

WORKING P A P E R

Accountability Elements of the No Child Left Behind Act: Adequate Yearly Progress, School Choice, and Supplemental Educational Services

JULIE A. MARSH, HEATHER R. BARNEY, AND
JENNIFER L. RUSSELL

WR-258-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.



EDUCATION

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.

Introduction

Standards-based accountability (SBA) under the No Child Left Behind (NCLB) Act of 2001 relies in large part on a complex set of technical definitions and measures of school and district progress at bringing all students to proficiency in language arts, mathematics, and science by 2014. The law requires states, districts, and schools to achieve Adequate Yearly Progress (AYP) objectives for all students and student subgroups, and imposes a set of consequences for those failing to meet these targets. Drawing on RAND's Implementing Standards-Based Accountability (ISBA) project, this paper presents preliminary data on the interpretation and implementation of these accountability elements—AYP and the early sanctions of school choice and supplemental educational services—at the state, district, and school levels in three states.

Background

To gauge school and district progress at meeting the ultimate goal of 100 percent of students proficient by 2014, NCLB requires each state to define its own set of benchmark proficiency targets or annual measurable objectives (AMOs) to determine whether schools and districts are making Adequate Yearly Progress (AYP) toward teaching students state standards and bringing all students up to proficiency. To ensure that test results reflect the achievement of the school and district's entire student population, the law also requires that at least 95 percent of all students in each school and district take the tests. In addition, states must set one additional measure of academic progress that is factored into the determination of whether schools and districts meet AYP targets annually. Under NCLB, Title I schools failing to meet AYP for two years in a row in the same subject (i.e., those identified for improvement) are required to provide students with the option of transferring to higher performing schools, and for low-income students in Title I schools not meeting AYP for three consecutive years, the opportunity to receive free academic tutoring in the non-school hours (i.e., supplemental educational services). Districts with schools failing to meet AYP for two or more years are also required to spend up to 20 percent of their Title I funds on transportation for transferring students and on tutoring services.

The NCLB requirement that states develop standards and assessments was not a new element, but rather an expansion of previous requirements of federal law. However, several of the more technical accountability elements of AYP and related sanctions did represent dramatic changes for many states. In particular, the requirement that all student sub-populations—

including students with disabilities and English Language Learners—meet the same benchmark proficiency standards was a new development for most states and state accountability systems. The general logic of standards-based accountability embedded in NCLB is that public reporting of these disaggregated data and the threat of sanctions create added incentives for states, districts, and schools to improve the services and quality of education for all students, regardless of background. To many observers, the subgroup requirement represents the law’s greatest contribution to public education. As one organization noted, “NCLB shines a bright spotlight on these underserved students” and creates high expectations for all students (Education Trust-West, 2004).

Yet, how states define the various technical elements of the law can greatly influence the extent to which these various subgroups are in fact held accountable for high standards and the extent to which schools and districts achieve proficiency targets for all students. Research has shown that states vary widely in their interpretation of NCLB requirements and the implementation of SBA components (Erpenbach, Forte-Fast, & Potts, 2003; Linn, Baker, & Betebenner, 2002). For example, one study found numerous differences in technical definitions of AYP—ranging from how states defined the AYP denominator (e.g., including only tested students versus including all enrolled students) to what was the minimum number of students required to qualify as a subgroup (e.g., uniform definitions vs. definitions adjusted for school size). This study also noted that states with "more sophisticated" accountability systems that incorporated strategies to enhance the reliability and validity of AYP decisions appeared to have greater state capacity, such as staff available to conduct these analyses, resources, and rich data bases (Erpenbach et al., 2003). Another study demonstrated that state progress at achieving proficiency targets varied greatly, in large part due to variation in the content of state standards, the rigor of state tests, and the stringency of performance standards (Linn et al., 2002).

Data and Methods

The introductory paper in this symposium describes the design, data, and methods used in the ISBA study. The overall study focuses on elementary and middle school science and mathematics in California, Georgia, and Pennsylvania. The results presented in this paper rely primarily on superintendent and principal survey and interview data from the 2003-04 school year, as well as data from state websites for 2002-03 and 2003-04. In a few instances we also draw on data from teacher surveys, interviews conducted in case study schools, and interviews

with state policymakers. The sample sizes for the estimates reported in this paper are reported in Table 1. In California, a larger percentage of districts originally selected declined to participate than in Georgia or Pennsylvania. The smaller number of participating districts in California resulted in smaller samples of principals and teachers compared to the other two states. Schools that included both elementary and middle-level grades (e.g., K-8 schools) are included in both the middle and the elementary school samples.¹

Table 1. Sample Sizes for Each State

	California	Georgia	Pennsylvania	Total
Districts				
-agreed to participate	19	25	24	68
-superintendents responding to survey	18	20	22	60
-superintendents participating in interviews	15	12	16	43
Elementary Schools				
-agreed to participate	35	52	54	141
-principals responding to survey	27	44	50	121
-principals sampled within participating schools	361	684	619	1664
-teachers responding to survey	262	588	548	1398
Middle Schools				
-agreed to participate	28	56	42	126
-principals responding to survey	24	44	38	106
-teachers sampled within participating schools	331	838	454	1623
-teachers responding to survey	225	730	378	1333

State-specific weights were applied to all survey data to make the responses reflect the state as a whole, reflecting both the sample design for each state and the patterns of survey non-response. As a result, the statistics reported in this paper 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.

