

# WORKING P A P E R

---

## Implementing Standards- Based Accountability (ISBA)

### Study Design, State Context, and Accountability Policies

BRIAN M. STECHER AND SCOTT NAFTEL

WR-380-EDU

April 2006

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 at the symposium “Implementing No Child Left Behind: New Evidence from Three States” at the annual meeting of the American Educational Research Association, April, 2006, San Francisco.

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.



-

## Contents

Preface.....	iii
Background and Goals.....	1
Conceptual Framework for ISBA Study.....	2
Study Design and Methods.....	5
Sampling.....	5
Analyses.....	11
State Context and Accountability Policies.....	11
Prior Accountability Systems.....	12
State Academic Content Standards.....	17
State Testing Programs.....	20
Adequate Yearly Progress (AYP) Definitions.....	22
NCLB Status of Schools.....	26
Summary.....	30
Remainder of the Symposium.....	31
References.....	33



## Figures and Tables

Figure 1. ISBA Conceptual Framework .....	4
Table 1. Size of K-12 Public School Systems, 2003-04 .....	6
Table 2. Student Demographic Characteristics, 2003-04 .....	7
Table 3. District Sample and Cooperation, 2004-05.....	8
Table 4. School Sample and Cooperation, 2004-05 .....	8
Table 6. Superintendent Survey Responses, 2004-05.....	10
Table 7. Principal and Teacher Survey Responses, 2004-05.....	10
Table 8. Content Standards in English Language Arts, Mathematics, and Science, by State and Grade Level.....	18
Table 9. State Testing Programs, 2004-05.....	20
Figure 2. Annual Measurable Objectives for Reading/ELA, by State (2002-2014).....	23
Figure 3. Annual Measurable Objectives for Math, by State (2002-2014) .....	24
Table 10. State Accountability Systems and Key Definitions (2002-03, 2003-04).....	25
Table 11. Percentage of Schools Making Adequate Yearly Progress, 2002-03 to 2004-05.....	27
Table 12. Percentage of All Schools Identified for Improvement, 2002-03 to 2004-05.....	29
Table 13. Percentage of Districts Making Adequate Yearly Progress, 2002-03 to 2004-05.....	30

## **Background and Goals**

The test-based accountability provisions of the No Child Left Behind Act (NCLB) of 2001 (20 U.S.C. § 6311 et seq.) are one of the primary policy initiatives affecting schools and districts in the United States. NCLB requires states to adopt content and achievement standards, to measure student progress toward those standards, and to implement a series of interventions in schools and districts that fail to meet their targets. Together, these standards, assessments, and consequences constitute a standards-based accountability (SBA) system. The law's SBA requirements affect every public school in the nation, and have already had a substantial impact on school and district actions (Center for Education Policy, 2004).

Advocates of SBA often claim that these policies will encourage educators to focus their efforts on improving achievement in core subjects and boost student proficiency, whereas detractors worry that attaching higher stakes to test results will lead to adverse consequences, such as narrowing of the curriculum or excessive test preparation resulting in invalid test scores (Hamilton, 2003; Koretz, Linn, Dunbar & Shepard, 1991). Whether NCLB's accountability provisions ultimately turn out to be beneficial or detrimental is likely to depend in large part on how states, districts, and schools implement the provisions and respond to the incentives.

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

-  
gathering data from 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 symposium provides a description of the ISBA project and presents descriptive results from the second year of data collection that address the first two of these four questions. Future reports will use data from multiple years to address the remaining questions.

### **Conceptual Framework for ISBA Study**

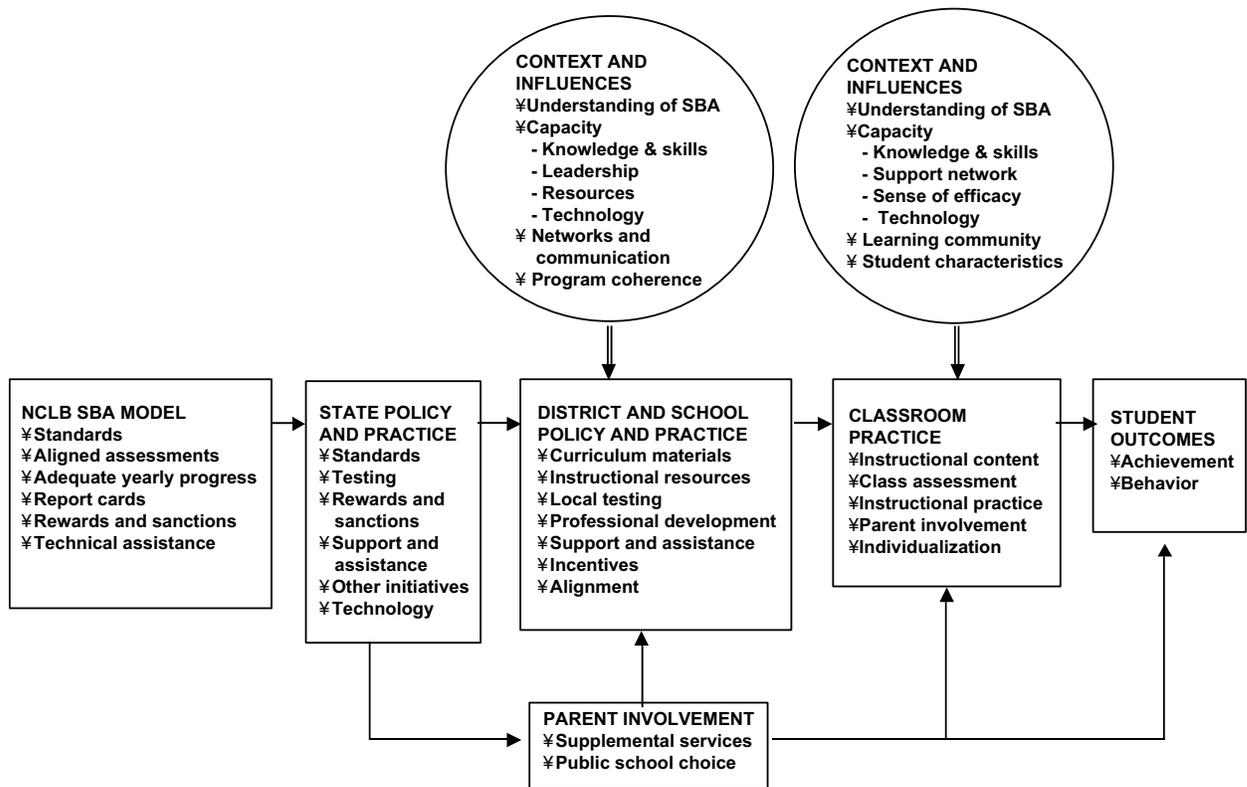
Our study design and data collection procedures are guided by a conceptual framework derived from research on implementation and scale-up of reforms, and by the

