

# WORKING P A P E R

---

## Introduction to First-Year Findings from the Implementing Standards-Based Accountability (ISBA) Project

BRIAN STECHER, LAURA HAMILTON, AND  
SCOTT NAFTEL

WR-255-EDU

April, 2005

This product is part of the RAND Education working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Education but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

### **Preface**

This paper was presented as part of the Symposium “Implementing Standards-Based Accountability: Results from Classrooms, Schools and Districts in Three States” at the annual meeting of the American Educational Research Association, Montreal, Canada, April 13, 2005.

This material is based on work supported by the National Science Foundation under Grant No. REC-0228295. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

### **Purpose of Project**

Perhaps the largest scale-up of an educational program occurring in the United States at the present time is the implementation of standards-based accountability (SBA) in all 50 states as a result of the No Child Left Behind Act of 2001 (P.L. 107-110). The Implementing Standards-Based Accountability (ISBA) project is designed to identify factors that enhance the implementation of SBA systems, foster changes in school and classroom practice, and promote improved student achievement. We are following selected states, districts, schools, and teachers longitudinally for three years to address four broad questions:

- What strategies are used in implementing SBA at the state, district, and school levels?
- Which state, district, and school implementation strategies are associated with changes in classroom practice?
- Which features of SBA implementation are associated with student attainment of academic standards?
- How valid are the standards-based test scores as indicators of changes in student achievement?

Answers to these questions will help policymakers, educators, and researchers understand the ways SBA policies are implemented at all levels of the system; explain relationships among implementation, instructional practices, and student achievement; and identify successful practices that can be promulgated to make SBA function more effectively. This paper provides a description of the study and presents some contextual information on the three states—California, Georgia, and Pennsylvania—that are participating in the study. It is the first paper in a five-paper symposium that presents findings from the initial year of data collection.

## **Background**

### **Standards-Based Accountability**

Standards-based accountability is the amalgamation of three ideas--standards, assessments, and accountability--that have been part of the educational landscape for some time. For example, the call for explicit “world-class” standards to guide educational practice was sounded loudly by the nation’s governors in 1989 in response to the poor performance of U.S. students on international comparative assessments (McKnight, et al., 1987; National Governors Association, 1989; Travers and Westbury, 1989). After much study and discussion (National Council on Education Standards and Testing, 1992), content standards became a formal requirement of the 1994 reauthorization of the Elementary and Secondary Education Act (The Improving America’s Schools Act). Standardized achievement testing is a far older idea. Standardized tests began to be used in the U.S. in the early twentieth century, and they have been the most common method for monitoring educational performance for decades (Resnick, 1982). Accountability for outcomes in education can be traced back more than a century to early pay-for-performance agreements in England (Kirst, 1990). The notion of accountability has taken a number of forms in the U.S. in recent years, including school report cards and school choice. Yet, the combination of these elements into a single integrated system is relatively new, as is the use of the term “standards-based accountability.”

Standards-based accountability is built around a multi-step feedback mechanism. Content and performance standards define the goals for the educational system. Schools are expected to use these statements about “what students should know and should be able to do” to establish policies regarding curriculum, professional development, and other school functions, and teachers are expected to use them to guide instructional planning. In this way, the

coordinated efforts of administrators and teachers should lead to students' mastery of the content and skills specified in the standards. The standards also guide the development of student assessments. Student performance on the assessments is used as an indicator of school success. Schools that do well are rewarded, which should reinforce good practice. Schools that do poorly are sanctioned and offered assistance, which should lead them to change practice and improve their services to students. This feedback loop is intended to improve educational practices and student outcomes. These few basic components are at the heart of SBA, although they can be operationalized in different ways by choosing different standards, tests, assistance policies, and reward structures.

**SBA in the Context of No Child Left Behind.** Most educators are familiar with the accountability provisions of the 2001 reauthorization of the Elementary and Secondary Education Act, also known as the No Child Left Behind Act (NCLB).<sup>1</sup> The law requires that states establish accountability systems that include: academic standards (indicating what students should know and be able to do); assessments (aligned with standards); performance standards (indicating the level of test performance that corresponds to “proficient” and other levels); annual measurable objectives (indicating the percentage of students expected to be proficient each year until all are proficient in 2012); adequate yearly progress (AYP) targets for schools; and rewards and sanctions for schools based on their AYP status. It is easy to become lost in the details of the law, and it is worth remembering that the law emphasizes four key principles: high standards for all, accountability, flexibility, and parental choice. The accountability policies are designed to ensure that all students, not just some, are expected to become proficient. (Thus, equity is also a key principle of NCLB.) On inspection it is clear that the accountability model

---

<sup>1</sup> A more detailed description of accountability under NCLB can be found in Stecher, Hamilton and Gonzalez (2003).

enacted by NCLB is a form of standards-based accountability. The SBA core of the NCLB accountability model is described succinctly in this discussion by the National Research Council (1999):

The centerpiece of the system is a set of challenging standards. By setting these standards for all students, states would hold high expectations for performance; these expectations would be the same regardless of students' backgrounds or where they attended school. Aligned assessments to the standards would allow students, parents, and teachers to monitor student performance against the standards. Providing flexibility to schools would permit them to make the instructional and structural changes needed for their students to reach the standards. And holding schools accountable for meeting the standards would create incentives to redesign instruction toward the standards and provide appropriate assistance to schools that need extra help (National Research Council, 1999, pp. 2-3).