¹ Note, however, that the estimates reported here are based only on teachers in the relevant grades (e.g., the middle school estimates include only teachers in grades 6 through 8).

Research Questions

Building on existing research, this paper examines how the three study states have interpreted and implemented NCLB accountability elements of AYP and early sanctions, along with the perceived impact and attitudes about these components of SBA at the district and school levels. It addresses five broad research questions:

- How have the three states defined AYP?
- What progress have schools and districts made in achieving AYP targets?
- What specific provisions or technical elements of AYP account for schools not meeting AYP?
- What is the level of understanding, perceived validity, and reported impact of AYP at the district and school levels?
- What have been district and school experiences with the parental choice/transfer and supplemental services provisions of NCLB? To what extent are these options utilized and what are the perceived impacts?

Organization of Paper

The following sections address each of the research questions in order. First, we detail how the three states defined AYP and the extent to which the new requirements of NCLB build on existing state accountability systems. Next, we examine the way in which the AYP provisions played out at the district and school level across the three states, including: how many schools and districts met AYP, the reasons for not meeting AYP, and the reported understanding and impact of, as well as attitudes about, AYP. Finally, we analyze the extent to which the transfer and supplemental educational services options have been utilized, along with the perceived impact of these options on schools and districts.

As the following sections illustrate, many of the differences we identify across the three states stem from differences in technical issues related to AYP definitions. As such, differences do not necessarily indicate that one state is performing academically better or making more progress than another. Our intent is not to be evaluative, but instead to illustrate patterns that emerged within and across states, to suggest possible explanations, and to highlight how technical interpretations and issues influence the way in which SBA plays out at the state, district, and school levels.

Background on State Systems and AYP Definitions

State Accountability Systems Prior to NCLB

In all three states, the implementation of standards-based accountability was shaped by the existing policy environment and prior state accountability systems (for a more detailed discussion of the broader state context, please see the paper by Stecher in this symposium). Each state's history with accountability influenced both how states chose to implement NCLB and how they operate their NCLB-compliant accountability systems. For some, NCLB represented an entirely new policy; for others it confronted an existing state accountability system that had to be modified or dismantled.

California. Stemming from the Public School Accountability Act of 1999, California had implemented a standards-based accountability system for years prior to NCLB that included standards-based assessments in grades 2-8, public reporting of results, and rewards and sanctions for schools based on their ability to meet targets. Unlike NCLB, however, this system was founded on a growth model of student achievement, whereby schools were rated according to progress made on the Academic Performance Index (API) based on student test scores from one year to the next. The system also required demonstration of progress for various student subgroups, based on race/ethnicity and socio-economic status. Many of these features were incorporated into the state's system under NCLB. As we discuss below, the API was used as the additional indicator, thus retaining to some degree the state's prior gain-based approach to accountability. The overlay of the federal system onto the existing state system, nevertheless, created some confusion for stakeholders at all levels. This is a topic we return to later.

Georgia. Georgia had implemented an accountability system for the state's Title I schools in the late 1990's in response to the 1994 reauthorization of the federal ESEA. Under this system, Title I schools were required to achieve an annual eight-percent "move out" on the nationally norm-referenced Iowa Test of Basic Skills or, after 2000, the standards-based CRCT, or face being identified for improvement (CPRE, 2000). Georgia's legislature enacted a statewide accountability system in 2000 modeled on the systems in Texas and North Carolina (Georgia's A Plus Education Act - HB1187). The A+ system awarded a letter grade to each school (A-F) and included 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. In 2002-03, Georgia tested students in grades 4,

6, and 8, but expanded to grades 3-8 in 2003-04. Georgia is currently in the process of completely revising its curriculum standards, which will take effect in the 2005-2006 school year.

Pennsylvania. Prior to NCLB, the Pennsylvania legislature had passed “Empowerment Legislation” under which districts were identified for being in need of improvement if more than 50 percent of students in the district scored at the “Below Basic” level on the state test for two years. Pennsylvania’s system also placed responsibility for school improvement in the hands of districts and local boards, who had wide options for improving the school, including reopening as a charter, reconstitution, entering into a management contract, or other major restructuring (Ziebarth, 2004). Another feature of this system was the Pennsylvania Performance Index (PPI), which measures and rewards improvement across the full range of academic achievement (i.e., it is not limited solely to the proficient level). In response to NCLB, Pennsylvania transferred schools’ status from the “empowerment legislation” over into the NCLB system. They also retained the PPI as a component of the new system, as discussed below. In 2002-03 and 2003-04, Pennsylvania was testing students only in grades 5 and 8.

