, clear technologies, and timely and precise performance feedback (DiMaggio and Powell 1983; Meyer and Rowan 1977). They tend to be more concerned with general survival than with the achievement of specific goals (Rowan and Miskel 1999) and thus adopt formal structures that are defined by government and the professions (Scott 1995). For example, in a study of district influences on curriculum and teaching in fourth-grade mathematics, Floden and his colleagues (1988) concluded, “The picture that emerges is one of districts with a vague intention to direct instructional content, but without any considered strategy for doing so. Districts do not leave teachers to their devices, but neither do they make systematic use of the tools available to persuade teachers to adopt patterns of content decision making. Rather than deciding to set central instructional goals and then trying to communicate those goals through all available means, districts tend to make unconnected decisions that do not lead to any clear pattern of curriculum policies” (Floden et al. 1988, 98). Other authors have gone further to conclude that school districts do not just neglect to support instructional reform but are actually an impediment (Elmore 1993; Hill 1995).

Despite the evidence that school districts have not traditionally played strong roles in teaching and learning, many have long thought that local districts are essential to the implementation of instructional reform. An early study of federal policies designed to improve curriculum and teaching found that successful projects were distinguished by the active involvement of district level administrators (Berman and McLaughlin 1978). Other studies have also shown that when districts take an active role in promoting instruction, new activity occurs at the school level (David 1990).

As school-based reform strategies have been shown to be insufficient (Bryk et al. 1998), some researchers have argued that the current standards-based HSA reforms are even more dependent on local district capacity than were previous reform initiatives (Elmore and Fuhrman 1994). District influence in HSA policy implementation has gained importance as researchers realized that districts acted as intermediaries of external policies that target teaching and learning (Craciun and Snow-Renner 2003; Massell 2000; Skrla et al. 2001; Spillane and Thompson 1997). McLaughlin and Talbert (2003) have shown in their study of Bay Area districts that supporting schools, rather than initiating reform, was an essential role in school reform success. Elmore and Burney’s (1998a) research on District #2 in New York City illustrated that district instructional focus and targeted professional development strategies

influence the practices of principals and teachers and ultimately impact student achievement. Other recent case studies of districts have claimed that their actions can improve overall student outcomes and reduce ethnic and socioeconomic gaps in performance (Hernandez 2003; Manpower Development Research Corporation 2002; McKenzie Group 1999; Skrla et al. 2001; Togneri 2003). Thus, case-based evidence has shown that school districts can impact teaching and learning; however, it is as yet unclear whether the districts studied are anecdotes or instances where real district effects that are broadly feasible and replicable have been observed.

Despite the uncertainty about districts' ability to systematically affect student learning, scholars have often assumed that an effect occurs and then concentrated on identifying the parameters of district actions. Three sets of research studies have begun to identify the context for district response and the actions taken to improve instruction. First, different researchers have attempted to ascertain when districts become engaged in instructional improvement. Geortz, Massell, and Chun (1998) identified clear responses by districts to HSA, and the pattern of these responses was related to the strength or weakness of the accountability system—prescriptiveness, rigor of measures, alignment of system with other policies, and stability of the system itself. Similarly, Firestone and Fariman (1998) found that sanctions, such as reconstitution, were more likely to engender positive district responses than just “embarrassment” from reporting poor results publicly. More recently, a survey conducted by Mayo and McIntyre (2003) found that in school districts located in states with high stakes testing, district leaders spent more time on instructional leadership than those in states with low stakes testing. The initial evidence from these studies suggests that accountability systems can trigger district actions when the accountability system in place has sufficient strength and longevity.

A second group of researchers has attempted to categorize the types of responses that districts make to support instructional improvement. Spillane and Thompson (1997) have identified three dimensions of district activity to support teaching and learning—human capital, social capital, and physical capacity. Human capital actions were defined as those engendering personal commitment to learn new practice and to do so collectively; social capital actions involved developing organizational norms that promoted trust, collaboration, and internal and external networks for information; and physical capacity actions involved the allocation of resources to staffing, time, and materials. Supporting this work, Gamoran et al. (2003) drew on sociocultural theory to propose that groups (social), practice (human), and organizational resources (material) are the building blocks for district actions and effects on teacher change.

In addition to the categories of action identified by Spillane and Thompson and also by Gamoran and his colleagues, a third set of studies have focused

Impact of High Stakes Accountability

on documenting specific district-level activities thought to be important to instructional improvement. According to Elmore (2003), the activities that districts take to impact teaching and learning attend to:

- How much teachers know about their subject and the pedagogical knowledge required to improve student knowledge in that subject;
- How leadership is defined and distributed;
- How schools are organized and how people in the organization solve problems related to instructional practice;
- The resources available to the school including time, money, information, materials, and external support;
- Internal accountability, or coherence—the shared norms, values, expectations, structures, and processes that provide the relationship between individual actions and collective results.

For example, in one of the earliest studies on district influence, Murphy and Hallinger (1988) studied 12 California school districts that were deliberately selected because of their demonstrated ability to produce high levels of student achievement. The study indicated that three-fourths of the districts “had a preferred approach to instruction that they expected all teachers to emphasize,” two-thirds of the districts had curriculum objectives, half of the districts screened and selected principals based on their knowledge of curriculum and instruction, about 40 percent of the staff-development activities focused specifically on district priorities, and in 11 of the 12 districts superintendents had a direct role in instructional oversight (177–78). The study also found that “although there was substantial evidence that the rational elements in these school systems were a product of district direction and coordination, the elements appeared to work because these systems were living, adaptive organisms rather than collections of codified procedures” (178).

McLaughlin and Talbert’s (2003) study of Bay Area, reforming school districts identified similar activities that characterize these districts, including:

- Identifying the unit of reform as the system;
- Creating a learning community at the central office level;
- Maintaining a coherent focus on teaching and learning;
- Supporting professional learning and instructional improvement;
- Using data-based inquiry and accountability.

Other studies have identified district level activities related to school leadership, unity of purpose, a clear focus, parent and community involvement, professional community, sustained attention by staff, and shared values for student learning (Bryk et al. 1990, 1993; Coleman et al. 1982; Elmore 1996; Hill and Celio 1998; Newmann and Wehlage 1995).

Coherent Instructional Policy

The single most consistent finding across all the studies of district activity is the presence of a coherent strategy that relies on a combination of both bureaucratic activities—those initiated by district administrators—and professional activities—those that rely on teachers to define and enforce standards of practice (O’Day 2002). Districts that have been shown to be effective at implementing HSA and improving student achievement emphasize, as Rowan (1990) has shown, both commitment and control activities. Or, in the words of Smith and O’Day (1991), “What is needed is neither a solely top-down nor a bottom-up approach to reform but a coherent systemic strategy” (234). Some argue that without coherence in the activities undertaken by districts what occurs is “compliance without capacity” (Debray et al. 2003, 84). Others claim that a lack of coherence in the use of resources—especially professional development—at the disposal of districts has negated their possible influence (Corwin and Borman 1988; Firestone et al. 2005; Floden et al. 1988). Yet most recognize that “such coherence is quite rare in the blizzard of often divergent guidance for instruction that typically blows over the U.S. public schools” (Cohen and Hill 2001, 9).

Various scholars have defined coherence in different ways, but all are unified in their insistence that to support higher student achievement, policy and practice must be unified. For example, in early accountability research, Smith and O’Day (1990) found that by aligning a broad range of policies, achievement by poor and minority students could be improved. In a volume edited by Fuhrman (1993), several authors documented the effects and importance of coherence across various levels of the educational system.