This description summarizes the beliefs of the framers of NCLB that this manifestation of standards-based accountability would lead to improvements in the education of all students. Yet, empirical research on SBA suggests that this approach does not assure success.

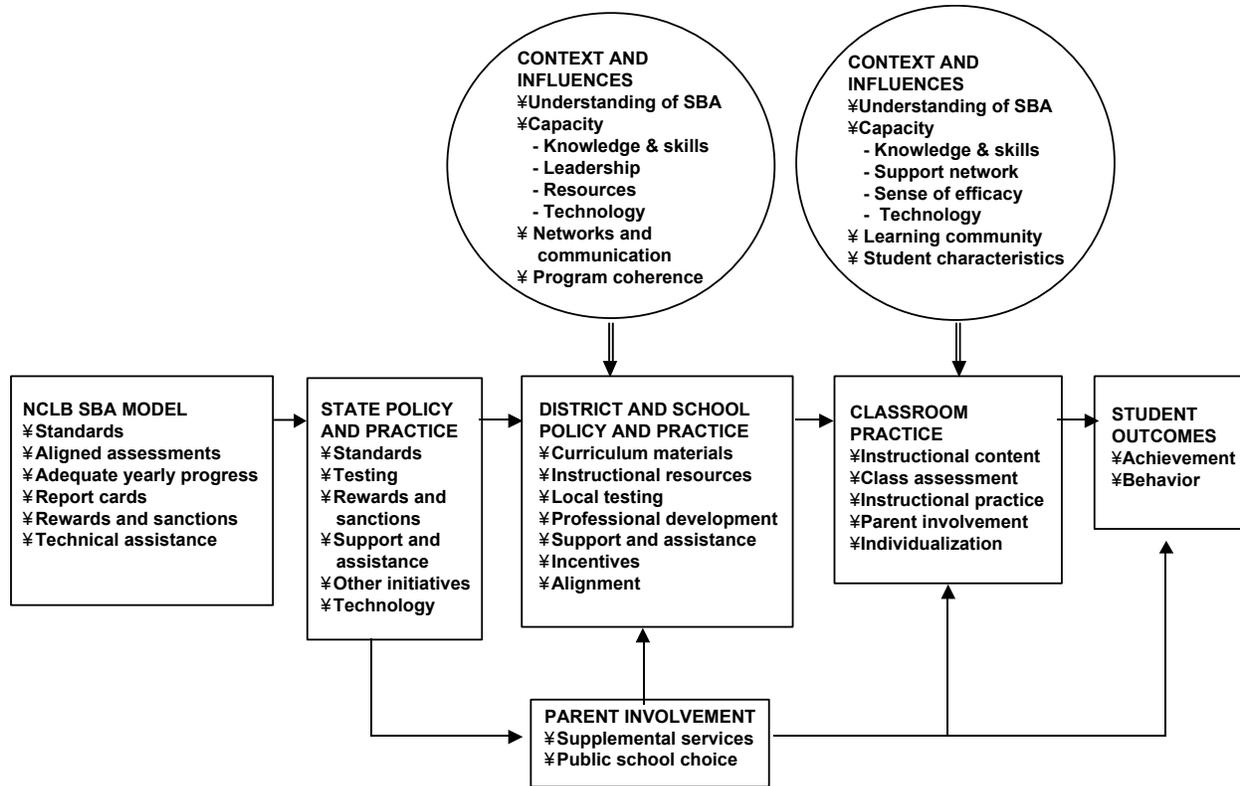
**Evidence for the Effectiveness of Standards-Based Accountability.** The utility of standards-based accountability was demonstrated in a few pioneering states in the 1990s. Two of the most prominent examples of SBA occurred in Texas and North Carolina, where test scores rose dramatically after the introduction of SBA systems (Grissmer & Flanagan, 1998; Texas Education Agency, 2000). Improved performance of NAEP has been interpreted by some observers as providing confirmatory evidence of the success of SBA policies in these states (Grissmer et al., 2000), although the increases in NAEP have not been as large as might have been expected given the magnitude of gains on the statewide tests (Klein et al., 2000). International studies also provide evidence that students in school systems with high-stakes testing programs outperform students in systems that do not have high-stakes testing programs (Bishop, 1997). These domestic and international results were important factors in the passage of NCLB by Congress.

In addition, there is a mounting body of research suggesting that SBA leads to significant changes in teachers' instructional practices and content coverage (Hamilton, 2004; Stecher, 2002). Tepper (2001) found that teachers in Chicago shifted content emphasis and time spent on test preparation in response to the introduction of a high-stakes testing program, and Swanson and Stevenson's (2002) analysis of state-level initiatives suggests that standards and assessments may influence instructional practices in different ways depending on features of the state policies. SBA policies also may improve learning by increasing student motivation (Betts & Costrell, 2001; Roderick & Engel, 2001), though the motivational effects are likely to vary according to the age of the student (Phillips & Chin, 2001) as well as the nature of the incentive system.