State Definitions of AYP

As Table 2 illustrates, the three study states varied in how they defined the requirements for schools and districts (or school systems, as they are called in Georgia) to meet Adequate Yearly Progress (AYP) goals of NCLB. 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.

Table 2. State Accountability Systems and Key Definitions (2002-03, 2003-04)²

	California	Georgia	Pennsylvania
AMO: Proficiency targets	ELA – 13.6% Math – 16.0%	ELA/Reading – 60% Math – 50%	Reading – 45% Math – 35%
Multi-year averaging	None for AMOs	Three-year average for AMOs	Two-year average allowed
Confidence interval	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
Safe Harbor	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
Minimum group size	100, or 50 if that makes at least 15% of enrollment	40	40
Participation Rate	2003-04: 2-year average allowed	2003-04: three-year average allowed	Two-year average allowed
Additional indicator	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

Annual Measurable Objectives. Definitions of proficiency targets or annual measurable objectives (AMOs) varied significantly across the three states—indicating different growth trajectory patterns to achieve the federal goal of 100 percent of students proficient by the year 2014. All three states utilized the same method for determining AMO starting points based on

² 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/aprinfo4guide04.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

schoolwide average proficiency levels from 2001-02.³ In 2002-03, this starting point in California meant that 13.6 percent of students in each subgroup had to reach proficiency in English Language Arts (ELA) 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, with the expectation that 60 percent of students in each subgroup reach proficiency in combined ELA/reading and 50 percent in math.

As Figures 1 and 2 illustrate, the overall trajectories are more gradual in Georgia and Pennsylvania compared to California. The AMO targets for reading/ELA and math start out higher in Georgia and Pennsylvania and ratchet up in three-year increments until 2010, after which the targets increase annually. In California, proficiency targets start out lower in both subjects, increase gradually in three-year increments until 2007, and then increase each year thereafter.

In 2002-03, Georgia and Pennsylvania allowed for multi-year averaging in calculating AMOs. That same year, Georgia allowed for confidence intervals to factor into these calculations for all schools and subgroups, while California allowed this only for schools with fewer than 100 valid scores. Pennsylvania added a confidence-interval provision in 2003-04.

All three states included "safe harbor" provisions in their definitions of AYP. As such, even if schools did not achieve state AMOs they could meet AYP by reducing the percent of students who were not proficient by 10 percent from the previous year (while also meeting the second indicator). In 2003-04, Pennsylvania amended its safe harbor provisions to allow for confidence interval calculations.

³ 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 20th 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.

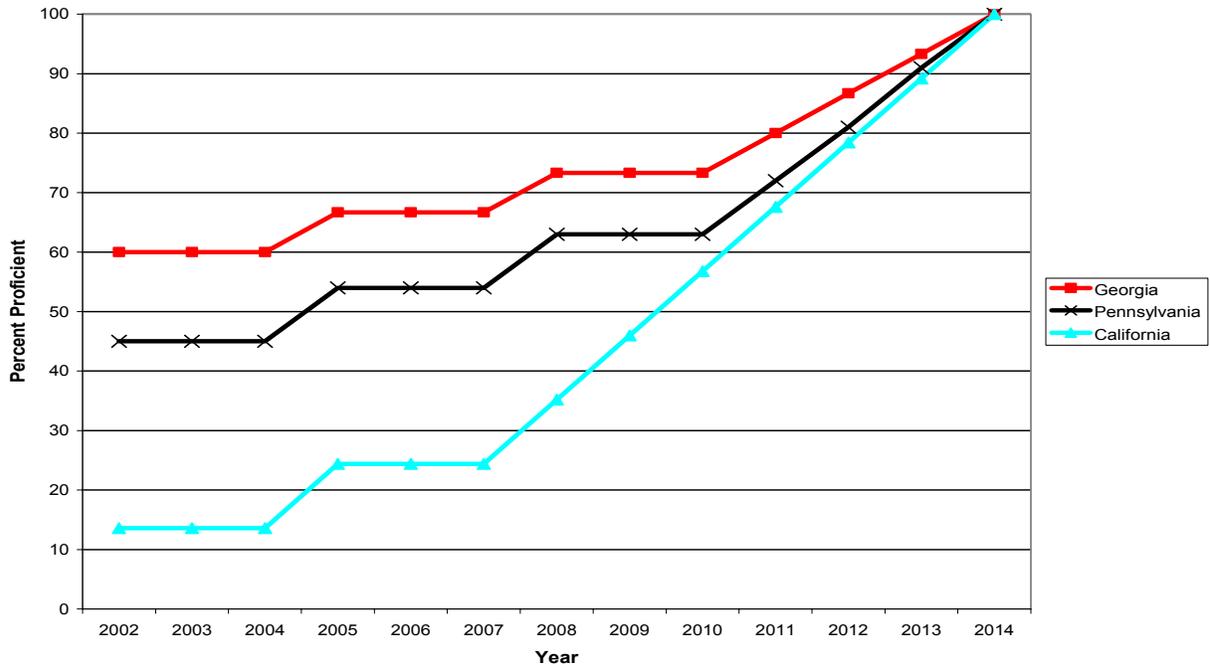


Figure 1. Annual Measurable Objectives for Reading/ELA, by State (2002-2014)