Professional Development as Coherent Instructional Policy

More recently, this work has shown that coherence applies directly to the professional development practices offered by the district. For example, Cohen and Hill’s (2001) study of California’s efforts to improve math teaching and learning provides substantial evidence that coherent professional development is necessary for reform policies to impact teaching practice and ultimately student achievement. Their study found that California’s policy was effective only when teachers had significant opportunities to learn how to improve teaching. The association between policy reform and improvements in student achievement was mediated by what teachers knew and did in their classrooms as a result of professional development opportunities. Specifically they found that when teachers reported having more opportunities to learn, student

Impact of High Stakes Accountability

achievement was higher. Cohen and Hill thus conclude that “Many efforts to ‘drive’ instruction by using ‘high-stakes’ tests [fail] either to link the tests to the student curriculum or to offer teachers substantial opportunities to learn” (151). Thus, one form of instructional coherence for school districts is the provision of coherent professional development activities that link reform efforts, instructional practices, and assessment with student achievement.

As Spillane and Thompson (1997) claim, “the LEA’s capacity to support ambitious instructional reform [is] primarily as a capacity to learn the substantive ideas at the heart of the new reforms and help teachers and others within the district learn these ideas” (199). An illustration of their claim is Elmore and Burney’s study of District #2 (1998a, 1998b, 1999). They describe coherence between professional development and the routine administrative functions of the district. For example, monthly principals’ meetings focused on student performance data and instructional practice, and “walk-throughs” involving classroom observation and reflection were conducted by central administrators and principals together. They also describe the variety of devices adopted by the district to sustain instructionally focused connections between principals, teachers, and the district, including intervisitation, mentorship, and internships. Elmore and Burney clearly show how districts contributed to school improvement through coherent administrative action, professional development, and the establishment of professional community.

Firestone et al. (2005) have similarly noted the important role of the district in the provision of coherent professional development. In their study of three school districts in New Jersey, Firestone and his colleagues describe the ways that district leadership can provide coherent teacher professional development, which in turn influences the quality of instruction provided by teachers. In their districts, coherent professional development had three elements that echoed the professional development research more generally (e.g., Birman et al. 2000; Desimone et al. 2002; Garet et al. 2001):

- Consistency of focus—allowing teachers to develop in-depth knowledge on a specific subject or topic;
- Extended and distributed time spent on professional development—to promote long-term change, it was not “one-shot”;
- Learning opportunities that modeled the instructional approaches teachers were expected to employ—including problem solving, situated learning, and the use of actual student work.

Firestone and his colleagues conclude that districts play a key role in supporting instructional reform by being the primary designers and deliverers of learning opportunities for teachers, and if they do so in a focused, coherent fashion, they can influence teaching practice (Firestone et al. 2005).

Hightower’s (2002) study of San Diego also indicates the important role of

district coherence of professional development in developing capacity for change. In contrast to the scattered character of traditional professional development, the professional development offered by the San Diego district clearly focused on instructional goals and was configured in a coherent, cumulative, multiyear program. Other studies have identified a lack of coherence, as evidenced by disjointed and diverse district activities, as the primary reason for failure of school reform (Allington and Johnson 1989; Cohen and Ball 1996; Elmore 1996; Newmann and Wehlage 1995; Smylie et al. 1998). Thus, coherent instructional policies within districts are often evidenced by professional development activities specifically tied to district goals for instruction and larger reform efforts.

Given the research to date on school districts, the importance of coherent professional development in the implementation of reform efforts and the improvement of student achievement, and the role of districts in providing coherent instructional policy, including professional development, we set out to determine whether a district effect was present in the implementation of HSA systems in six southern states and whether that effect was accompanied by the types of activities previously identified. We tested a theory of action that assumed that HSA would cause school districts to develop coherent instructional strategies that would be evidenced by:

- provision of coherent, high-quality professional development;
- alignment of district policy and resources in support of school improvement.

These activities on the part of districts would then improve student achievement as measured by state tests.

Policy Profiles in Six Southern States

The district response from HSA is expected to vary as a function of the individual policy contexts of each of the states in our study. To understand the variation that exists across the states, policy profiles were constructed. The profiles described each state's policy context in relation to testing policies, professional development policies, and HSA policies. These three areas of policy were examined because of their relationship to elements in our model—HSA leverages district action targeting instructional improvement (in the form of high-quality teacher professional development) that results in improved student achievement as measured by state tests. In order to ascertain the responses to the accountability system in place, we collected data on the system that was implemented in each state in 1999–2000. While many of the systems were in flux during the 2000–2001 school year when teacher survey data were

Impact of High Stakes Accountability

collected, we looked at the HSA systems as implemented, including testing, rewards, sanctions, and labeling.

In addition to describing these policies for each state, the policies were also examined to identify levels of cohesion that exist within and between policies, the maturity of the policy system, the stability of the policies over time, and the strength of the policies (especially the accountability policies). Each of the types of policy and these policy conditions are discussed below.

Testing Policies

In 1999–2000 there was a great deal of variation in the testing policies of our studied states (see table 1 for an overview of these tests). For example, four states gave norm-referenced tests and four states gave criterion-referenced tests, four states gave writing assessments, five states gave high school exit or competency exams, and only one state gave end-of-course tests. The absence of criterion-referenced tests in two states and end-of-course exams in most states indicated that instructional standards and the testing policies that existed in these states were not completely aligned. This lack of alignment could weaken the connection between HSA, professional development, and student achievement—in essence making it hard for schools and districts to identify the content-specific objectives on which they need to focus teachers' attention and professional development.

Additionally, the testing policies were very unstable in 1999–2000. The states in our sample were undergoing significant changes in their testing policies, with the exception of Kentucky. Most of the states were adding combinations of criterion-referenced, end-of-course exams, and gateway exams in order to hold students or schools accountable and to measure student performance. On the whole, the states' testing policies were increasing the number, types, and frequency of testing. By 2006, all the states expected to have a combination of criterion-referenced tests and/or end-of-course exams that would cover all school levels—elementary, middle, and high school. In the meantime, significant changes in the test content and delivery were under way. The rapid expansion of their testing programs may be accompanied by increased pressures on schools and teachers to understand tests and testing, utilize test information to improve instruction, and, ultimately, improve student performance. This pressure may manifest itself in multiple forms of teacher and school level responses. As these new tests are implemented, we would expect to see additional professional development offerings or requirements. However, because of the rapidity and amount of change, we may also expect to see some confusion among schools and teachers about the tests and their

TABLE 1

Testing Policies by State

	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
What types of tests were required by state in 1999–2000?	Norm-referenced, writing assessment, high school exit exam	Criterion-referenced, norm-referenced, writing assessment	Norm-referenced and high school exit exam	Criterion-referenced, writing assessment, high school competency exam, end-of-course tests	Criterion-referenced test and high school exit exam	Combined norm/criterion-referenced test, writing assessment, high school competency exam
What was the schedule of tests given in 1999–2000?	Norm-referenced (grades 3, 5, 8); writing assessment (grades 3, 5, 8, 11); high school exit exam (grade 11)	Norm-referenced (grades 3, 5, 10); writing assessment (grades 4, 7, 12); criterion-referenced (grades 4, 5, 7, 8, 10, 11)	Norm-referenced (grades 5 and 8); high school exit (grade 11)	Criterion-referenced (grades 3–8); writing assessment (grades 4 and 7); high school competency (grades 9 and 10); end-of-course exams (grades 9–12)	Criterion-referenced (grades 3–8); high school exit exam (grade 10)	Norm/criterion-referenced (grades 3–8); writing assessment (grades 4, 7, 11); high school competency exam (grade 9)
What changes to testing program are planned?	Criterion-referenced tests (grades 1–8); end-of-course exams to replace high school graduation test	None planned	Criterion-referenced tests (grades 2–8); writing assessment (grades 4, 7, 10); subject area exams to replace high school exit exam	High school exit exam to be added in 2004; gateway exams to be added in grades 3, 5, and 8 by 2005	Criterion-referenced tests in science and social studies in grades 3–8 added by 2003; criterion-referenced tests to replace high school exit exam by 2003	End-of-course exams that will serve as gateway tests to replace high school competency exam in 2001

Impact of High Stakes Accountability

instructional relevance, and this may lead to diffuse rather than focused professional development.