### **Research Relevant to the Implementation of SBA**

Our study design and data collection procedures are guided by a conceptual framework derived from research on implementation and scale-up. Figure 1 illustrates this conceptual framework and our understanding of how NCLB policies may get translated into classroom practice and student learning outcomes. The lists of policies, practices, and influences in the figure are not exhaustive but illustrate some of the features we are studying. The figure highlights the multi-level nature of this process. In NCLB, accountability policies are developed and promulgated primarily at the state level, but actions at all levels of the educational system—state, district, school, and classroom—affect the ways in which these policies are implemented and their impact on student learning. Our study is designed to gather information on implementation at all of these levels with an emphasis on understanding variations in implementation across schools and classrooms. Examining what happens in the classroom is

critical for understanding how SBA affects student learning, and that is where we place the greatest emphasis.



**Figure 1**  
**ISBA Conceptual Framework**

Studies of other state initiatives suggest that instruction tends to be resistant to externally-mandated policies because instructional activities are so distant from the source of the policy initiatives and tend to be poorly monitored and evaluated (Meyer & Rowan, 1978; Rowan, 1982; Weick, 1976). Furthermore, the incentive structures necessary to ensure fidelity of implementation are typically missing (Elmore, 1996). This lack of connection between policy and instruction makes any kind of systemic reform difficult (Elmore & Fuhrman, 1995). Unlike many other policy initiatives, SBA directly addresses these problems through ongoing monitoring and incentive systems. Whether these policies are effective, however, will depend on

their specific features and how they are enacted through the other levels of the system. Research points to particular issues that are relevant at each node of the conceptual framework, and these are the variables we tried to incorporate into the study. We reviewed the research and identified key variables that were relevant to implementation and scale-up at the state, district, school and classroom levels, as well as key features of parental involvement.

At the state level we are gathering information about the specificity and rigor of the content and achievement standards, the format and content of the tests, the alignment of tests with standards, the targets for performance, and the specific rewards and sanctions attached to performance. In addition, our state-level data collection efforts will be designed to provide a more careful and detailed examination of state capacity and state support and technical assistance practices. At the district level, we are collecting information on the district's role in transmitting state-mandated policies to schools, and the type and level of support, training, and resources districts provide to help schools and teachers translate state policy into classroom practice. Schools have similar responsibilities under NCLB, and we will be collecting data at the school level about improvement strategies and support for teacher change. For example, we will examine school decisions relating to curriculum, textbooks and instructional materials, professional development, supplemental testing, and extra student support services. The research literature also identifies a number of important contextual features that influence the actions of districts and schools that we will incorporate into our data collection. These include such features as capacity, instructional leadership, program coherence, resources, and the distribution of authority and responsibility.

Classroom practice is the key leverage point for SBA and a major focus of our study. Although teaching styles are fairly resistant to change, research has shown that policy

interventions can influence both what is taught and how it is taught. We will gather data on changes in curriculum and instruction in mathematics and science, and we will also examine the use of test scores in planning and the use of locally-adopted assessments. Finally, since NCLB is designed to provide increased options and accountability to parents, we will collect some information about parent perceptions of changes arising from NCLB. However, resource constraints limit how much information we can gather from parents.

### **Methods**

The study uses a combination of large-scale, quantitative data collection and small-scale case studies to examine SBA implementation and outcomes in three states. In each state, we trace SBA implementation at the state, district, school, and classroom level, and gather information on student achievement so that we can associate implementation factors with student outcomes. We focus on elementary and middle school science and mathematics. The choice of subjects was made in consultation with the National Science Foundation. Mathematics was chosen as a focus because it is one of the two required NCLB content areas. Science was chosen because states still had three years to comply with the regulations regarding science standards and assessments, and this would provide an opportunity for the project to study scale up as it unfolded.

### **Sampling**

**States.** Three states were selected to represent a range of approaches to implementing SBA, and to provide both geographic and demographic diversity. The states are California, Pennsylvania and Georgia. (The accountability systems and demographic features are described below.)

**Districts.** To sample districts and schools, we obtained from the NCES Common Core of Data for 2001-2002 a comprehensive list of schools and districts for each of the three states. To obtain our sampling frames, we restricted our attention to the populations of "regular" schools by eliminating vocational, charter, alternative and special education schools. All districts that contained regular schools were included in the district sampling frames. Power calculations assuming various configurations of between and within district variances indicated that taking fewer than 25 districts per state would not provide sufficient power to detect significance of district-level relationships. We cross-classified our districts based on the number of elementary schools and middle schools (roughly proportional to the size of the district), and for each state divided the districts into five strata based on this cross-classification. Stratification was necessary to obtain sufficient representation of the many small districts, while at the same time ensuring a sufficient number of total schools in our sample. The Los Angeles Unified district in CA was sampled with probability 1.