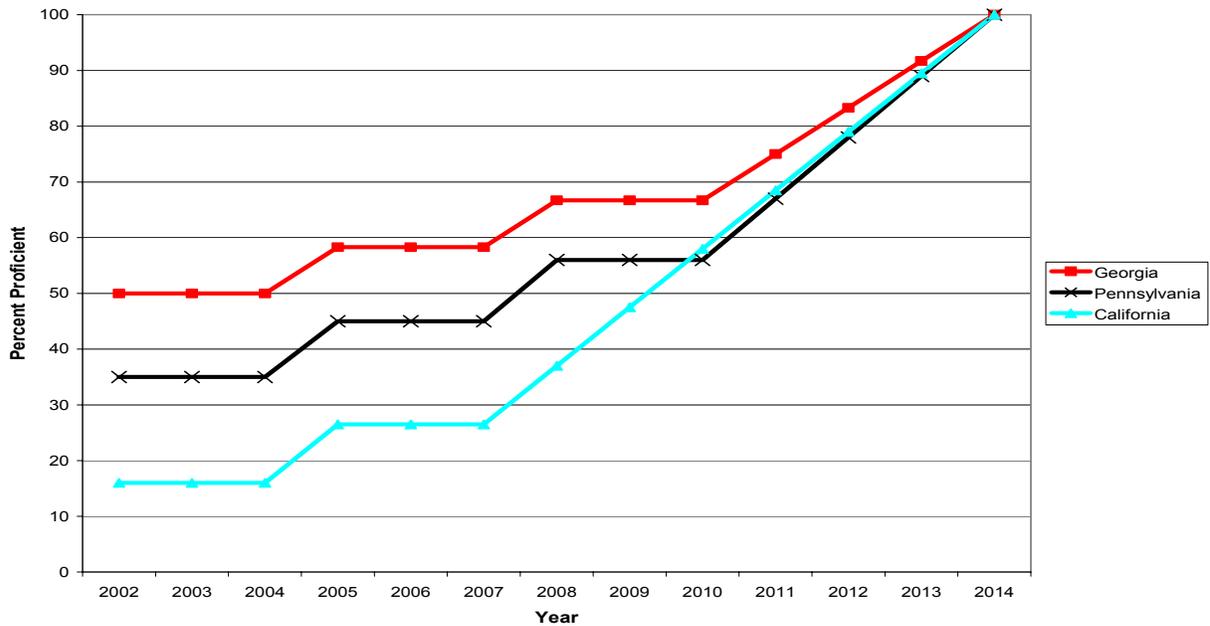


Figure 2. Annual Measurable Objectives for Math, by State (2002-2014)

Subgroups. All three states specified a similar set of student subgroups, including black, American Indian/Alaskan, Hispanic, white, Asian (California defined 3 separate subgroups, Asian, Filipino, Pacific Islander), economically disadvantaged, students with disabilities, and

limited English proficient. Georgia and Pennsylvania also added the subgroup of “multi-racial.” The states differed somewhat in their definitions of the minimum group size to qualify as a student “subgroup.” In Georgia and Pennsylvania, 40 students in any one category constituted a subgroup. California set a higher threshold, defining a minimum group size as 100 students or 50 if that number constituted 15 percent of total student enrollment in a school.

Participation Rates. While all three states started out with similar rules regarding the federal requirement that at least 95 percent of all students and subgroups of students take the state test, relaxation of federal rules allowed for slight alterations over time. By 2003-04 all three states were allowing schools to average participation rates over a two-year (California and Pennsylvania) or three-year (Georgia) period.

Additional Indicator. Georgia and Pennsylvania set attendance as the additional measure of academic progress in elementary and middle schools. In 2003-04, however, both states altered their requirements: Pennsylvania decreasing the attendance target from 95 percent to 90 percent; Georgia allowing districts to select among several options, including attendance and writing test scores, percent of students exceeding on reading, math, science, or social studies tests; and percent proficient on science or social studies tests (the choice remains in effect for three years, after which districts can select a different indicator, if they choose). California selected the state’s Academic Performance Index (API) as the second indicator, requiring schools to either achieve a schoolwide API of 560 (this number increases over time) or demonstrate improvement by at least one point over the previous year.