Professional Development Policies

The professional development policies have little connection to accountability or testing policies in almost all of our sample states—there is no coherence between these policies. Further, few states made provisions for professional development concerning new accountability or testing policies to be offered to teachers. While five of six states, with Mississippi being the exception, provide some state financial support for professional development, the responsibility for financially supporting professional development rests primarily on the availability of federal and district funds. North Carolina is the only state that financially supports the majority of professional development provided within the state and its districts. (Table 2 provides an overview of the states' professional development policy elements.) In most cases, choices about professional development activities rest with either districts or schools. The decentralization of professional development requirements has great potential for significant variation in both the type of professional development offered to teachers and its connection to accountability and testing policies. Further, this decentralization places increased importance on district and school level capacity in selecting and offering high-quality professional development. The decentralization of professional development also means that there is very little information at the state level about the types of professional development offered, the quality of professional development, and the teachers who are receiving the assorted offerings.

High Stakes Accountability Policies

The states had varying degrees of experience with accountability policies (see table 3 for a comparison of state accountability policies). Some, such as Kentucky and North Carolina, had been holding schools accountable for a decade. Others, such as Georgia and Mississippi, had only recently passed accountability policies. Across these states, this variation persists in both the amount and type of accountability structure in place. For example, two of our states provide some form of reward and two states (but not exactly the same states) provide some form of sanction for school performance. Also, four out of our six states categorized performance levels in 1999–2000, but two of these states applied the label to schools and two of these labeled districts. At the time of

TABLE 2

Professional Development Policies

	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Who makes decisions about teacher PD?	State and districts	Schools and teachers	Districts based on state requirements	Schools	Districts and schools	Districts and schools
What types of financing are available for PD?	Federal, state, and district	Federal, state, and district	Federal and district	Federal and state	Federal, state, and district	Federal, state, and district
Was additional PD funding provided for accountability implementation?	No	Yes	No	Only for schools designated as low performing	Yes	No

NOTE.—PD = professional development.

TABLE 3

Accountability Policies

	Georgia	Kentucky	Mississippi	North Carolina	South Carolina	Tennessee
Dates of accountability legislation	2000	1990	1999	1993	1998	1992
Date of full accountability system implementation	2004	1994	2002	1998	2001	1997
Did the state label in 1999–2000?	No	Yes	Yes (districts only)	Yes	Yes (districts only)	No
Did the state give rewards in 1999–2000?	No	To schools and teachers	No	To teachers	No	Legislated but not funded
Did the state give sanctions in 1999–2000?	No	No	Loss of district accreditation for low student performance	Yes	No	No
Was additional financing attached to accountability implementation?	No	Yes	No	Yes	Yes	Yes

our study, the states had different approaches to the consequences associated with inadequate performance:

- Georgia and Tennessee had no consequences for the performance of schools or school districts on the states' assessments;
- South Carolina only labeled districts based on their performance;
- Mississippi both labeled and sanctioned districts but not schools;
- Kentucky labeled schools and provided rewards to teachers and schools;
- North Carolina had all the accountability elements (school labels, school sanctions, and rewards to teachers but not schools).

The differing structures in accountability policies may trigger different responses from districts, schools, and teachers, and, ultimately, different policy impacts. Additionally, not every state provided fiscal support for the implementation of their accountability policies. The absence of this support when combined with varying degrees of district-level capacity and school-level capacity for implementation may impact the extent to which the accountability systems can leverage effective reforms.

The variation in HSA policies found in the states provided two scenarios for analysis. First, the variation of elements used by the states allowed us to look for relationships between elements and/or combinations of elements and professional development that may be expected to affect student achievement. Because of the variation in HSA policy structure across the states, we could pose a series of questions that would help determine the importance or possible impact of accountability elements. For example:

- Does labeling impact professional development?
- Do rewards cause teachers to pursue more effective professional development?
- Do sanctions trigger district level responses?
- Do criterion-referenced tests lead teachers to pursue more coherent professional development or influence schools to support teachers' use of assessment data?

A second scenario for analysis that resulted from the variation in HSA across the states involved the cascading degree of strength across the accountability systems. In our sample of states, the strength of the systems ranged from no stakes at all (Georgia and Tennessee) to having implemented all the elements of HSA (North Carolina). In order to examine the relationship between the overall strength of the accountability system, professional development of teachers, and student achievement, we relied on the accountability index developed by Martin Carnoy and Susanna Loeb (2002). Their 0–5 scale indicates the strength of accountability pressure applied by states via testing and HSA requirements. In their index:

Impact of High Stakes Accountability

- States receive a 0 if they do not have statewide testing or do not set statewide standards for schools or districts.
- States receive a 1 if they require testing at the elementary and middle school levels, report the results, but have no rewards or sanctions attached.
- States receive a 2 if they have testing at the elementary and middle school levels with moderate rewards or sanctions attached *or* they have a high school exit exam.
- States receive a 3 if they have elementary and middle grades testing with moderate rewards or sanctions *and* they have a high school exit exam.
- States receive a 4 if they have testing and strong rewards or sanctions but no high school exit exam.
- States receive a 5 if they have testing in elementary and middle grades with strong rewards or sanctions attached *and* they have a high school exit exam.

The Carnoy and Loeb index provided a measure of the overall strength of each state’s accountability system that we used to determine the relationship between HSA and the six systemic responses described above (see table 4 for an overview of our states and their accountability index scores). We also ran separate HLM models to determine if four elements of HSA systems make an important difference: criterion-referenced tests, labels, rewards, or sanctions.

Study Data and Methods: Sample, Survey Measures, Analytic Techniques

The study and methods were developed to estimate the effects of HSA policies on teachers’ professional development activities and their attitudes about dis-

TABLE 4

Elements of High Stakes Accountability and State-by-State Index Values

	Labeling	Rewards	Sanctions	CRT	Accountability Index*
North Carolina	X	X	X	X	5.0
Kentucky	X	X		X	4.0
Mississippi	X (district)		X (district)	X	3.0
South Carolina	X (district)				3.0
Georgia					2.0
Tennessee				X	1.5

NOTE.—CRT = criterion-referenced test.

* Source: Carnoy and Loeb 2002.

district level and school level support for teaching and learning. We pursued a strategy that would maximize the likelihood of finding district effects, if these effects occurred, in the states chosen for the study. To foreshadow the detailed description that appears later in this section, our strategy involved purposeful selection of states and probability selection of districts and schools within the chosen states. States were selected to maximize the contrasts between the focal elements of the HSA systems, including testing students using criterion-referenced exams, labeling, sanctioning, and rewarding without creating confounds that could result from states having the same combinations of the HSA elements. Districts were placed into cells to increase the opportunities to find district effects, but the sample selected was a disproportionately stratified probability sample. Schools were selected from the districts to ensure that high, low, and adequate performing schools were included in the final sample. Finally, we attempted to obtain survey responses from all teachers in the schools that were selected. In this section, we describe the sample selection, measures, and survey procedures that were used to obtain estimates of the dependent variables for the study, that is, responses to HSA. In addition, we present the analytical procedures used to estimate the effects of HSA on teachers' behaviors and attitudes.