Initially, we sampled 27 districts per state. However, we were not able to enroll as many of these districts as we had anticipated. Refusals were particularly high in California (see Table 1). Districts gave a number of reasons for not participating, including additional pressures related to NCLB. As a result, a supplemental sample of 23 districts (primarily in California) was drawn and recruitment efforts were extended. Eventually, we recruited 68 districts to participate in the 2003-04 school year, representing a cooperation rate of 65%. The final sample is slightly lower than our design specification, and we are making efforts to increase it in subsequent years.

**Table 1.**  
**District Sample and Cooperation, 2003-04**

	California	Georgia	Pennsylvania	Total
Initial Sample	27	27	27	81
Replacement Sample	13	5	5	23
Total Sample	40	32	32	104
Cooperating Districts	19	25	24	68
Cooperation Rate	47.5%	78.1%	75%	65.4%

**Schools.** We aimed to study approximately 100 schools per state, which involved sampling about 125 schools per state and expecting about an 80% participation rate. To select schools, we first designated each school in a sampled district as either an elementary school, a middle school, or a combined school. Elementary schools contained grades 3 and 4, middle schools contained grades 7 and 8, and combined schools contained grades 3, 4, 7, and 8. We excluded very small schools by requiring that schools have on average 10 or more students in each of these grades.<sup>2</sup> Then we randomly sampled elementary schools and middle schools from the sampled districts, selecting between one and five schools of each type from each district according to a prearranged pattern based on district size. Schools were not contacted until districts agreed to participate. In California, a larger percentage of districts declined to participate than in Georgia or Pennsylvania. As a result, the California sample was smaller than the sample in the other two states. Districts usually assume responsibility for approving research requests, and once district cooperation is obtained, school cooperation is usually easier to obtain. Overall, we recruited 267 schools to participate in the study in 2003-04, representing a school cooperation rate of 90% (see Table 2).

**Table 2.**  
**School Sample and Cooperation, 2003-04**

	California	Georgia	Pennsylvania	Total
Sample	78	116	103	297
Cooperating Schools	63	108	96	267
Cooperation Rate	80.7%	93.1%	93.2%	89.9%

**Teachers.** Once schools were recruited, we obtained complete rosters of all teachers in the relevant subjects and grades. For elementary schools we asked for lists of all teachers who taught mathematics or science to students in grades three, four or five. The same request was made of middle schools focusing on grades seven and eight. All teachers in these grades and subjects were included in the study.

### **Instrumentation**

State-level data collection has been done through semi-structured, face-to-face interviews with key personnel in the state Departments of Education, including the state superintendent or designee, the director of assessment, the director of curriculum, and others. We also interviewed other policy actors, including legislators, staff members from the education committees of the legislature, members of staff from the state board of education, leaders of the teachers' union and the state school boards' association. In addition to the interviews, we are collecting relevant documents, such as copies of the state content standards and studies the state conducted to evaluate alignment between standards and tests.

At the district level, we gathered information from superintendents using both semi-structured telephone interviews and paper-and-pencil surveys. The surveys include questions on

---

<sup>2</sup> In California the criterion was 11 or more students per grade to mirror a limit used by the state for reporting purposes.

district actions related to SBA, as well as on the level of support districts receive from the state and on challenges related to the implementation of NCLB provisions, such as parental choice and supplemental services. The interviews captured more specific information on state-to-district contacts and district-to-school contacts than could be obtained through surveys.

At the school level, we gathered information from principals and teachers using surveys. These surveys addressed local actions related to the components of NCLB (standards, assessments, etc.) as well as the contextual factors that may influence these actions, such as the quality of leadership and the degree of collaboration among teachers. We developed a separate teacher survey for each state, which allowed us to use state-specific terminology, though the underlying questions were the same. We administered a common principal survey across the three states. In addition, the principal and teacher surveys include a number of questions that were administered to both sets of respondents, permitting us to compare principals' responses with those of teachers.

The draft surveys were all pilot tested with relevant groups of respondents (superintendents, principals, teachers). The pilot testing included detailed interviews, in which similar respondents in the participating states talked through their responses and informed us of any problems they had answering the questions. We made a number of revisions to the surveys as a result of these interviews. These survey development procedures provide us with a reasonable degree of confidence that our survey items are capturing the intended constructs.

In addition to the large-scale data collection, we conducted case study visits to 18 schools (three schools in each of two districts in each of the three states).<sup>3</sup> During these visits we interviewed principals and teachers, and we conducted focus groups with small groups of parents. The principal and teacher interview protocols used during these visits focused on

NCLB-related changes in the school, the nature and adequacy of technical assistance received, and the use of information from the accountability system. The parent focus group protocol asked parents about their perceptions of the school, their understanding of NCLB, their responses to the choice and supplemental services provisions, and their use of information from state tests. The case studies were primarily intended to provide richer information than we were able to collect through surveys. In addition, we used the case study information to revise our surveys for 2004-05 data collection in cases where the surveys failed to capture something important that was uncovered during the school visits. We will visit these schools again during spring 2005 and spring 2006.