Appeals process. All three states specified appeals processes for schools and districts wanting to dispute initial AYP determinations based on data errors and special circumstances. In Pennsylvania, schools could meet AYP on appeal on several grounds, including demonstration of significant growth on the PPI, which, as noted above, measures growth across all performance levels, not just at the proficient level (Socolar, 2004; also, www.paayp.com/about_ayp.html).

Status, Understanding, and Perceptions of AYP

As illustrated above, the three states varied in their interpretation of elements of NCLB and AYP. These different technical interpretations and contexts affected the results and attitudes we observed both within and across the states. The following sections examine the way in which the AYP provisions of NCLB played out at the district and school level across the three states, including: how many schools and districts met AYP, the reasons for not meeting AYP, and the

reported understanding and impact, as well as attitudes about, AYP. What emerges is a pattern in which elementary school principals experienced less AYP failure and more positive attitudes about AYP than their middle school counterparts in all three states. In addition, we find that the AYP requirements have made an earlier impact on Georgia. According to our survey data, more middle schools and districts failed to meet AYP and were facing NCLB consequences, and a greater proportion of individuals reported skepticism and concerns. Much of this “failure,” however, may be due to the technical definitions and aspects of AYP and not necessarily academic reasons.

School and District AYP Status

Statewide data indicate that more than half of all schools in the three states met AYP in 2002-03. As Table 3 illustrates, a much greater proportion of elementary schools met these targets than did middle schools in all three states, although the difference is even more substantial in Georgia. It is also notable that very few districts in Georgia met AYP in 2002-03: 7 percent compared to about one-fourth in Pennsylvania and almost half in California.

Table 3: Percentage of All Schools and Districts in State Making AYP 2002-03⁴

	California	Georgia	Pennsylvania
All Schools	54	64	61
Middle Schools	33	20	43
Elementary Schools	67	89	73
All Districts	43	7	24

In all three states, approximately 10 percent of all schools statewide were identified as needing improvement (NI) based on 2002-03 AYP status (i.e., failed to make AYP for two consecutive years), as shown in Table 4. Unlike the other two states, Georgia schools had experienced more years of being identified as NI. Of the NI schools in California and

⁴ Statewide data for this table and this section were collected from the Department of Education websites for California, Georgia, and Pennsylvania accessed March 24, 2005. Middle and elementary school breakdowns for Pennsylvania were provided via personal communication with the research staff of the Pennsylvania Department of Education, March 24, 2005. In general, our school survey data reveal similar patterns to the state data, with more elementary schools making AYP compared to middle schools and fewer Georgia middle schools and districts reported to make AYP than schools and districts in California and Pennsylvania. According to the survey data, 90 percent of Georgia district superintendents reported having at least one school identified as NI, compared to only 5 percent in Pennsylvania and 60 percent in California (standard errors respectively: 8, 3, 14). Although our estimates are weighted to reflect distribution of AYP status in each state, they show more schools in California and Pennsylvania making their AYP than the actual state-wide data reflects. This slight discrepancy is worth keeping in mind when interpreting the survey data reported in the remainder of this paper.

Pennsylvania, the range of years in NI was one to four years. In Georgia, five percent of all schools were in either their fifth or sixth years of NI status. This condition exists in part because Georgia implemented its Title I accountability system earlier than the other two states.

Table 4: Percentage of All Schools Statewide Identified as Needing Improvement for 2003-04 (based on 2002-03 assessment results)⁵ by Years in Improvement Status

	California	Georgia	Pennsylvania
Year 1 ¹	7	3	5
Year 2	2	2	0
Year 3	4	1	0.3
Year 4	0.1	1	4
Year 5		3	
Year 6		2	
Making progress ²			1
Total	13	12	10

Notes:

¹Percentages included are calculated as the number of schools in each year of NI status out of a total of all schools in the state.

²Pennsylvania categorizes schools as “making progress” if they made AYP in 2002-03 but required an additional year of making AYP to move out of NI status. It is not clear what year of NI these schools fit into.

Finally, as displayed in Table 5, statewide data sources indicate that in all three states, a greater percentage of schools made AYP in 2003-04 than the previous year. Many observers believe the increase, particularly at the secondary level, was due to greater attention and efforts to boost test participation rates, as well as changes in participation and second indicator requirements. The following section further examines the reasons why schools and districts failed to make AYP and what accounts for differences observed across states, levels of schools, and time.

Table 5. Percentage of Schools Statewide Making AYP 2002-03, 2003-04⁶

	California	Georgia	Pennsylvania
2002-03	54	64	61
2003-04	65	80	81

⁵ Data for this table and this section were collected from the Department of Education websites for California and Pennsylvania accessed March 24, 2005. Georgia data were provided by Georgia Department of Education staff on November 5, 2003. Note that in public reports, California and Georgia more often report these data as a percentage of Title I schools only. As such, these data indicate that 22 percent of Title I schools in California were identified as NI, while 21 percent of Title I schools in Georgia were identified as NI. Given Pennsylvania’s unified system and decision to apply NCLB accountability requirements to all schools, this figure is not readily available or reported.