goal of understanding classroom-level responses. 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. As noted above, examining what happens in the classroom is critical for understanding how SBA affects student learning, and that is where we place the greatest emphasis.

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. The study is designed to gather descriptive information on each of these variables and to investigate relationships among them and with student achievement.

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



## **Study Design and 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 levels, 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, though we also collect some information on reading instruction and achievement because reading is a focus of states' accountability systems. The study is longitudinal in nature with three waves of data collection, 2003-04, 2004-05, and 2005-06. This symposium is based primarily on results from survey and case study data collected in the spring of the 2004-05 school year.

### ***Sampling***

**States.** Three states were selected to represent a range of approaches to implementing SBA and to provide both geographic and demographic diversity. The study is being conducted in California, Georgia, and Pennsylvania. Table 1 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 that are quite small—128 districts have fewer than 2,000 students.

**Table 1.**  
**Size of K-12 Public School Systems, 2003-04**

<b>Feature</b>	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>
Number of Districts	1,059	180	500
Number of Schools	9,222	2,033	3,253
Number of Teachers	297,480	99,535	119,888
Number of Students	6,298,413	1,553,437	1,821,146

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 and ethnicity, while California is the most diverse with more than half of all students coming from “minority” groups (see Table 2). Georgia’s overall school enrollment is 38 percent black, 7 percent Hispanic, and 50 percent 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 2.**  
**Student Demographic Characteristics, 2003-04**

<b>Characteristic</b>	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>
White	33%	51%	76%
Hispanic	46%	7%	6%
African-American	8%	38%	16%
Asian	8%	3%	2%
Free/Reduced Lunch Eligible	49%	46%	28%
English Learners	25%	4%	2%
Students with Disabilities	11%	12%	14%

Source: Department of Education Web Sites for California, Georgia and Pennsylvania, accessed May 21, 2005. Pennsylvania percentage for English learners is for 2001-02. Pennsylvania percentages for free/reduced lunch eligible and students with disabilities are from the NCES Common Core of Data for 2003-04.

**Districts.** In the first year of the study, we initially sampled 27 districts per state, stratified by size, for a total of 81 districts. However, we were not able to enroll as many of these districts as we had anticipated. For reasons we do not fully understand, refusals were particularly high in California. 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 percent.

For the 2004-05 school year we added to the sample of districts drawn for the 2003-04 school year to both increase the overall number of districts and to increase the number of districts where high percentages of schools were not meeting NCLB achievement goals. For this refresher sample we used the same stratified sampling techniques as in the first year, but restricted the sample to districts with lower achieving

schools. The refresher sample contained 28 districts, increasing the total sample for 2004-05 to 132 districts. Ninety-two districts agreed to cooperate, yielding a cooperation rate of about 70 percent (see Table 3.)

**Table 3.**  
**District Sample and Cooperation, 2004-05**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>	<b>Total</b>
Total Sample	56	37	39	132
Cooperation	31	30	31	92
Cooperation Rate	55%	81%	80%	70%

**Schools.** 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. Overall, we recruited 301 schools to participate in the study in 2004-05, representing a school cooperation rate of 85 percent (see Table 4).

**Table 4.**  
**School Sample and Cooperation, 2004-05**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>	<b>Total</b>
Sample	122	124	107	353
Cooperation	91	111	99	301
Cooperation Rate	75%	90%	93%	85%

**Teachers.** Once schools were recruited, we obtained complete rosters of all teachers in the relevant subjects and grades. All teachers who taught mathematics or science to students in grades three, four, five, seven or eight 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 Department 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 or staff from the state board of education, leaders of the teachers' union and the state school boards' association.

At the district level, we gathered information from superintendents using both semi-structured telephone interviews and paper-and-pencil surveys. At the school level, we gathered information from principals and teachers using surveys.

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>1</sup>. During these visits we interviewed principals and teachers, and we conducted focus groups with small groups of parents.