### **Data Collection**

The schedule for data collection in 2003-04 is shown in Table 3. The bulk of our data collection activities occurred between late January and late May.

**Table 3.**  
**Data Collection 2003-2004**

<b>Activity</b>	<b>Dates</b>
District Recruitment	July – October, 2003
School Enrollment	October – December, 2003
State Interviews	September – November, 2003
Survey Pretest	November, 2003
Principal and Teacher Survey Administration	January – April, 2004
Superintendent Survey Administration	February – May 2004
School Site Visits	February – May, 2004
District Superintendent Interviews	February – May, 2004

---

<sup>3</sup> In one district we were only able to arrange visits to two schools.

**Survey Responses.** With a few exceptions, survey response rates were quite high. This may be due to the fact that principals and teachers who returned completed surveys received \$25.00 as an acknowledgment of their time and effort. School coordinators who helped in the distribution and collection of teacher surveys received an additional \$25.00. Superintendent response rates were also high, although no honoraria were paid to superintendents. Table 4 shows the response rates for superintendents in participating districts, by state. Overall, the response rate for participating districts was 88%.

**Table 4.**  
**Superintendent Survey Responses, 2003-04**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>	<b>Total</b>
Cooperating Districts	19	25	24	68
Completed Superintendent Survey	18	20	22	60
Survey Response Rate	94.7%	80.0%	91.6%	88.2%

Table 5 shows the corresponding response rates for principals and teachers in cooperating schools. Overall, 85% of principals and 83% of teachers in these schools returned completed surveys.

**Table 5.**  
**Principal and Teacher Survey Responses, 2003-04**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>	<b>Total</b>
Cooperating Schools	63	108	96	267
Principals Survey Responses	51	88	88	227
Principal Response Rate	80.9%	81.5%	91.7%	85.3%
Teacher Sample	692	1522	1073	3287
Teacher Survey Responses	487	1318	926	2731
Teacher Response Rate	70.4%	86.6%	86.3%	83.1%

**Achievement Data.** In future years, the project will look for associations between policies and practices as measured by our surveys and changes in student performance as measured by state achievement tests in mathematics and science. At the present time, the only information we have about student achievement is school-level summary statistics provided by states to comply with the accountability provisions of NCLB. We are still in the process of assembling a database that breaks down these scores by grade level and NCLB-defined subgroups (where the number of such students is deemed to be significant). In the future, we hope to be able to conduct more refined analyses using individual student scores linked over time, but it is not yet clear whether states will be able to provide us with these data.

**Analyses.** Different methods were used to analyze the results from the surveys, the superintendent interviews, and the case studies. For the surveys, state-specific weights were generated to make the responses reflect the state as a whole. The weights reflect both the sample design for each state and the patterns of survey non-response. As a result, the statistics reported in other papers in this symposium represent estimates of the responses of superintendents, principals and teachers statewide. One of the consequences of our sampling strategy, in which teachers and principals are nested within schools and schools are nested within districts, is that the number of responses grows progressively smaller as we move from teachers to principals to superintendents. As a result, the summary statistics based on teacher responses are more precise than those based on principal responses, which are more precise than those based on superintendent responses. To help the reader interpret the results, we have included estimates of the standard errors in all tables.

The 43 superintendent interviews were analyzed using the QSR N6 qualitative analysis software package. The interview notes were entered into the electronic database. A team of

three researchers developed a detailed coding scheme based on the various questions from our interview protocol, the study's overarching conceptual framework, and the major topic areas that emerged from an initial review of the interview notes. The electronic files were coded using this framework, and we used the software to generate detailed analyses of responses and identify themes within and across states.

The case studies were used primarily to obtain concrete examples to illustrate patterns revealed by the surveys and to afford fresh insight into implementation issues that might be incorporated in future data collection. Each interview was recorded on audiotape, and the tapes were transcribed to provide a searchable archive. In addition, the site visit team of two researchers wrote a summary of each school visit as well as an overall summary of each district visit. These summaries were read by most members of the research team as part of the analysis process.

## **Results**

The other papers in this symposium summarize results from the interviews and surveys conducted during 2003-04. However, before presenting those results, we describe some aspects of the educational systems of the three states to provide important background for the findings regarding standards-based accountability. In this section, we briefly describe contextual differences that are relevant to NCLB and accountability in California, Georgia and Pennsylvania.

### **Demographic features**

Table 6 summarizes some important demographic features of K-12 public education in California, Georgia and Pennsylvania. The states differ in a number of ways that might affect their implementation of SBA, including the size of their K-12 systems and the racial/ethnic

characteristics of their students. California is a large, diverse state that enrolls approximately one out of every eight students in the United States. It has about one thousand districts that vary greatly in enrollment. There are over 100 districts with more than 10,000 students each, but about one-half of the districts in the state have fewer than 1,000 students. Georgia is predominantly a county system of school districts with enrollments ranging from 351 to 13,769; most districts fall in the 2,500 to 10,000 range. Pennsylvania has few large districts – only two have more than 25,000 students—and many are quite small—128 districts have fewer than 2,000 students.