Sampling

The sampling procedures were designed to yield a probability sample of elementary school teachers from each of the six states included in the study. We used a probability sampling design with unequal probability of selection to oversample both high and low performing schools relative to those whose performance was considered average or meeting but not exceeding expectations in each of the six states. To minimize differences that could be expected to occur by level of school, and to focus on the level of schooling where testing policies were ubiquitously implemented at the time of our study, we limited the teachers surveyed to those working in public elementary schools. In the focal six states, there were 4,383 elementary schools operating in both 2000–2001 and 2001–2002, which were located in districts with at least two elementary schools. These schools were the population for the study. For this study, elementary schools were defined as offering grades K–5.

In addition to differences in the policy environments in the six states, we hypothesized that prior levels of school performance on state tests could affect responses to the accountability. Therefore, to obtain a sample that varied with respect to prior performance, we chose a stratified cluster sample approach. We divided the elementary schools into three groups: one high performing, one low performing, and one adequately performing. The distinctions were

Impact of High Stakes Accountability

based on the individual criteria used in each state or, when those criteria were not available, by designating the 10 percent of schools that performed best on statewide elementary assessments as high performing and the 5 percent of the schools that performed worst on the assessments as low performing. The choice of 10 percent for high performing and 5 percent for low performing was based on the approximate percentage of schools falling into each of those categories in states that labeled schools high or low performing. The latest available test scores were used for categorizing school performance, in most cases from the spring of 2001.

Schools were stratified to maximize the contrasts between districts in terms of district effectiveness. Our assumption was that districts with no low performing schools were the districts where positive district effects were most likely to be observed. Conversely, the districts where no schools were high performing were the districts where positive district effects were least likely to be observed. To achieve this contrast, we divided all the districts in the six states into four cells:

- Cell 1. Districts with both high and low performing schools;
- Cell 2. Districts with low performing but no high performing schools;
- Cell 3. Districts with high performing but no low performing schools;
- Cell 4. Districts with neither high performing nor low performing schools.

While cell 4 contained exclusively average performing schools, average performing schools were included in all cells. Average performance for the purpose of this study was considered to be the adequately performing schools according to each state's criteria. Then within each cell the schools were categorized by their level of performance.

In each state, 24 schools were selected for participation in the survey, six from each of the district cells. Within each of the four district cells, the allocations were for two high, two low, and two moderately performing schools in cell 1; for three low and three moderately performing schools in cell 2; for three high and three moderately performing schools in cell 3; and for six moderately performing schools in cell 4. The omissions of school performance types in cells 2–4 resulted from the absence of those types of schools in the districts within the cells. The districts in cells 2 and 3 were the districts in which we expected to maximize the likelihood of not observing and observing positive district effects, respectively. All responses were weighted in the analysis to adjust for the unequal probabilities of selection in the sampling design.

Measures Included in the Questionnaire

The questionnaire was designed to collect data from teachers about district responses to HSA and their participation in "high-quality" professional de-

velopment. The questionnaire was designed in a three-step process. First, existing literature was examined to find surveys and to identify items about professional development that had been found to relate to instructional practices or student academic performance. In particular, instruments developed for the evaluation of the Eisenhower Professional Development Program—Title 2 of the Elementary and Secondary Education Act and the Student and Staffing Survey, administered by the National Center for Education Statistics, provided several useful items.

In the second step of the process, information from qualitative case studies in the six states was received (see Berry et al. [2003] for a description of this research) and scrutinized to assist in prioritizing the items for the surveys and improve wording for ease of understanding and greater clarity. Finally, the items were formatted into a questionnaire appropriate for obtaining teacher responses.

Items about professional development activities that are expected to improve instruction and student achievement (referred to as “high quality professional development”) cover:

1. Time spent in professional development activities,
2. Reform versus traditional format of activity,¹
3. Duration of activity,
4. Collective participation by teachers within school,
5. Active learning opportunities,
6. Coherence,
7. Instructional resources provided to teachers,
8. Communication of content or instructional methods.

In addition, teachers were asked about how effective their professional development had been during the last year. Items were asked concerning all professional development activities and about the single activity in which they spent the most time undertaken between June 2001 and May 2002.² Other items included district support for teaching and learning; school-wide use of assessment information; school administration support for professional development; coherence of state, district, and school plans for improving instruction; opinions about the state accountability system; and many teacher characteristics.

Survey Administration

Following our sample design, we drew a sequential probability sample of schools within each state designed to yield a sample of 24 schools per state, or 144 overall. District superintendents of the sampled schools were contacted

Impact of High Stakes Accountability

by letter and asked to approve participation of the sampled school in the project. School principals at the selected schools were contacted and their cooperation requested. If either the superintendent or the principal declined to cooperate, another school district was randomly selected from the appropriate cell to replace it. At each school, the study team with the assistance of the principal identified a faculty member who served as a liaison for the study and assisted in obtaining responses from all regular classroom teachers.

Packets were mailed to the local liaison containing instructions for administering the surveys, questionnaires for teachers, a brief school information form, and a self-addressed, stamped envelope for returning the completed surveys. We obtained cooperation from 133 schools (92 percent), and 107 actually returned completed surveys from their teachers (74 percent).

Analysis

Because the teachers surveyed for this study were nested within schools, the analysis must account for the likelihood that the teachers' responses within a school were more highly correlated as a result of the nesting. Therefore, we used two-level, hierarchical linear models for the study: the first level was teachers and the second was schools. In the first level, several teacher characteristics were included: credential (regular), teaching experience (0–2 and 3–10 years), graduate degree, grade taught (K–2), race (African American and other minority), and graduate of school system where currently teaching. The second-level equations include two dichotomous control variables for location (urban and rural; suburban was the omitted or reference group) and four continuous control variables: number of students in school, percentage of students receiving free or reduced-price lunch, percentage mobility of student population, and ratio of students to teachers. The variables of interest for this study were included at the second level and included either the accountability index in the first specification or dichotomous variables that indicated the presence or absence of a specific element of the HSA system in the second specification.

The analysis described here was used to compare teachers' responses to numerous constructs that related to accountability and professional development activities across the six states, as well as to test the district theory of action. In the following sections we present only those findings related to teachers' engagement in and use of instructional innovation, school support for testing, school support for teaching and learning, and district support for teaching and learning in each of the six states and the district-related theory of action.

Findings: Teachers' Perceptions of Engagement in Instructional Innovation, School Support, and District Support

Teachers were asked about the extent to which teachers were engaged in improving instruction and innovation within their schools (see table 5). Teachers agreed that the teachers in their schools were engaged in instructional improvement and were encouraged to stretch and grow. Rather than being oppressed or pressured by the accountability system, teachers expressed a “can do” attitude on behalf of their colleagues. They also indicated that their colleagues were trying to improve their instruction. Notably, teachers in North Carolina were least likely to agree that they and their colleagues are engaged in improving instruction and instructional innovation, but their NAEP scores show the greatest improvement and the highest levels of educational achievement of these six states.

Schools were not frequently making more systematic efforts to use test scores, according to teachers in these six states (see table 6). Kentucky and North Carolina appeared to differ with respect to the use of test scores. Kentucky teachers reported more frequently discussing test scores and ways to improve them, while North Carolina teachers indicated that they met on these topics infrequently. Georgia teachers most frequently were provided with materials to improve students' test-taking skills.