### **Survey Response Rates for 2004-05**

Principal and teacher survey response rates were quite high. Table 6 shows the survey response rates for superintendents in participating districts, by state. Overall, the

---

<sup>1</sup> In one site we were only able to arrange visits to two schools.

response rate for superintendents in 2004-05 was 73 percent, which was somewhat lower than the 88 percent response rate in 2003-04.

**Table 6.  
Superintendent Survey Responses, 2004-05**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>	<b>Total</b>
Cooperating Districts	31	30	31	92
Completed Superintendent Survey	24	24	19	67
Survey Response Rate	77.4%	80.0%	61.3%	72.8%

Table 7 shows the corresponding response rates for principals and teachers in cooperating schools. Overall, 86 percent of principals and 87 percent of teachers in these schools returned completed surveys in 2004-05. These rates were slightly higher than the response rates for 2003-04.

**Table 7.  
Principal and Teacher Survey Responses, 2004-05**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>	<b>Total</b>
Cooperating Schools	91	111	99	301
Principals Survey Responses	78	95	87	260
Principal Response Rate	86%	86%	88%	86%
Teacher Sample	1013	1605	1050	3668
Teacher Survey Responses	826	1409	938	3173
Teacher Response Rate	81.5%	87.8%	89.3%	86.5%

## **Analyses**

Survey weights were generated at the district, school, and teacher levels based on differing probabilities of selection and non-response rates. The weights were designed to compensate for the differences between samples and populations resulting from the sampling design and response patterns and give survey data estimates that are representative of the states. As a result, the survey statistics in this report represent estimates of the responses of superintendents, principals and teachers statewide. To help the reader interpret the results, we have included estimates of the standard errors in all tables.

In 2003-04, the first year of the study, we also interviewed superintendents, and the 43 superintendent interviews that were conducted were analyzed using the QRS N6 qualitative analysis software package. Some of the results from these analyses are reported where relevant in this symposium, as well.

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. State-level interviews were used to provide background and policy context specific to each of the three states.

### ***State Context and Accountability Policies***

States have many decisions to make in implementing a standards-based accountability system consistent with NCLB (Stecher, Hamilton & Gonzalez, 2003). In this paper, we summarize some of the key elements that differentiate the context and

-

accountability policies of California, Georgia, and Pennsylvania, including the accountability systems that were in place prior to NCLB implementation, academic content standards, the testing programs states have adopted to comply with NCLB, states' methods for defining AYP, and the status of their schools with respect to meeting AYP targets. This information provides relevant background for interpreting the results to be presented in subsequent presentations.

### *Prior Accountability Systems*

In all three states, the implementation of standards-based accountability was shaped by the existing policy context, including state budget constraints, the structure of educational governance in the state, and the political environment in which the educational reforms were operating. Perhaps the most important factor in determining the direction each state took with respect to SBA was the state's prior experience with testing and accountability. In some states, NCLB accountability represented a fairly dramatic shift in education policy; in others the legislation was implemented in the context of elaborate state testing and accountability systems that had to be modified or dismantled. The implementation of NCLB on top of pre-existing state accountability systems led to tensions in some states as policymakers attempted to comply with NCLB while maintaining continuity with earlier approaches to evaluating schools. The co-existence of two SBA systems forced educators to deal with sometimes-divergent information provided by each system (Linn, 2005). The next sections describe the existing SBA systems in each of our participating states and discuss ways in which the

states have incorporated features of these systems into their new NCLB-compliant accountability systems.

**California.** California had extensive experience with SBA before NCLB was implemented. The Public School Accountability Act of 1999 (PSAA) created a test-based accountability system for all public schools in the state. This system was intended to provide the public with access to information on school performance and to create incentives for educators to focus on improving student achievement in their schools. The state adopted academic standards and a comprehensive testing program that combined commercially available norm-referenced tests with custom-developed standards-based tests. The emphasis placed on the latter relative to the former gradually increased as the system matured. Under PSAA, each school was assigned an Academic Performance Index (API) based on student test scores. Schools were required to increase their API each year by an amount that was equal to 5 percent of the distance between their prior scores and the state interim target of 800 (on a scale with a maximum value of 1000). The system included rewards and sanctions based on annual gains in a school's API score rather than on attainment of a specific score level. For the first two years of the program, schools that made the greatest gains received large financial rewards that went directly to teachers. 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 intervention program provided schools with expert help to develop and implement improvement plans. If these plans 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 created some problems in California. Educators were frustrated by changes in the way accountability was determined and the consequences that followed from it. More troubling was the fact that the two systems sometimes led to different conclusions about the performance of individual schools. It was common for a school to be told it had achieved significant growth according to the state system but had not met the AYP 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. Despite this fact, California continued to report both AYP and API results and many educators continued to believe that the state's system was superior to the federal system (O'Connell, 2004).

**Georgia.** Georgia was implementing test based accountability for the first time just as NCLB was enacted. In 2000, Georgia's legislature enacted a school accountability system (Georgia's A Plus Education Act - HB1187) modeled on the systems in Texas and North Carolina. Georgia had an extensive state testing program on which to build the accountability system, including tests in reading, mathematics and science in grades 3-8. The A Plus system was supposed to award a letter grade (A-F) to each school based on both overall student achievement and gains in student achievement; these grades were supposed to be the basis for assigning rewards and interventions.