⁶ Data for this table were collected from the Department of Education websites for California, Georgia, and Pennsylvania accessed March 24, 2005.

Technical Reasons Accounting for Schools Failing to Make AYP

What aspects of the AYP requirements account for schools' inability to meet their overall AYP targets, particularly in Georgia and among middle schools across the three states? As noted above, under NCLB, schools faced targets in three areas: annual measurable objectives (AMOs), participation rate, and an additional indicator that varied by state. The following section examines which targets were most problematic for those schools not meeting AYP in the three states (estimates based on weighted data from our study's sample).⁷ As we discuss in this section, we found that:

- AMO subgroup requirements appeared to be the most consistent obstacle across the three states, particularly the students with disabilities AMO in Georgia;
- Test participation was a challenge in 2002-03, particularly in Georgia and Pennsylvania, but less so the following year due to regulation changes and greater focus; and
- Additional indicators were not an obstacle in California, but were considerable challenges in Georgia and Pennsylvania.

Table 6 provides an overview of the specific targets that schools found relatively more or less difficult to meet. As illustrated, schools most frequently missed the subgroup AMOs in all three states. There was greater variation across the states in the degree to which schools met the other targets. We examine each of these categories in more depth in the sections that follow.

Table 6. Schools Failing to Meet Specific Targets, as a Percentage of All Schools Not Meeting AYP

	California		Georgia		Pennsylvania	
	2002-03	2003-04	2002-03	2003-04	2002-03	2003-04
<i>Schools missing ...</i>						
Participation Rates	15 (7)	2 (2)	45 (11)	6 (4)	39 (9)	4 (4)
Additional Indicator	0	0	35 (9)	36 (10)	46 (11)	15 (11)
Schoolwide AMOs	60 (17)	30 (9)	11 (8)	7 (5)	26 (11)	55 (15)
Subgroup AMOs	95 (4)	100	77 (8)	81 (9)	83 (9)	100

Note: Standard errors are given in parentheses.

⁷ Statewide data for these analyses were not readily available in all three states. As a result, we utilized estimates based on our weighted school data. This school sample included any school in which a principal responded to our survey. In the future, we hope to conduct these analyses using data for all schools in each state.

Annual Measurable Objectives. As shown in Table 6, schools that failed to make AYP most often missed the AMO subgroup targets, both in 2002-03 and in 2003-04. Interestingly, schools in Pennsylvania and California also frequently missed schoolwide AMOs, much more so than schools in Georgia.

The total number of AMOs each school was required to meet varied, and ranged in theory from 2 to 22, as schools were responsible for meeting AMOs in both reading and math for the school as a whole and for any subgroup large enough to meet the state’s minimum number for AYP analysis. Table 7 shows the distribution of schools by the number of subgroups required to meet AMOs for 2002-03 and 2003-04 for each of our three states. Schools from California and, in 2003-04, Georgia were required to meet more subgroup AMOs than schools in Pennsylvania. These differences are important because schools with more subgroups have been shown to face larger odds of not meeting AYP, even after accounting for family poverty and schools’ average test scores (Novak & Fuller, 2004).

Table 7. Percentage of Schools Required to Meet Subgroup AMO’s

<i>Schools required to meet ...</i>	California		Georgia		Pennsylvania	
	2002-03	2003-04	2002-03	2003-04	2002-03	2003-04
No Subgroups	2 (2)	2 (2)	23 (6)	0	15 (6)	14 (4)
1 Subgroup	17 (7)	13 (6)	44 (5)	10 (6)	59 (9)	60 (7)
2 Subgroups	14 (6)	17 (7)	17 (5)	19 (5)	21(7)	19 (5)
3 Subgroups	25 (11)	20 (11)	12 (2)	36 (7)	3 (2)	4 (3)
4 Subgroups	29 (12)	25 (7)	3 (2)	25(6)	1 (1)	1 (.4)
5 Subgroups	9 (4)	14 (5)	0	4 (2)	1 (1)	2 (2)
6 Subgroups	5 (3)	8 (5)	1 (1)	4 (2)	0	0
7 Subgroups	0	0	0	1 (1)	0	0
8 Subgroups	0	0	0	1 (1)	0	0

Note: Standard errors are given in parentheses.

Although each individual school’s enrollment and diversity played a large role in determining the number of subgroups and AMOs that schools were responsible for, state policies such as the number of tested grades played a role as well. Georgia serves as a vivid example of the effect of testing policy. In 2002-03, the CRCT was administered only in grades 4 and 8 due to technical problems. As a result, schools’ tested populations were relatively small, and 84 percent of the schools were responsible for two or fewer subgroups. In 2003-04, however, testing expanded to all grades 3-8, which increased the number of subgroups required to meet AMOs for schools in our sample dramatically: 71 percent were now responsible for three

subgroups or more. Pennsylvania continued to test in only two grades in 2003-04, and its schools were required to meet relatively few subgroup AMOs. However, when the state test expands to all grades 3-8 as is planned for 2005-06, schools will face the added challenge of higher tested enrollment and increasing numbers of subgroup AMOs.