**Table 6.**  
**Size of K-12 Public School Systems**

<b>Feature</b>	<b>California (2003-04)</b>	<b>Georgia (2003-04)</b>	<b>Pennsylvania (2001-02)</b>
Number of Districts	1,059	180	501
Number of Schools	9,222	2,033	3,248
Number of Teachers	297,480	99,535	118,470
Number of Students	6,298,413	1,553,437	1,821,627

Source: Department of Education Web Sites for California, Georgia and Pennsylvania, accessed May 21, 2005.

Pennsylvania is the least diverse of the three states in terms of student race/ethnicity, while California is the most diverse with more than half of all students coming from “minority” groups (see Table 7). Georgia’s overall school enrollment is 38% black, 7% Hispanic, and 50% white. All three states have roughly the same proportion of students with disabilities. Neither Georgia nor Pennsylvania has large numbers of students who are English language learners; however, serving the needs of English learners is a major educational challenge in California.

**Table 7.**  
**Student Demographic Characteristics**

<b>Characteristic</b>	<b>California (2003-04)</b>	<b>Georgia (2003-04)</b>	<b>Pennsylvania (2001-02)</b>
White	32.5%	51%	77.7%
Hispanic	46.0%	7%	4.8%
African-American	8.1%	38%	4.8%
Asian	8.0%	3%	2.1%
Free/Reduced Lunch Eligible	49.0%	46%	--
English Learners	25.4%	4%	1.7%
Students with Disabilities	10.7%	12%	12.3%

Source: Department of Education Web Sites for California, Georgia and Pennsylvania, accessed May 21, 2005. Information about free/reduced price lunch eligible students was not found for Pennsylvania.

### **Prior State Accountability Systems**

In all three states, the implementation of standards-based accountability was shaped by the existing policy environment. While many state policy issues are relevant to the implementation of SBA—e.g., state budget, educational governance structure, political volatility—the most important may have been the states’ prior experience with testing and accountability. In some states, NCLB represented an entirely new policy; in others it confronted existing state testing and accountability systems that had to be modified or dismantled.

**California.** The Public School Accountability Act of 1999 (PSAA) created a test-based accountability system in California prior to NCLB. The state had already adopted academic standards and a comprehensive testing program (combining norm-referenced tests with standards-based tests). Under PSAA, each school was assigned an Academic Performance Index (API) based on student test scores. Schools were required to increase their scores each year by

an amount that was equal to 5% of the distance between their prior score and the state interim target of 800. The system included rewards and sanctions, which were based on gains in a school's API score, not on attaining a specific score level. For the first two years of the program, schools that made the greatest gains could receive large financial rewards that went directly to teachers. However, after 2000, the economic outlook worsened considerably and the reward funding was discontinued. Schools that did not meet their targets were subject to interventions to improve their performance. The interventions provided expert help to develop and implement improvement plans. If the interventions were not successful, then sanctions could be applied, including five forms of restructuring: reopening as a charter school, reconstitution, entering into a management contract, state takeover, or other major restructuring (Ziebarth, 2004).

Many of the features of the prior system were incorporated into the state's system under NCLB. For example, the API was used as the additional indicator required by NCLB (in addition to test scores), thus retaining to some degree the state's gain-based approach to accountability. However, the overlay of the federal system onto the existing state system has created some problems in California because the two systems can produce different judgments about individual schools. It is common for a school to be told it has achieved significant growth according to the state system but failed when measured against the federal benchmark. In fact, in 2004 the state Superintendent reported that 317 schools increased 30 points or more on API (a large gain), yet failed to make AYP.

**Georgia.** Georgia's legislature enacted an accountability system in 2000 modeled on the systems in Texas and North Carolina (Georgia's A Plus Education Act - HB1187). The A+ system was supposed to award a letter grade to each school (A-F) and assign rewards and interventions based on overall student achievement as well as gains in student achievement.

However, the grading system was not fully implemented as the state responded to the NCLB requirements and there were changes in the political leadership of the state. For several years, accountability has been an issue of contention between the Governor's office and the Department of Education. The A Plus Education Act created a new accountability system for Georgia schools, and it removed the responsibility for it from the Department of Education and placed it in the Governor's office. In fact, the newly-created Office of Educational Accountability (OEA) was responsible for drafting the state NCLB plan and submitting the state workbook. The contentious situation lessened after the election of a new Governor in 2003, and the current administration is beginning to re-integrate the functions of the OEA (which has been renamed the Office of Student Accountability) into the Department of Education.

The state's system is now fully compliant with NCLB, with AYP as the primary measure of school performance. The system includes sanctions for schools that persistently perform poorly. A school that is a poor performer for two years has a "school master" or management team appointed by the state. If the school's performance remains low for three or more consecutive years, it can be restructured, including reopening as a charter, reconstitution, or state takeover (Ziebarth, 2004).