-

However, the grading system was not fully implemented by the time the state had to respond to NCLB, and the plans were changed to create a system that was in compliance with the federal requirements although it retained some elements of the A Plus system.

Changes in the accountability system reflected changes that were occurring in the political leadership of the state. For several years, accountability had been a point of contention between the Governor's office and the Department of Education. In addition to creating an accountability system, the A Plus Education Act removed the responsibility for accountability from the Department of Education and placed it in a new Office of Educational Accountability that reported to the Governor. The Office of Educational Accountability (OEA) was actually given the responsibility for drafting the state NCLB plan and submitting the state workbook. In 2003, a new Governor was elected from the other party, and efforts were begun to reverse the changes in educational governance that had been enacted under the previous Governor. For example, the functions of the OEA (which was renamed the Office of Student Accountability) are being re-integrated into the Department of Education.

After much political wrangling accompanied by frustration on the part of the schools, the state's accountability system is 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).

**Pennsylvania.** Pennsylvania has traditionally been a “local control” state with a limited state testing program. This was starting to change prior to NCLB, when the state enacted “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 Pennsylvania System of School Assessment (PSSA) for two years. (The PSSA was administered only in grades 5, 8 and 10.) 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, reconstituting, or entering into a management contract (Ziebarth, 2004). As a result of the Empowerment Legislation, the state also created the Pennsylvania Performance Index (PPI), a continuous improvement measure that was supposed to be used by the state to recognize “significant growth.”

However, this system was far from compliant with NCLB, and Pennsylvania had to make significant changes to meet the federal guidelines. The changes included revised content standards, additional tests, new performance standard setting, and a new formula for identifying poorly performing schools. In the first year, each schools’ status from the “empowerment legislation” was transferred over into the NCLB system. In addition, Pennsylvania retained the PPI by allowing schools and districts to use it to appeal AYP status by demonstrating significant growth.

Pennsylvania planned to have its accountability provisions apply equally to all schools and districts rather than just to Title I schools, as required by NCLB. In particular, state leaders wanted to provide the same technical assistance and sanctions to all schools that failed to make AYP, regardless of Title I status. However, budget

-

constraints made this impossible; NCLB sanctions such as school choice and supplemental education services (SES) are only being offered to eligible students in Title I schools.

As noted above, Pennsylvania has a long history of local control, and the state Department of Education has traditionally had limited direct involvement in local districts and schools. However, under the new NCLB-mandated accountability system, the Department is taking a much more active role in guiding local education policy. There is a debate in the legislature about how much control the Department of Education should have over districts. The increased role requires more funding, but the legislature has been reluctant to expand the budget of the state Department of Education.

Pennsylvania has recently implementing a pilot program to provide schools with estimates of student achievement based on value-added models. This program, the Pennsylvania Value-Added Assessment Program (PVAAS), is being implemented in phases, but it will eventually be implemented in all of the state's districts. Whether it will be incorporated formally into the state's accountability system remains to be seen.

### ***State Academic Content Standards***

Academic content standards define the goals for each state's educational system. Schools are expected to use these statements about "what students should know and 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. Like a majority of states, California, Georgia, and Pennsylvania had already

adopted content standards in certain grades and subjects prior to NCLB. However, the grades and subjects in which content standards were in place in 2003-04 varied across the three states. Table 8 indicates the grades in which each state had implemented standards in English language arts, mathematics, and science as of the 2003-04 school year.

California's content standards outline the skills and knowledge that all students should master at each grade level for all three of these subjects plus a number of others, including history and social science and visual and performing arts. They have been praised by a number of organizations for their clarity and comprehensiveness. In particular, California was one of only three states to receive a grade of "A" from the Fordham Foundation in its evaluation of state mathematics standards (Klein, 2005). Although there were contentious debates about the standards when they were adopted, they have now been in place for a number of years and educators and policymakers are generally satisfied with them.

**Table 8.**  
**Content Standards in English Language Arts, Mathematics,**  
**and Science, by State and Grade Level**

<b>State</b>	<b>English Language Arts</b>	<b>Mathematics</b>	<b>Science</b>
California	K-12	K-12	K-12
Georgia	K-8, High school	K-8, High school	K-8, High school
Pennsylvania	3, 5, 8, and 11	3, 5, 8, and 11	4, 7, 10, and 12

Georgia's state content standards were revised in 2004-05. The Quality Core Curriculum (QCC), which was in place when NCLB was adopted, described expectations for English language arts, mathematics, science, social studies, foreign languages, fine arts, health, physical education, technology/career education, and agriculture. The QCC contained grade-by-grade standards for grades K–8 and grade cluster standards for 9–12 (i.e., descriptions that apply to a range of grades). However, an external audit conducted by Phi Delta Kappa (PDK) found that the QCC lacked depth, covered too many topics, and was not aligned with national standards. Georgia revised its curriculum standards, and the new standards, called the Georgia Performance Standards (GPS), received high marks in a subsequent PDK audit. The GPS took 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's academic standards define what students should know and be able to do in reading, writing, speaking, listening, and mathematics in grades 3, 5, 8, and 11 and in science in grades 4, 7, 10, and 12. The state is now expanding the standards to delineate content in each of the tested grades and subjects required under NCLB. However, rather than rewriting the standards, the state produced a supplemental document in 2004-05 called the Assessment Anchors. The Pennsylvania Department of Education does not intend for the Assessment Anchors to replace the standards; rather the anchors are to be used to clarify which standards are going to be assessed at which grade levels on the PSSA.