Teachers indicated that their schools strongly supported teaching and learning (see table 7). They agreed that standards for both teaching and student learning were high in their school and were slightly less likely to agree that they were pressed to implement what they learned in their professional development. Across the states, the responses about school support were more consistent and higher than the responses about district support for teaching and learning.

Teachers viewed their district's support for teaching and learning less positively than they viewed their school's support, but the responses across states for district support were also much more varied (see table 8). The variation was best reflected in the responses to the district support for using student assessments to guide instructional improvement, with three states above average (Kentucky, Mississippi, and South Carolina) and three states below (North Carolina, Georgia, and Tennessee). It is somewhat confusing to find that North Carolina and Tennessee teachers indicated below average district support on every item and that these states had the most and least well developed accountability systems according to the index used by Carnoy and Loeb (2002). However, findings about systematic differences should be based on the analysis of the hierarchical linear models that follows.

TABLE 5

Teacher Engagement in Improving Instruction and Innovative Practices at Schools in Six Southern States

Engagement in Improving Instruction and Innovative Practices at School	Overall	North Carolina	Kentucky	Mississippi	South Carolina	Georgia	Tennessee
Teachers have a “can-do” attitude (1–4)*	3.27	3.07	3.23	3.26	3.34	3.37	3.29
Teachers are encouraged to stretch and grow (1–4)	3.37	3.20	3.32	3.40	3.42	3.46	3.34
Teachers are continuously learning and seeking new ideas (1–4)*	3.28	3.05	3.20	3.33	3.41	3.35	3.23
Teachers participate in important educational decisions (1–4)	2.95	2.75	3.00	2.96	3.03	2.94	2.95
Teachers are willing to take risks to make the school better (1–5)*	3.42	3.18	3.40	3.48	3.55	3.38	3.43
Teachers are eager to try new ideas (1–5)*	3.52	3.22	3.45	3.62	3.63	3.52	3.54
Teachers are really trying to improve their teaching (1–5)*	3.71	3.45	3.67	3.78	3.80	3.71	3.75
Teacher engagement in improving instruction and innovative practices index (1–4)	3.44	3.20	3.39	3.49	3.54	3.47	3.45

NOTE.—Bold indicates that the state’s mean is significantly different than overall mean ($p < .05$).

* Item is included in a teacher engagement in improving instruction and innovative practices index ($\alpha = .87$).

TABLE 6

School Administration Support for Test Utilization in Six Southern States

School Support for Test Utilization (1–5)	Overall	North Carolina	Kentucky	Mississippi	South Carolina	Georgia	Tennessee
Holds staff meetings to review test scores*	2.99	2.85	3.21	3.17	2.91	2.74	3.09
Discusses ways to improve test scores*	3.40	3.24	3.60	3.48	3.31	3.26	3.50
Discusses ways to strengthen instruction in areas where test scores show weakness*	3.34	3.17	3.56	3.38	3.23	3.24	3.47
Provides students materials for test-taking skills*	3.09	3.05	2.98	3.07	2.96	3.23	3.21
Provides special assistance to teachers to improve test scores*	2.75	2.77	2.69	2.85	2.69	2.67	2.84
Checks that teachers emphasize skills for which past test results showed weakness*	3.00	2.86	3.03	3.19	2.88	2.94	3.05
School administration support for test utilization index	3.09	2.99	3.18	3.19	3.00	3.02	3.18

NOTE.—Bold indicates that the state's mean is significantly different than overall mean ($p < .05$).

* Item is included in a school administration support for test utilization ($\alpha = .92$).

TABLE 7

School Administration Support for Teaching and Learning in Six Southern States

School Support for Teaching and Learning (1–4)	Overall	North Carolina	Kentucky	Mississippi	South Carolina	Georgia	Tennessee
Carefully tracks student academic progress*	3.35	3.30	3.36	3.34	3.37	3.36	3.36
Understands how children learn*	3.38	3.29	3.36	3.38	3.35	3.50	3.36
Presses teachers to implement what they learned in professional development*	3.28	3.15	3.24	3.29	3.30	3.33	3.32
Communicates a clear vision for the school*	3.44	3.37	3.38	3.42	3.48	3.52	3.42
Sets high standards for student learning*	3.54	3.48	3.52	3.53	3.57	3.61	3.48
Sets high standards for teaching*	3.52	3.43	3.50	3.55	3.57	3.59	3.44
Makes clear expectations for meeting instruc- tional goals*	3.44	3.39	3.39	3.49	3.46	3.50	3.40
School administration support for teaching and learning index	3.42	3.34	3.39	3.43	3.44	3.49	3.40

NOTE.—Bold indicates that the state's mean is significantly different than overall mean ($p < .05$).

* Item is included in a school administration support for teaching and learning index ($\alpha = .93$).

TABLE 8

District Support for Teaching and Learning in Six Southern States

District Support for Teaching and Learning (1–4)	Overall	North Carolina	Kentucky	Mississippi	South Carolina	Georgia	Tennessee
Helps schools develop and maintain high standards*	3.11	2.95	3.15	3.20	3.23	3.08	2.94
Helps principals promote and nurture a focus on teaching and learning*	3.06	2.92	3.09	3.22	3.12	3.03	2.90
Provides support to enable teachers to adjust the curriculum and instruction to meet students' individual needs*	2.92	2.82	2.97	3.09	2.94	2.88	2.79
Helps schools use information about student achievement to improve instruction*	3.12	3.02	3.23	3.23	3.18	3.01	3.02
Helps schools set benchmarks and evaluate progress toward school and district standards*	3.25	3.13	3.28	3.44	3.35	3.13	3.08
Helps establish systems of governance and decision making that include participation by key stakeholder groups*	2.99	2.87	3.13	3.02	3.07	2.96	2.81
Helps maintain open communication with and public accountings to key stakeholders regarding student and school performance*	3.01	2.88	3.06	3.07	3.15	2.97	2.87
Promotes teacher leadership across school districts*	2.94	2.78	3.00	3.03	3.10	2.88	2.78
District support for teaching and learning index	3.05	2.92	3.11	3.16	3.14	2.99	2.90

NOTE.—Bold indicates that the state's mean is significantly different than overall mean ($p < .05$).

* Item is included in a district support for teaching and learning index ($\alpha = .93$).

Impact of High Stakes Accountability

TABLE 9

Effects of High Stakes Accountability on Teacher, School, and District Actions

Teacher, School, and District Actions	Accountability Index	Rewards	Labels	Sanctions	CRT
Teacher engagement in improving instruction and innovative practices	-.04 -1.35	-.13 -1.60	.22 2.79	-.20 -2.70	-.19 -2.57
School administration support for teaching and learning index	.02 .80	.06 .81	-.07 -.86	.10 1.62	.03 .40
School administration support for test utilization index	.05 1.21	.18 1.42	-.01 -.69	.00 .03	-.03 -.22
District support for teaching and learning index	.05 2.13	-.04 -.54	.24 3.59	-.05 -.59	-.06 -.93

NOTE.—Bold indicates that coefficient is significantly different than zero ($p < .05$). Values are unstandardized regression coefficients and t -values.