Subgroup AMO requirements also may help to explain why middle schools fared worse than elementary schools in all three states. Middle schools had larger, more diverse enrollments and as a result were generally responsible for more subgroups. The differences between middle and elementary schools was most apparent in Georgia, where 54 percent of middle schools were required to meet AMOs for three or more subgroups in 2002-03, compared to practically no elementary schools.

The specific subgroup AMOs that schools missed varied across our three states, as shown in Table 8. Schools in Georgia were particularly likely to miss the AMOs for students with disabilities. Though not shown in the table, in 2003-04, 97 percent (standard error: 3) of Georgia schools that missed any subgroup AMO missed the students with disabilities AMO, and less than 10 percent missed any other additional AMOs. Schools in California and Pennsylvania, on the other hand, missed a variety of different AMOs, and were more likely to miss AMOs for more than one subgroup.

Table 8. Schools Missing Specific Subgroup AMOs, as a Percentage of All Schools Not Meeting AYP

<i>Schools missing AMOs for...</i>	California		Georgia		Pennsylvania	
	2002-03	2003-04	2002-03	2003-04	2002-03	2003-04
Black/African American	23 (14)	9 (8)	14 (8)	0	14 (5)	71 (19)
Hispanic/Latino	74 (14)	73 (17)	0	0	11 (11)	31 (24)
Students with Disabilities	28 (13)	50 (21)	72 (9)	79 (9)	10 (4)	34 (16)
Econ. Disadvantaged Students	76 (12)	39 (11)	NA ⁸	7 (5)	80 (10)	55 (15)
English Language Learners	79 (15)	69 (18)	3 (3)	0	1 (1)	0
Only One Subgroup	17 (9)	31 (18)	65 (10)	77 (9)	56 (15)	21 (8)
More than One Subgroup	77 (9)	69 (18)	13 (7)	4 (4)	27 (10)	79 (8)

Note: Standard errors are given in parentheses.

Participation Rate. As Table 9 illustrates, test participation rate was a challenging target for some schools in our three states in 2002-03, particularly in Georgia and Pennsylvania. However, even in 2002-03, participation was rarely the make or break factor for schools: among

⁸ Georgia did not require schools to meet AMOs for economically disadvantaged students in 2002-03 because it did not have baseline data on this subgroup from 2001-02 that would have been needed to make safe harbor calculations.

schools that did not make AYP, participation rate was the only missed target for between 3-5 percent of schools in California and Georgia and none in Pennsylvania..

Table 9. Schools Missing One or More Participation Rate Target(s), as a Percentage of All Schools Not Meeting AYP

	California		Georgia		Pennsylvania	
	2002-03	2003-04	2002-03	2003-04	2002-03	2003-04
Schools Missing One or More Participation Rate Targets	15 (7)	2 (2)	45 (11)	6 (4)	39 (9)	4 (4)
Schools in which Participation Rate(s) was the Only Missed Target	5 (4)	0	3 (3)	0	0	0

Note: Standard errors are given in parentheses.

In 2003-04, all three states took advantage of the new flexibility offered by the U.S. Department of Education that allowed for multi-year averaging of participation rates and permitted students to be excluded from the testing enrollment count in cases of serious medical emergency. The states took other steps to improve schools' participation rates, as well. For example, Georgia convened a special task force on test participation to draft proposed changes to state law (Robelen, 2004). Also, several of our case study schools reported concerted efforts and investments to boost participation, such as instituting raffles for prizes on the day of state testing as a way to entice students to attend. Most schools also made strong efforts to communicate the importance of testing and test participation to parents. These policy changes and other efforts seem to have paid off in 2003-04, as less than 10 percent of schools that missed AYP in all three states missed the participation rate targets in the second year of NCLB implementation.

Additional Indicator. As Table 10 shows, practically no California schools missed the additional indicator, the state's API, for either 2002-03 or 2003-04. In 2002-03, however, the additional indicator proved challenging for schools failing to meet AYP in Georgia and Pennsylvania, where 35 percent and 46 percent of schools respectively missed their target. When interviewed, many principals and superintendents in these two states complained about the difficulty of meeting attendance goals, noting that they had little control over whether students came to school.

**Table 10. Schools Missing Additional Indicator Target,
as a Percentage of All Schools Not Meeting AYP**

	California		Georgia		Pennsylvania	
	2002-03	2003-04	2002-03	2003-04	2002-03	2003-04
Schools Missing Additional Indicator	0	0	35 (10)	36 (10)	46 (11)	15 (11)
Schools in which Additional Indicator was the Only Missed Target	0	0	12 (7)	19 (9)	7 (6)	0

Note: Standard errors are given in parentheses.