### *State Testing Programs*

Each state had a testing program prior to NCLB, but they varied considerably in the number of grades and subjects tested, and, hence, whether they satisfied the requirements of NCLB. Georgia's testing system provided all the achievement information required by NCLB. However, as the state's content standards were revised, the tests had to be revised to remain aligned with the standards. California tested students in reading and mathematics in most of the required grades prior to NCLB, but it did not test in science. The state had been gradually shifting from norm-referenced tests to standards-based tests in all subjects, and this change was accelerated by NCLB. In addition, California implemented the required science testing in 2004-05. Pennsylvania was only administering a common state test in "milepost" grades (i.e., one elementary grade, one middle school grade, one high school grade) prior to NCLB. As a result of the law, it will add reading and mathematics tests in the other grade levels in 2005-06. In addition, Pennsylvania will add science tests in 2006-07. Table 9 summarizes some of the key features of states testing programs as of 2004-05, but as this summary indicated, many changes are occurring.

**Table 9.**  
**State Testing Programs, 2004-05**

<b>Feature</b>	<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)

<b>Feature</b>	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>
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 (GHSGT) - 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)

### *Adequate Yearly Progress (AYP) Definitions*

The three study states differed in how they set the Annual Measurable Objectives (AMO) that schools and districts (called school systems in Georgia) had to meet each year and in how they computed AYP for schools and districts.<sup>2</sup> All three states utilized the same method for determining AMO starting points based on schoolwide average proficiency levels from 2001-02.<sup>3</sup> In 2002-03, this starting point in California meant that 13.6 percent of students in each subgroup had to reach proficiency in reading (English Language Arts) and 16 percent in mathematics. Pennsylvania's initial proficiency targets were somewhat higher that year with 45 percent for reading and 35 percent for math. Georgia's starting points were higher still, with the expectation that 60 percent of students in each subgroup reach proficiency in reading/ELA and 50 percent in math.

The states selected different growth trajectory patterns to achieve the federal goal of 100 percent of students proficient by the year 2014. All three states adopted "stair step" trajectories in which the AMO remains the same for a three-year period before increasing (see Porter, Linn, & Trimble, 2005, for a discussion of variability in trajectories across states). After two or three steps, the AMO increases linearly to 100 percent in the final years. However, California's trajectory has just two steps whereas Georgia's and Pennsylvania's have three. Thus, after 2007, student performance in

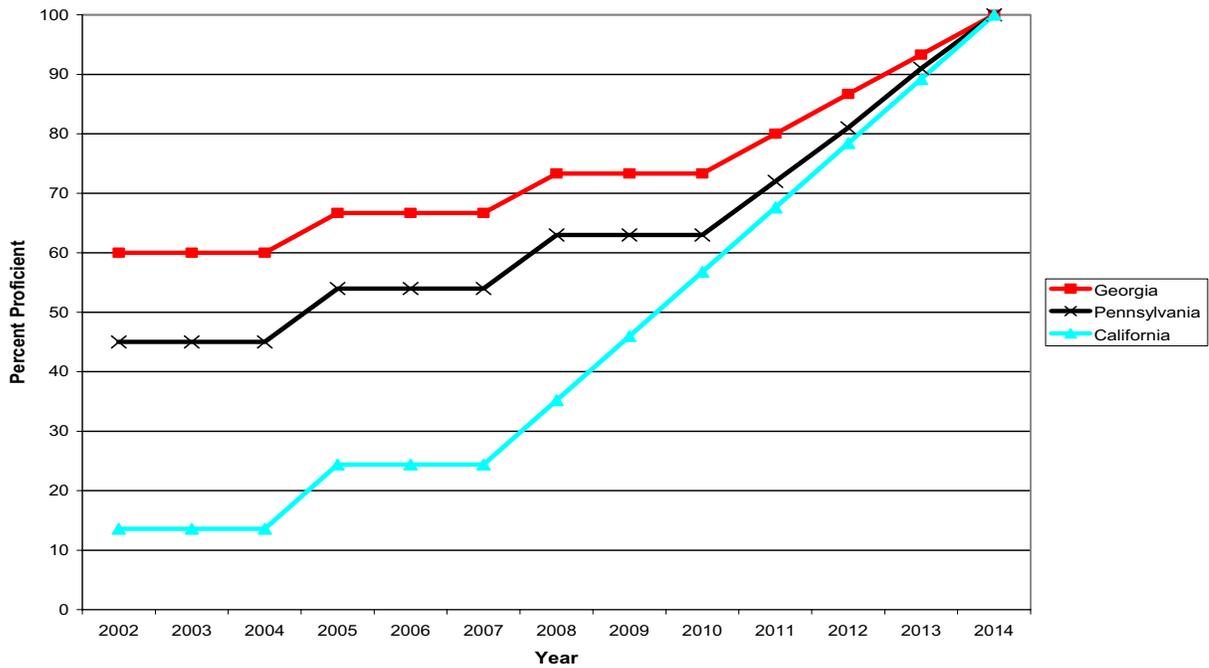
---

<sup>2</sup> Given our study's focus, we examine basic definitions for the elementary and middle school levels only. Requirements for districts are similar to those of schools: they generally have to meet the same targets, based on all individual student data.