Findings: Hierarchical Linear Models Analysis of District Responses to High Stakes Accountability

In this section of the article, we present the results of the HLM of the responses to HSA. We examine the impacts of HSA in two ways: (1) the relationship between an accountability index (Carnoy and Loeb 2002) and the average response of teachers within a school, controlling for the teacher and school characteristics listed below, is tested, and (2) the relationship between each of four important components of HSA—criterion-referenced tests, rewards, labels, and sanctions—and the systemic response in each school is tested in another model. Table 9 presents the coefficients from the second-level models that estimate the effects of the index and each of the components on the dependent variable. The index is tested in a separate equation from the four components, but the four components are tested in one model. The first-level model includes eight dichotomous independent variables measuring individual teacher characteristics: credential (regular), teaching experience (0–2 and 3–10 years), graduate degree, grade taught (K–3), race (African American and other minority), and graduate of school system where currently teaching. The second-level models include two dichotomous variables indicating system location (urban and rural; suburban is reference group) and four continuous variables: number of students in school, percentage of students receiving free or reduced-price lunch, percentage mobility of student population, and ratio of students to teachers. The coefficients of interest for this study represent the extent to

which the accountability index or component is systematically related to the dependent variable, including teachers' engagement in improving instruction and innovation, school support, and district support, after controlling for the teacher and school characteristics listed above.

We assess the relationship of HSA on four dependent variables: teacher engagement in improving instruction and innovative teaching practices, school administration support for teaching and learning, school administration support for using test information, and district administration support for teaching and learning. All of the dependent variables are composites of the variables presented in tables 5–8 and each is listed as the last variable in these tables.

More developed accountability systems were associated with a systemic response in one of the ways anticipated by standards-based reformers—district support for teaching and learning. However, it does not appear that the level of development of HSA within a state has an effect on teaching or school support for teaching and learning or using assessment data. Consistent with previous case study research, districts appear to act as intermediaries of external policies (Craciun and Snow-Renner 2003; Massell 2000; Skrla et al. 2001; Spillane and Thompson 1997) and therefore may play an essential role in reform success (Elmore and Burney 1998a; McLaughlin and Talbert 2003).

According to prior research, district action can be triggered by various elements of HSA policies (Geortz et al. 1998). However, only labeling was positively and significantly related to district support for teaching and instruction. Sanctions, rewards, and use of criterion-referenced tests were negatively and insignificantly related to teachers' perceptions of more district involvement. This finding appears to be consistent with Meier et al. (2000), who find that districts responded to low test scores in Florida. This finding is also consistent with the work of Mayo and McIntyre (2003) on how superintendents spend their time in high stakes and low stakes testing environments. It appears that more highly developed HSA systems trigger districts and their leaders to concentrate on teaching and learning to a greater extent than they would have in the absence of these systems.

The finding concerning labeling contrasts with Firestone and Fariman (1998), who found that stronger consequences were necessary to motivate districts to become actively engaged in improving teaching and learning. In addition, in our study, labeling poor-performing districts or schools was positively related to teacher engagement in improving instruction and innovation. On the other hand, sanctions and using criterion-referenced tests appear to lower teacher engagement in improvement. It is noteworthy that teachers' perceptions of actions undertaken by school administrators do not vary systematically with more developed accountability systems or any of the specific elements of these systems. This suggests that schools are not responding to the accountability systems or are unable to respond without district actions.

Impact of High Stakes Accountability

The variation in impact of accountability elements in our study is predicted by those who have studied the effects of policy instruments (e.g., Bemelmans-Videc et al. 1998; McDonnell and Elmore 1987). Research on the impact of policy instruments has consistently shown, for example, that sanctions create adversarial relationships between actors at various levels in the implementation system. It is therefore not surprising that the presence of sanctions and labels, while inducing school districts to become engaged, may cause teachers to view their districts less favorably and lower their engagement in instructional improvement and comfort with innovation. Further, the research on policy instruments has also indicated that the use of rewards may exaggerate differences between the groups implementing the policy rather than narrow variation in outcome.

Conclusions

As the new century began, the states in the South, along with the rest of the nation, were reforming their education systems by raising the stakes associated with performance on statewide assessments. Southern states lagged behind most of the nation in student achievement, and most avidly pursued HSA as a means to gain ground. The six states included in this study had implemented some components of the system by 1999–2000, but the implementation was uneven. For example, North Carolina and Kentucky had led the nation in the development of HSA systems, while Georgia had not yet implemented a major component, other than a weak high school graduation exam. The differences in policies and their implementation presented an opportunity to examine how teachers and districts were responding to these new systems.

Understanding the responses to HSA is essential to explaining the association between these systems and improved test scores. This study points to one potential systemic response that seems to be triggered by more developed systems—districts become more engaged with teaching and learning at the school level. The findings from this study support the previous case study research on districts by showing a systematic district effect that results from the implementation of HSA systems. While a district response is consistent with other research, especially in-depth study of the changes in Texas, the study has its own internal anomaly—districts in North Carolina, the state with the most fully developed system and greatest increase in student achievement, registered lower marks in district involvement. This district response anomaly may have many explanations, among them:

- District involvement was uneven in North Carolina, and it occurred only in certain districts (i.e., those with the lowest performing schools, those

with leaders who felt pressure to improve performance, those who felt they had a good chance of leveraging significant improvement, etc.).

- District engagement occurs systematically but is unrelated to improving student achievement.

Thus, while this study indicates a district effect, more pieces of the puzzle must be put together in future research to determine when district engagement occurs, why it occurs, and the impacts it has on schools, teachers, instruction, and student achievement.

At first glance, our findings indicate that accountability systems result in sometimes contradictory systemic responses. For example, districts become engaged in instructional improvement efforts as a result of accountability systems. However, in the most mature system (North Carolina), districts were the least engaged. Beyond the two possible explanations given above, previous research provides some direction for understanding this anomaly. One possible explanation for these findings is that they may illustrate a cycle, and possibly unintended consequence, of accountability systems that emerge over time—initial resistance, then acceptance, comfort, and ultimately complacency. When accountability systems are first introduced, professional development may be positively impacted and districts may engage in helping schools improve. As teachers live with the system they may come to accommodate its components into their own schema and accept it as valid. However, as this acceptance occurs, districts and teachers might become complacent. Districts may disengage from improvement and teachers may attend less to the system specifics and plan less for professional development integration.

High stakes accountability systems have not had sufficient longevity for us to determine the cycle (if any) that these policies undergo and the differential impacts of these policies over time. Mazmanian and Sabatier's (1981, 1989) work on implementation would indicate that we should expect some cycle of implementation attention and varying impact over the life of these policies. Exactly what this implementation cycle looks like, how districts and teachers accommodate these policies, and how these policies vary in impact over time is unclear and in need of further study.

Another possible cause for the irregularity in the North Carolina findings may be related to the impact of policy instruments. It has previously been stated that sanctions and rewards impact policy implementers differentially (Bemelmans-Videc et al. 1998; McDonnell and Elmore 1987). While this literature provides some guidance in predicting the impact of rewards and sanctions when used in isolation, it provides no guidance on how these instruments work in combination. And, in our study, North Carolina was the only state to have implemented all HSA instruments available. Thus, further research is needed to understand how policy instruments work together to impact implementation and the actors involved in the process.

Impact of High Stakes Accountability

Additional explanations for the North Carolina result come from the research on the role of context in the emerging district effect literature. Specifically, there is some indication that district size is an important element in the development of coherence within districts. Counterintuitively, smaller districts may struggle with the development of coherent instructional policies because of a lack of capacity, access to technical assistance, and lower levels of program understanding (Hannaway and Kimball 2001). In our study, we included variables that accounted for school size but not district size. The relatively large number of small districts in North Carolina may account for our discrepant findings. Supporting this as a possible explanation is our finding that North Carolina teachers had lower levels of understanding of their state's accountability system than other teachers in our study. The role of district size in the development of coherence, and the district effect more generally, would be an important direction for future research.