Georgia is also in the process of completely revising its curriculum standards in response to an external audit (conducted by Phi Delta Kappa) that found the state's Quality Core Curriculum (QCC) lacked depth, covered too many topics, and was not aligned with national standards. The new performance-based Georgia Performance Standards (GPS) have received high marks in a subsequent PDK audit. The new standards will take effect in the 2005-2006 school year, giving teachers time to participate in professional development on the new

curriculum and also providing time to continue the revision of the state's assessments to align them with the standards.

**Pennsylvania.** Pennsylvania has a long history of local control, and the state Department of Education has traditionally had low involvement with local districts and schools. Prior to NCLB, the Pennsylvania legislature had passed “Empowerment Legislation” under which districts were identified as “educational empowerment districts” if more than 50 percent of students in the district scored at the “Below Basic” level on the state test—Pennsylvania System of School Assessment (PSSA)—for two years. About 12 districts were identified for improvement under this law in its first year, and their local boards were given wide latitude for improving schools, including reopening as a charter, reconstitution, or entering into a management contract, (Ziebarth, 2004). Another key feature of this system was the Pennsylvania Performance Index (PPI), a continuous improvement measure that was supposed to be used by the state to recognize “significant growth.”

In response to NCLB, Pennsylvania developed a new system that complied with the federal guidelines. Schools' status from the “empowerment legislation” was transferred over into the NCLB system. The PPI has been relegated to a minor role in the new system; schools that do not meet AYP but make substantial gains on the PPI may appeal the AYP status decision to the Department of Education, which may change their status “on appeal.” However, this procedure is now under review by the U.S. Department of Education.

Pennsylvania chose to have their accountability system apply equally to all schools and districts. Under this unified system, Pennsylvania plans to provide the same technical assistance and sanction to all schools that fail to make adequate yearly progress, not just Title I schools. (However, choice is only being offered to students in Title I schools.) Under the new NCLB-

mandated accountability system, the Department is taking a much more active role, and there is a debate in the legislature about how much control the Department of Education should have over districts. The larger role requires more funding, but the legislature has been reluctant to expand the budget of the state agency.

### **State Testing Programs**

Each state had a testing program prior to NCLB, but none was completely adequate to meet the basic NCLB requirements. California and Georgia were closest, testing students in reading and mathematics in most of the required grades prior to NCLB, while Pennsylvania was only testing in “milepost” grades (i.e., one elementary grade, one middle school grade, one high school grade). All three states have made changes to their testing program to comply with NCLB, and most still have additional changes in the works. For example, California has been gradually shifting from norm-referenced tests to standards based tests over the past decade. Each state has been modifying their testing programs during this period, and a year by year summary is extremely complex. At the present time, all three state testing programs meet the NCLB requirements in all areas except science. Georgia has been testing in science in the required grades for many years. California will implement required science testing in 2004-05, while Pennsylvania science tests will not be operational until 2006-07. Table 8 summarizes some of the key features of states testing programs as of 2004-05.

**Table 8.**  
**State Testing, 2004-05**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>
Test Name	California Standards Test (CST)	Criterion-Referenced Competency Test (CRCT)	Pennsylvania System of School Assessment (PSSA)
Comprehensive Testing (Grade)	Mathematics (3-8) Reading (3-8) Science (5) (8 and 10 in 2005)	Reading (1-8) Mathematics (1-8) Science (3-8)	Reading (3, 5, 8, 11) Mathematics (3, 5, 8, 11) Writing (6, 9, 11) Science (2007-08)
Additional Testing	California Achievement Test-6 (4, 7)	Stanford-9 (3, 5, 8) Writing (3, 5)	Locally developed tests (4, 6, 7)
End of Course Testing (High School)	Mathematics Science	Mathematics Science	
High School Exit Exam	California High School Exit Exam (CAHSEE)	Georgia High School Graduation Tests (GHS GT) - English/language arts, mathematics, writing, science, social studies.	Proficient level in reading, writing, and mathematics on either state or local assessment.
Second Language Testing	Spanish Assessment of Basic Education, Second Edition (SABE/2)		
Alternative Assessment	California Alternative Performance Assessment (CAPA)	Georgia Alternative Assessment for students with severe disabilities	Pennsylvania Alternative Student Assessment (PASA)

### **NCLB Status**

Based on 2003-04 testing, more than one-third of the schools in California did not make adequate yearly progress (see Table 9). In Pennsylvania the rate was 22%; in Georgia it was

18%. Yet, Georgia had a higher percentage of schools in each of the other NCLB designated conditions, identified for improvement, corrective action and restructuring. This condition exists, in part, because Georgia implemented its Title I accountability system earlier than Pennsylvania or California, so schools that are struggling are further along in the identification and remediation process. Pennsylvania has the lowest percentage of schools in each category. While all three states must be concerned about poorly performing schools, Georgia is facing the greatest current challenges as far as required district and state intervention is concerned.