<sup>3</sup> States generally ranked all schools in the state from highest to lowest performance based on average student proficiency levels in 2001-02. They then selected the one school in that list representing the 20<sup>th</sup> percentile for student enrollment (i.e., this school and all those below it on this ranked list of higher to lower performing schools enrolled 20 percent of all students in the state). The level of proficiency in math and reading/ELA in that school became the starting point for AMO trajectories.

California must increase linearly to the target of 100 percent proficient by 2014. Georgia and Pennsylvania have another level plateau from 2008-2010, and then increase linearly to 100 percent proficient thereafter.

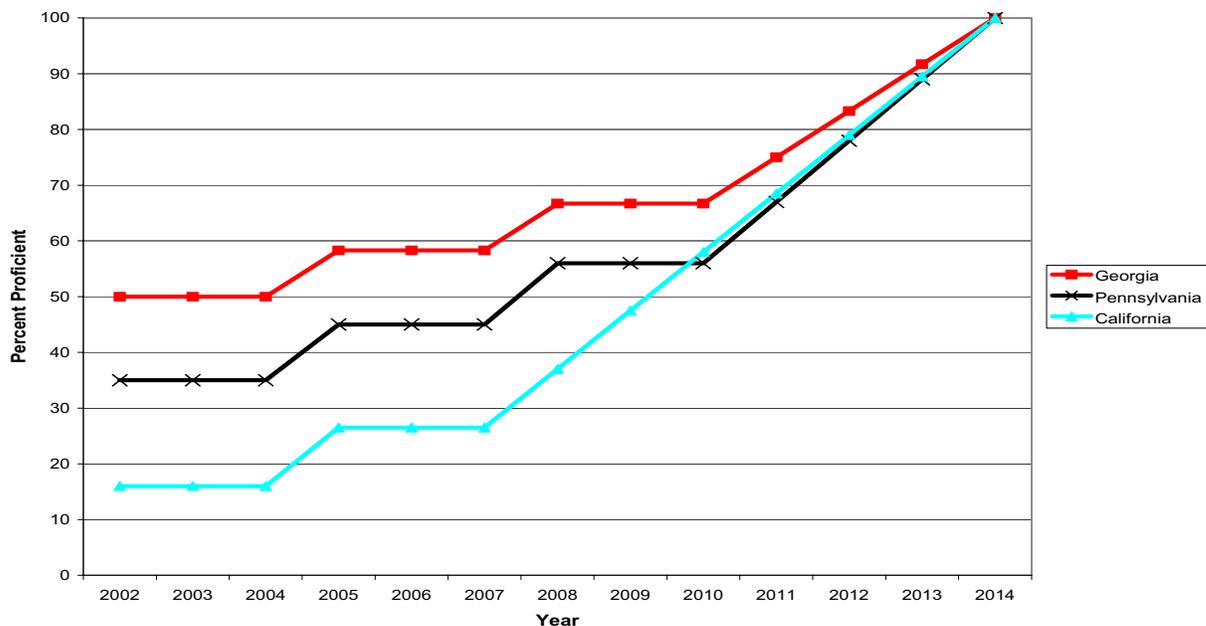
**Figure 2.**  
**Annual Measurable Objectives for Reading/ELA, by State (2002-2014)**



The computation of AYP is also different among the three states. AYP calculations are complicated, involving multiple tests, multiple subgroups, additional indicators, test participation rates, rules governing subgroup size, the use of confidence intervals (in some states), and a number of other features. For the purpose of this discussion, we mention two key ways in which the state AYP calculations differ: subgroup size and additional indicators. All three states report performance for the student subgroups required by law—including black, American Indian/Alaskan,

Hispanic, white, Asian, economically disadvantaged, students with disabilities, and limited English proficient.<sup>4</sup> However, they differ in the minimum number of students needed to qualify as a significant subgroup for reporting purposes. In Georgia and Pennsylvania, 40 students in any one category constituted a subgroup. California set a higher threshold, defining a minimum group size as either 100 students in any school or 50 students if that number constitutes 15 percent of total student enrollment in a school. As demonstrated by Porter, Linn, and Trimble (2005), decisions about subgroup size can have large effects on AYP determinations.

**Figure 3.**  
**Annual Measurable Objectives for Math, by State (2002-2014)**



---

<sup>4</sup> California defined three separate Asian subgroups: Asian, Filipino, Pacific Islander. Georgia and Pennsylvania also included the subgroup of “multi-racial.”

The states also varied in what they chose as their additional measure of academic progress in elementary and middle schools. Georgia and Pennsylvania initially selected attendance as the additional indicator. California selected the state’s Academic Performance Index (API) as the second indicator, requiring schools to either achieve a schoolwide API of 560 (this target increases over time) or demonstrate improvement by at least one point over the previous year. However, in 2003-04, both Georgia and Pennsylvania received permission to change their requirements. Pennsylvania decreased the attendance target from 95 percent to 90 percent. Georgia allowed districts to select among several options, including attendance as well as writing test scores, percent of students exceeding proficiency on reading, math, science, or social studies tests; and percent proficient on science or social studies tests. Table 10 summarizes these features as well as other aspects of state definitions and rules regarding AYP for the 2002-03 and 2003-04 school years.