Finally, there is also some evidence to suggest that when districts are effective at influencing teaching and learning, high levels of internal accountability are present. That is, those districts that have been identified in the case literature as effective at improving student achievement within systems of state accountability also have accountability that is self-enforced as part of the district culture (Firestone et al. 2005; Newmann et al. 1997; Skrla et al. 2001). As Elmore (2003) has claimed, "high internal accountability is a necessary precondition . . . to be successful in responding to the pressures of external accountability systems" (198). High levels of internal accountability, either initially or that develops as a consequence of a mature accountability system, may reduce the need for more overt actions on the part of districts. Thus, teachers may not perceive the district's support when high internal accountability exists.

More research also remains to be done to understand if, when, and how HSA affects student learning. The unprecedented federal legislation, No Child Left Behind, has increased the energy and anxiety focused on HSA. While the state initiatives have largely been overcome by federal requirements, research on state level effects will continue to have import for improving our understanding of the relationship between student achievement and the policies that are enacted to increase it. Accountability systems do not appear to improve the quality of instruction and related student outcomes directly. What accountability may do is to initiate a chain of events that may lead to improved instruction and student performance. It is clear from our study that the role of districts is a key link in this chain and deserves further attention.

Notes

This study was funded by the Spencer Foundation (grant no. 2001000295) and conducted at the Andrew Young School of Policy Studies at Georgia State University. The authors wish to acknowledge the support of John Williams and the contributions of Brandee Krug, Andrew J. Mashburn, Christina Stephens, and Marguerite Bryan. Special thanks are extended to Barnett Berry, Steven Clements, Dwight Hare, Charles Thompson, Alan Toms, and the other members of the UNC team.

1. Reform type activities tend to be school-based and include study groups, teacher networks, mentoring, teacher research, and so on in contrast to traditional activities that include workshops, courses, and conferences.

2. It is important to understand the timing of implementation of the accountability system and data collections. The attributes of the HSA systems were taken from the 1999–2000 school year. In most cases those results are made available after the beginning of the following school year, when most plans for professional development and other relevant school and district plans have been set for the year, in this case 2000–2001. Therefore, teachers were surveyed about their activities in 2001–2002, when the results of the HSA system would be able to have been incorporated into their activities.

References

- Allington, Richard, and Peter Johnson. 1989. "Coordination, Collaboration, and Consistency: The Redesign of Compensatory and Special Education Interventions." In *Effective Programs for Students at Risk*, ed. Robert Slavin, Nancy Karweit, and Nancy Madden. Needham Heights, MA: Allyn & Bacon.
- Bemelmans-Videc, Marie Louise, Ray Rist, and Evert Vedung. 1998. *Carrots, Sticks, and Sermons: Policy Instruments and Their Evaluation*. New Brunswick, NJ: Transaction.
- Berman, Paul, and Milbrey McLaughlin. 1978. *Federal Programs Supporting Educational Change*. Vol. 3 of *Implementing and Sustaining Innovations*. Report No. R-1589/8-HEW. Santa Monica, CA: Rand.
- Berry, Barnett, Laura Turchi, Dylan Johnson, Dwight Hare, Deborah Owens, and Steve Clements. 2003. "The Impact of High Stakes Accountability on Teachers' Professional Development: Evidence from the South. A Final Report to the Spencer Foundation." Unpublished manuscript, http://www.teachingquality.org/pdfs/Spencer_FinalReport.pdf.
- Birman, Beatrice, Laura Desimone, Andrew Porter, and Michael Garet. 2000. "Designing Professional Development That Works." *Educational Leadership* 57 (8): 28–33.
- Bryk, Anthony, Valerie Lee, and Peter Holland. 1993. *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.
- Bryk, Anthony, Valerie Lee, and Julia Smith. 1990. "High School Organization and Its Effects on Teachers and Students: An Interpretive Summary of the Research." In *Choice and Control in American Education*, vol. 1, ed. William H. Clune and John F. Witte. New York: Falmer.
- Bryk, Anthony, Penny Sebring, David Kerbow, Sharon Rollow, and John Easton. 1998. *Charting Chicago School Reform*. Boulder, CO: Westview.
- Carnoy, Martin, and Susanna Loeb. 2002. "Does External Accountability Affect Stu-

Impact of High Stakes Accountability

- dent Achievement? A Cross-State Analysis." *Educational Evaluation and Policy Analysis* 24 (4): 287–301.
- Cohen, David, and Deborah Ball. 1996. "Reform by the Book: What Is—or Might Be—the Role of Curriculum Materials in Teacher Learning and Instructional Reform?" *Educational Researcher* 25 (9): 6–8.
- Cohen, David, and Heather Hill. 2001. *Learning Policy: When State Education Reform Works*. New Haven, CT: Yale University Press.
- Coleman, James, Thomas Hoffer, and Sally Kilgore. 1982. *High School Achievement: Public, Catholic, and Private Schools Compared*. New York: Basic.
- Cooke, William. 1985. *Union Organizing and Public Policy: Failure to Secure First Contracts*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.
- Corwin, Ronald, and Kathryn Borman. 1988. "School as Workplace: Structural Constraints on Administration." In *Handbook of Research on Educational Administration*, ed. Norman J. Boyan. New York: Longman.
- Craciun, Kirstin, and Ravay Snow-Renner. 2003. *No Child Left Behind Policy Brief: Low-Performing Schools*. Denver: Education Commission of the States.
- David, Jane. 1990. "Restructuring in Progress: Lessons from Pioneering Districts." In *Restructuring Schools: The Next Generation of Educational Reform*, ed. Richard Elmore and Associates. San Francisco: Jossey-Bass.
- Debray, Elizabeth, Gail Parson, and Salvador Avila. 2003. "Internal Alignment and External Pressure: High School Responses in Four State Contexts." In *The New Accountability: High Schools and High Stakes Testing*, ed. Martin Carnoy, Richard Elmore, and Leslie Siskin. New York: Routledge Falmer.
- Desimone, Laura, Andrew Porter, Michael Garet, Kwang Yoon, and Beatrice Birman. 2002. "Effects of Professional Development on Teachers' Instruction: Results from a Three-Year Longitudinal Study." *Educational Evaluation and Policy Analysis* 24 (2): 81–112.
- DiMaggio, Paul, and Walter Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147–60.
- Elmore, Richard. 1993. "The Development and Implementation of Large-Scale Curriculum Reforms." Paper presented at the American Association for the Advancement of Science, Washington, DC, February.
- Elmore, Richard. 1996. "Getting to Scale with Good Educational Practice." *Harvard Educational Review* 66 (1): 1–17.
- Elmore, Richard. 2003. "Accountability and Capacity." In *The New Accountability: High Schools and High Stakes Testing*, ed. Martin Carnoy, Richard Elmore, and Leslie Siskin. New York: Routledge Falmer.
- Elmore, Richard, and Deanna Burney. 1998a. *Continuous Improvement in Community District #2, New York City*. Pittsburgh: Learning Research and Development Center, High Performance Learning Communities Project, University of Pittsburgh.
- Elmore, Richard, and Deanna Burney. 1998b. *School Variation and Systemic Instructional Improvement in Community District #2, New York City*. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.
- Elmore, Richard, and Deanna Burney. 1999. "Investing in Teacher Learning: Staff Development and Instructional Improvement." In *Teaching as the Learning Profession: Handbook of Policy and Practice*, ed. Linda Darling-Hammond and Gary Sykes. San Francisco: Jossey-Bass.
- Elmore, Richard, and Susan Fuhrman. 1994. "Governing Curriculum: Changing Patterns in Policy, Politics, and Practice." In *The Governance of Curriculum: 1994 Yearbook of the Association for Supervision and Curriculum Development*, ed. Richard Elmore and