**Table 9.**  
**2004-05 NCLB Status (Based on 2003-04 Testing)**

School Status	California	Georgia	Pennsylvania
Making AYP	64%	78%	82%
Identified for Improvement	13%	21%	8%
Corrective Action	2%	3%	0%
Restructuring	3%	4%	2%

### **Teacher Qualifications**

NCLB also requires that all students be taught by “highly qualified teachers.” States must take steps to bring all teachers up to this status. California is facing a much greater challenge than the other two states in terms of highly qualified teachers. Table 10 shows the overall percentage of classes in each state taught by teachers who were highly qualified in 2002-03, and the percentages in high- and low-poverty schools.

**Table 10.**  
**Percent of Classes Taught by Highly Qualified Teachers, 2002-03**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>
All Schools	48%	94%	97%
High-Poverty Schools	35%	95%	92%
Low-Poverty Schools	53%	95%	99%

### **Outline of Symposium**

The remainder of the symposium will present findings from our 2003-04 interviews and surveys.

The second paper, presented by McCombs, will focus on standards and assessments. Following that, Marsh, Russell, and Barney will describe accountability provisions (AYP, sanctions, etc.).

The fourth paper by Barney and Robyn examines technical assistance and support. Lastly,

Hamilton, Berends and Stecher will discuss classroom-level effects.

### References

- Betts, J. R., & Costrell, R. M. (2001). Incentives and equity under standards-based reform. In D. Ravitch (Ed.), *Brookings Papers on Education Policy: 2001* (pp. 9-74). Washington, DC: Brookings Institution Press.
- Bishop, J. (1997). *The effect of curriculum-based exit exam systems on student achievement*. CAHRS Working Paper 97-15.
- Education Commission of the States (2002, December). States conducting student competency testing for high school graduation (exit exams). Denver: Author. Downloaded from <http://www.ecs.org/clearinghouse/41/18/4118.htm> on March 21, 2005.
- Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review, 66*, 1-17.
- Elmore, R. F., & Fuhrman, S. H. (1995). Opportunity-to-learn standards and the state role in education. *Teachers College Record, 96*, 432-457.
- Grissmer, D W., & Flanagan, A. (1998). *Exploring rapid score gains in Texas and North Carolina*. Washington, DC: National Education Goals Panel.
- Grissmer, D. W., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP scores tell us* (MR-924-EDU) Santa Monica, CA: RAND.
- Hamilton, L. (2004). Assessment as a policy tool. *Review of Research in Education, 27*, 25-68.
- Kirst, M. W. (1990). *Accountability: Implications for state and local policymakers*. Washington DC: Office of Educational Research and Improvement, U.S. Department of Education.

- Klein, S. P., Hamilton, L. S., McCaffrey, D., F., & Stecher, B. M. (2001). *What do test scores in Texas tell us?* (IP-202). Santa Monica, CA: RAND.
- McKnight, C. C., Crosswhite, F. J., Dossey, J. A., Kifer, E., Swafford, J. O., Travers, K. J., and Cooney, T. J. (1987). *The underachieving curriculum: Assessing US schools mathematics from an international perspective*. Champaign, IL: Stipes Publishing Company.
- Meyer, J. W., & Rowan, B. (1978). The structure of educational organizations. In M. Meyer (Ed.), *Environments and organizations* (pp. 78-109). San Francisco: Jossey-Bass.
- National Council on Education Standards and Testing. (1992). *Raising standards for American education*. Washington, DC: Author.
- National Governor's Association (1989). *Results in education: 1989*. Washington, D. C.: author.
- National Research Council (1999). *Testing, Teaching, and Learning*. Washington DC: National Academy Press.
- Phillips, M., & Chin, T. (2001). Comment on Betts & Cotrell. In D. Ravitch (Ed.), *Brookings papers on education policy: 2001* (pp. 61-66). Washington, DC: Brookings Institution Press.
- Resnick, D. (1982). History of educational testing. In A. K. Wigdor and W. R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies, Part II*. Washington, D. C.: National Academy Press, pp. 173-194.
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23, 197-227.

- Rowan, B. A. (1982). Organizational structure and the institutional environment: The case of public schools. *Administrative Science Quarterly*, 27, 259-279.
- Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher, & S. P. Klein (Eds.), *Making sense of test-based accountability in education*. Santa Monica, CA: RAND (forthcoming).
- Stecher, B., Hamilton, L., and Gonzalez, G. (2003). *Working smarter to leave no child behind*. WP-138-EDU. Santa Monica: RAND.
- Swanson, C., & Stevenson, D. L. (in press). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis*.
- Texas Education Agency (2000). *Texas TAAS passing rates hit seven-year high; four out of every five students pass exam*. Press release, May 17.
- Tepper, R. L. (2002). *The influence of high-stakes testing on teacher behavior and instructional practice in Chicago*. Working paper.
- Travers, K. J., and Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curriculum*. Oxford: Pergamon Press.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1-19.
- Ziebarth, T. (2004, April). Models of state education governance. Denver, CO: Education Commission of the States. Downloaded from <http://www.ecs.org/clearinghouse/50/73/5073.htm> on March 21, 2005.
- Ziebarth, T. (2004, December). StateNotes: State policies for school restructuring. Denver, CO: Education Commission of the States.