**Table 10.**  
**State Accountability Systems and Key Definitions (2002-03, 2003-04)**

	<b>California</b>	<b>Georgia</b>	<b>Pennsylvania</b>
<b>AMO: Proficiency Targets</b>	ELA – 13.6% Math – 16.0%	ELA/Reading – 60% Math – 50%	Reading – 45% Math – 35%
<b>Multi-year Averaging</b>	None for AMOs	Three-year average for AMOs	Two-year average allowed
<b>Confidence Interval</b>	Only for schools with fewer than 100 valid scores	95% z-test can be used as alternate test for all schools and subgroups	2002 -03: None 2003-04: 95% z-test can be used as alternate test for all schools and subgroups

	California	Georgia	Pennsylvania
<b>Safe Harbor</b>	10% reduction in percent Not Proficient and meets Second Indicator	10% reduction in percent Not Proficient and meets Second Indicator	10% reduction in percent Not Proficient 2003-04: 75% CI using standard error of difference in proportions
<b>Minimum Group Size</b>	100, or 50 if that makes at least 15% of enrollment	40	40
<b>Participation Rate</b>	2003-04: 2-year average allowed	2003-04: three-year average allowed	Two-year average allowed
<b>Additional Indicator</b>	API: 560+ (increases over time) or growth of at least one point	2002-03 – Attendance: <15 percent absent for more than 15 days, or improvement from previous year 2003-04: menu of options	2002-03: Attendance: 95% or improvement 2003-04: Attendance: 90% or improvement from previous year

Notes:

Data sources: California 2003-04 Accountability Workbook; Georgia 2002-03 and 2003-04 Accountability Workbook; Pennsylvania 2002-03 and 2003-04 Accountability Workbook; Decision Letters on requests to amend accountability plans from Raymond Simon, Assistant Secretary, Office of Elementary and Secondary Education, to state superintendents and boards of education, available on-line at <http://www.ed.gov/admins/lead/account/letters/index.html>; CDE 2004 Accountability Progress Report Information Guide (August 2004) - <http://www.cde.ca.gov/ta/ac/ay/documents/aprinfoguide04.pdf>; CDE 2003 Adequate Yearly Progress Phase I Report Information Supplement - <http://www.cde.ca.gov/ta/ac/ay/documents/infosup03p1.pdf>  
GA DOE Adequate Yearly Progress FY2003 Calculation Guide for School-Level Data - <http://techservices.doe.k12.ga.us/ayp/aypcalculations.pdf>; and "About AYP in Pennsylvania" - [www.paayp.cpm/about\\_ayp.html](http://www.paayp.cpm/about_ayp.html)

***NCLB Status of Schools***

Overall, the performance of schools has been improving in all three states over the past three academic years. Table 11 shows the percentage of all schools that made adequate yearly progress in each state, as well as the percentage of elementary schools and middle schools that met their annual targets for student proficiency. It should be remembered, when reviewing these numbers, that the proficiency levels are not the same

-

in the three states nor are the annual targets that must be met for a school to have made AYP. (Occasionally, contradictory information was obtained from different state sources. For example, O'Connell (2005) and California Department of Education (2006) report different percentages of schools making AYP in 2004-05.)

Three features are noteworthy in Table 11. First, in each state and for each group of schools, the percentage of schools making AYP increased between 2003 and 2004, but in most cases the patterns flattened or reversed between 2004 and 2005. Second, the percentage of schools making AYP was much lower in California than in the other two states. This occurred despite the fact that California had a lower initial annual measurable objective. Third, the percentage of middle schools making AYP was much lower than the percentage of elementary schools making AYP in all three states. (The percentages were lower still for high schools.) Schools are having a more difficult time achieving NCLB targets with older students.

If a school does not make AYP in the same subject for two years in a row, it is identified for improvement and the state or district must initiate a set of steps to help it perform better. Continuing failure to make AYP results in a school being placed into corrective action, and, after five years of identification, into restructuring. These designations entail more intensive interventions and sanctions. Table 12 shows the percentage of schools identified for improvement in the three states in 2003 to 2005.

**Table 11.**  
**Percentage of Schools Making Adequate Yearly Progress, 2002-03 to 2004-05**

	<b>2003</b>	<b>2004</b>	<b>2005</b>
<b>California</b>			
All schools	54	65	60
Elementary	67	75	60
Middle	35	44	39
<b>Georgia</b>			
All schools	64	80	82
Elementary	88	95	95
Middle	20	50	57
<b>Pennsylvania</b>			
All schools	56	84	80
Elementary	70	90	86
Middle	49	76	66

Notes:

California data come from California Department of Education (2004, 2005, 2006), O’Connell (2005), Singh (2006) and <http://ayp.cde.ca.gov/datafiles.asp>. Georgia data downloaded on March 26, 2006 from <http://public.doe.k12.ga.us/aypnclb.aspx>. Pennsylvania data downloaded on March 26, 2006 from [http://www.paayp.com/state\\_report.html](http://www.paayp.com/state_report.html).

Although Georgia had a relatively low percentage of schools failing to make AYP in 2003-04, it had a relatively high percentage of schools identified for improvement. This situation probably reflects the fact that 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. In 2005, two percent of Georgia schools were in corrective action status and three percent were in restructuring status, requiring the most severe interventions. Pennsylvania has the lowest percentage of schools identified for improvement. In 2005, four percent of Pennsylvania schools were in corrective action status. We were unable to determine the percentage of Pennsylvania schools that were in restructuring status. In California, we calculated that

16 percent of schools receiving Title I funding were in corrective action and 14 percent were in restructuring status in 2005 (this represents 10 percent and nine percent of all schools, respectively). Thus, California faces the greatest challenge with respect to intervention. (Although states must report on student performance and make information about school performance available to parents, they do not always summarize the information in similar ways or make statewide summaries readily available.)