Opfer, Henry, and Mashburn

- Susan Fuhrman. Alexandria, VA: Association for Supervision and Curriculum Development.
- Firestone, William, and Janet Fariman. 1998. "The District Role in State Assessment Policy: An Exploratory Study." Paper presented at the annual meeting of the American Educational Research Association, San Diego, April.
- Firestone, William A., Melinda Mangin, Cecilia Martinez, and Terris Polovsky. 2005. "Leading Coherent Professional Development: A Comparison of Three Districts." *Educational Administration Quarterly* 41 (3): 413–48.
- Floden, Robert, Andrew Porter, Linda Alford, Donald Freeman, Susan Irwin, William Schmidt, and John Schwille. 1988. "Instructional Leadership at the District Level: A Closer Look at Autonomy and Control." *Educational Administration Quarterly* 24 (2): 96–124.
- Fuhrman, Susan. 1993. *Designing Coherent Education Policy: Improving the System*. San Francisco: Jossey-Bass.
- Gamoran, Adam, Charles Anderson, Paula Quiroz, Walter Secada, Tona Williams, and Scott Ashmann. 2003. *Transforming Teaching in Math and Science: How Schools and Districts Can Support Change*. New York: Teachers College Press.
- Garet, Michael, Andrew Porter, Laura Desimone, Beatrice Birman, and Kwang Yoon. 2001. "What Makes Professional Development Effective? Results from a National Sample of Teachers." *American Educational Research Journal* 38 (4): 915–45.
- Geortz, Margaret, Diane Massell, and Tammi Chun. 1998. "District Response to State Accountability Systems." Paper presented at the annual meeting of the Association for Public Policy and Management, Washington, DC, March.
- Grissmer, David, and Ann Flannagan. 1998. *Exploring Rapid Achievement Gains in North Carolina and Texas*. Washington, DC: National Education Goals Panel.
- Hannaway, Jane, and Kristi Kimball. 2001. "Big Isn't Always Bad: School District Size, Poverty, and Standards-Based Reform." In *From the Capitol to the Classroom: Standards-Based Reform in the States*, ed. Susan Furham. Chicago: University of Chicago Press.
- Hanushek, Eric. 1994. *Making Schools Work: Improving Performance and Controlling Costs*. Washington, DC: Brookings Institution.
- Hernandez, Cynthia. 2003. "Eliminating the Achievement Gap: The Study of One Texas School District." PhD diss., University of Texas at Austin, College of Education.
- Hightower, Amy. 2002. "San Diego's Big Boom: Systemic Instructional Change in the Central Office and Schools." In *School Districts and Instructional Renewal*, ed. Amy Hightower, Michael Knapp, Julie Marsh, and Milbrey McLaughlin. New York: Teachers College Press.
- Hill, Paul. 1995. *Reinventing Public Education*. Santa Monica, CA: Rand.
- Hill, Paul, and Mary Celio. 1998. *Fixing Urban Schools*. Washington, DC: Brookings Institution.
- Hirschman, Charles, and Kim Blankenship. 1981. "The North-South Earnings Gap: Changes during the 1960's and 1970's." *American Journal of Sociology* 87 (2): 388–402.
- Manpower Development Research Corporation. 2002. *Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement*. Washington, DC: Council of Great City Schools.
- Massell, Diane. 2000. "The District Role in Building Capacity: Four Strategies." CPRE Policy Briefs, RB-32. Consortium for Policy Research in Education, Philadelphia.
- Mayo, C. Russell, and Lisa McIntyre. 2003. "Superintendent Time on Task: High Stake Tests vs. Low Stake Tests." Paper presented at the annual meeting of the University Council for Educational Administration, Portland, OR, November.

Impact of High Stakes Accountability

- Mazmanian, Daniel, and Paul Sabatier. 1989. *Implementation and Public Policy: With a New Postscript*. Lanham, MD: University Press of America.
- Mazmanian, Daniel, and Paul Sabatier. 1981. *Effective Policy Implementation*. Lexington, MA: Lexington Books.
- McDonnell, Lorraine, and Richard Elmore. 1987. "Getting the Job Done: Alternative Policy Instruments." *Educational Evaluation and Policy Analysis* 9 (2): 133–52.
- McKenzie Group. 1999. *Student Achievement and Reform Trends in 13 Urban Districts*. McKenzie Group, Washington, DC (Eric Document Reproduction Service No. ED 441 016).
- McLaughlin, Milbrey, and Joan Talbert. 2003. "Reforming Districts: How Districts Support Reform." September. Center for the Study of Teaching and Policy, University of Washington, Seattle.
- Meier, Kenneth, J. L. Polinard, and Robert Wrinkle. 2000. "Bureaucracy and Organizational Performance: Causality Arguments in Public Schools." *American Journal of Political Science* 44 (3): 590–602.
- Meyer, John, and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83 (2): 340–63.
- Murphy, Joseph, and Philip Hallinger. 1988. "Characteristics of Instructionally Effective School Districts." *Journal of Educational Research* 81 (3): 175–81.
- Newmann, Fred, Bruce King, and Mark Rigdon. 1997. "Accountability and School Performance: Implications for Restructuring Schools." *Harvard Educational Review* 67 (1): 41–74.
- Newmann, Fred, and Gary Wehlage. 1995. *Successful School Restructuring: A Report to the Public and Educators*. Madison: Wisconsin Center on Education Research.
- O'Day, Jennifer. 2002. "Complexity, Accountability, and School Improvement." *Harvard Educational Review* 72 (3): 293–329.
- Raymond, Margaret, and Eric Hanushek. 2003. "Shopping for Evidence Against School Accountability." *Education Next* 3:1–13.
- Rowan, Brian. 1990. "Commitment and Control: Alternative Strategies for the Organizational Design of Schools." *Review of Research in Education* 16:353–89.
- Rowan, Brian, and Cecil Miskel. 1999. "Institutional Theory and the Study of Educational Organizations." In *Handbook of Research in Educational Administration*, ed. Joseph Murphy and Karen Louis. San Francisco: Jossey-Bass.
- Scott, W. Richard. 1995. *Institutions and Organization*. Thousand Oaks, CA: Sage.
- Skrla, Linda, James Scheurich, Joseph Johnson, and James Koschoreck. 2001. "Accountability for Equity: Can State Policy Leverage Social Justice?" *International Journal in Educational Leadership* 4 (3): 237–60.
- Smith, Marshall, and Jennifer O'Day. 1990. "Systemic School Reform." In *The Politics of Curriculum and Testing*, ed. Susan Fuhrman and Betty Malen. New York: Routledge.
- Smylie, Mark, Diane Bilcer, Julie Kochanek, Karin Sconzert, Dorothy Shipps, and Holly Swyers. 1998. *Getting Started: A First Look at Chicago Annenberg Schools and Networks*. Chicago: Consortium on Chicago School Research, Chicago Annenberg Research Project.
- Spillane, James, and Charles Thompson. 1997. "Reconstructing Conceptions of Local Capacity: The Local Education Agency's Capacity for Ambitious Instructional Reform." *Educational Evaluation and Policy Analysis* 19 (2): 185–203.
- Stamas, George. 1981. "The Puzzling Lag in Southern Earnings." *Monthly Labor Review* 104 (6): 27–36.
- Togneri, Wendy. 2003. *Beyond Islands of Excellence: What Districts Can Do to Improve Instruction and Achievement in All Schools*. Washington, DC: Learning First Alliance.