**Table 12.**  
**Percentage of All Schools Identified for Improvement, 2002-03 to 2004-05**

	<b>2003</b>	<b>2004</b>	<b>2005</b>
California	22	28*	30*
Georgia	27	20	17
Pennsylvania	5	8	6

Notes:

California data come from California Department of Education (2004, 2005, 2006), O'Connell (2005), Singh (2006) and <http://ayp.cde.ca.gov/datafiles.asp>.

\* percent of "eligible schools." Georgia data downloaded on March 26, 2006 from <http://public.doe.k12.ga.us/aypncfb.aspx>. Pennsylvania data downloaded on March 26, 2006 from [http://www.paayp.com/state\\_report.html](http://www.paayp.com/state_report.html).

Districts are required to meet the same AMOs as schools, and districts that do not make AYP for two years in a row are identified for improvement. In all three states, the percentage of districts making AYP increased dramatically from 2003 to 2005 (see Table 13). However, a substantial percentage of districts in California and Georgia failed to make AYP in 2005.

**Table 13.**  
**Percentage of Districts Making Adequate Yearly Progress, 2002-03 to 2004-05**

	<b>2003</b>	<b>2004</b>	<b>2005</b>
California	44	59	56
Georgia	--	--	45
Pennsylvania	19	58	90

Notes:

California data come from O'Connell (2005) and Singh (2006). Georgia data downloaded on March 26, 2006 from <http://public.doe.k12.ga.us/aypnclb.aspx>. Pennsylvania data downloaded on March 26, 2006 from [http://www.paayp.com/state\\_report.html](http://www.paayp.com/state_report.html).

### **Summary**

The three states that are participating in this study approached the accountability and teacher quality provisions of NCLB in different ways, and their choices were directly related to the policy context that existed in the states prior to NCLB. In particular, the states had different accountability systems in place when NCLB was passed, and this prior experience with SBA colored their implementation decisions. California and Georgia had testing systems that complied with many of the requirements of NCLB, whereas Pennsylvania had very limited state testing. California had a functioning accountability system based on growth, which had to be modified to comply with NCLB. The states also varied in the nature and breadth of their standards. These differences and others discussed in this chapter are illustrative of the variation in SBA implementation that would have been found across all states at the time of NCLB. The differences underscore the importance of considering state context when examining the effects of SBA on districts, schools, and students. They are especially relevant to the material

-  
presented in the next chapter on educators' reactions to state standards and assessment systems.

*Remainder of the Symposium*

The remainder of the symposium will present findings from our 2004-05 interviews and surveys. The second paper, presented by McCombs and Russell, will focus superintendent, principal, and teacher understanding and attitudes toward standards-based accountability. Following that, Hamilton and Berends will describe the effects of standards and assessments on teachers' classroom practices. The fourth paper, by Marsh and Robyn, examines school improvement efforts.



## References

- California Department of Education (March, 2005). State accountability report card, 2003-04. Downloaded from website March 30, 2006.
- California Department of Education (February, 2006). State accountability report card, 2004-05. Downloaded from website March 30, 2006.
- California Department of Education (July, 2004). State accountability report card, 2003-04. Downloaded from website March 30, 2006.
- Center for Education Policy (2004). *From the capitol to the classroom: Year 2 of the No Child Left Behind Act*. Washington D.C.: Author.
- Elmore, R. F. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, 66(1), 1-26.
- Elmore, R. F., & Fuhrman, S. H. (1995). Opportunity-to-learn standards and the state role in education. *Teachers College Record*, 96, 432-457.
- Hamilton, L.S. (2003). Assessment as a policy tool. *Review of Research in Education*, 27, 25-68.
- Koretz, D. M., Linn, R. L., Dunbar, S. B., & Shepard, L.A. (1991). The effects of high-stakes testing on achievement; Preliminary findings about generalization across tests. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.

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.

No Child Left Behind Act (NCLB) of 2001 (20 U.S.C. § 6311 et seq.).

O'Connell, J. (2004). "O'Connell releases data showing most California schools improve

API scores; meet federal AYP criteria." *California Department of Education*

*News Release #04-76*. (<http://www.cde.ca.gov/nr/ne/yr04/yr04rel76.asp>).

Downloaded March 30, 2006.

O'Connell, J. (2005). "O'Connell announces significant gains in state API results, mixed

progress in federal AYP results. *California Department of Education News*

*Release #05-103*. (<http://www.cde.ca.gov/nr/ne/yr05/yr05rel03.asp>).

Downloaded March 30, 2006.

Porter, A. C., Linn, R. L., & Trimble, C. S. (2005). The effects of state decisions about

NCLB Adequate Yearly Progress targets. *Educational Measurement: Issues and*

*Practice*, 24(4), 32-39.

Rowan, B. A. (1982). Organizational structure and the institutional environment: The

case of public schools. *Administrative Science Quarterly*, 27, 259-279.

Singh, J. (2006). California Department of Education. Personal communication.

Weick, K. E. (1976). Educational organizations as loosely coupled systems.

*Administrative Science Quarterly*, 21, 1-19.