Following revisions to standards for the additional indicator in 2003-04, Pennsylvania schools fared better: only 15 percent missed the additional indicator target. Despite changes to the definition of additional indicators in Georgia in 2003-04, this target remained a challenge in the state.

AYP Understanding, Impact, and Attitudes

What was the impact of AYP status and the requirements on individuals at the district and school levels? The following sections examine individual understanding of and attitudes about AYP. Overall, we found that individuals most affected by the AYP provisions and having failed to meet those requirements—most notably individuals in Georgia and middle school principals across the three states—were more skeptical about the prospects of achieving AYP, the validity of these measures, and the impact of AYP on staff.

Understanding of AYP and NCLB in General. Despite the complexities of state definitions and technical requirements, the majority of superintendents and principals in all three states reported having a strong understanding of AYP criteria and accountability requirements. As Table 11 displays, these views were consistent within and across states at the district and school leadership levels, and across both levels of schooling. In addition, principals in all three states reported receiving assistance to understand the state systems. Interviews with state policymakers confirmed that state departments of education invested in communicating state rules and regulations (e.g., via memos, email, websites), and that a lot of the state’s technical assistance (e.g., regional meetings) to schools focused on clarifying NCLB requirements.

Yet, our data also indicate that there was less understanding among teachers. Approximately half of the teachers at both levels across all three districts expressed confusion about the accountability system. During case study visits, teachers often reported not understanding technical requirements of AYP calculations.

Table 11. Percentage of Superintendents, Principals, and Teachers Agreeing to Statements of Understanding and Not Understanding

	California	Georgia	Pennsylvania
I have a clear understanding of AYP criteria			
Superintendents	80 (14)	77 (13)	100
Middle School Principals	90 (5)	89 (5)	99 (1)
Elementary School Principals	78 (9)	97 (2)	95 (4)
The district and/or state helps me to understand the state accountability system requirements			
Middle School Principals	92 (6)	77 (8)	88 (6)
Elementary School Principals	85 (10)	84 (5)	83 (7)

The state's accountability system is so complicated it is hard for me to understand			
Middle School Teachers	55 (4)	46 (2)	54 (4)
Elementary School Teachers	54 (5)	50 (3)	56 (3)

Note: Standard errors are given in parentheses. Survey response options for the items reported above include strongly agree, agree, disagree, and strongly disagree. The percentages in the table represent the combined responses, agree and strongly agree.

Reported Confidence in Meeting AYP. Compared to understanding of AYP, we found greater variation in the level of confidence leaders expressed in their district and school's ability to achieve AYP targets in the near and far term. In particular, Georgia superintendents were far more skeptical than their counterparts in other states about the prospects of meeting AYP in the coming year. As illustrated in Table 12, only half of the superintendents agreed that their district could attain AYP targets for 2003-04, compared to 85 percent in California and 74 percent in Pennsylvania.

Yet, when considering the relative number of districts failing to make AYP in each state, superintendent accounts in all three states are surprisingly optimistic. For example, 44 percent (standard error: 16) of Georgia superintendents failing to make AYP in 2002-03 felt they could make this target the coming year. Similarly, a sizeable number of superintendents in the other two states who missed AYP the past year reported confidence in attaining the targets the next year. One might speculate that many superintendents felt they had learned how to overcome some of the obstacles to participation and second indicators or would benefit from increased flexibility in the regulations of these targets instituted by the state and federal governments.

Table 12. Percentage of Superintendents and Principals Agreeing to Statements of Confidence in Meeting AYP

	California	Georgia	Pennsylvania
My district/school can attain the AYP targets in the 2003-2004 school year			
Superintendents	85 (13)	50 (14)	74 (11)
Middle School Principals	77 (10)	69 (9)	89 (6)
Elementary School Principals	89 (10)	97 (2)	89 (6)
My district/school can attain the AYP targets for the next five years			
Superintendents	35 (13)	45 (13)	40 (13)
Middle School Principals	56 (8)	65 (9)	44 (13)
Elementary School Principals	44 (13)	80 (6)	55 (10)

Note: Standard errors are given in parentheses. Survey response options for the items reported above include strongly agree, agree, disagree, and strongly disagree. The percentages in the table represent the combined responses, agree and strongly agree.

Nevertheless, almost all of the administrators at all levels and across the three states expressed greater skepticism about attaining AYP targets in the next five years compared to 2003-04. This drop in the level of confidence was strongest in California and Pennsylvania. In Pennsylvania, the greatest concern expressed in interviews was that as more grades were tested and schools became responsible for meeting AMOs for an increasing number of subgroups, AYP goals would be more difficult to attain. In all states, superintendents interviewed also frequently noted that moving beyond 80 percent proficiency was simply unrealistic—particularly in schools with larger Special Education and ELL populations.

Perceived Validity of AYP. Many administrators across the three states expressed doubts about the validity of AYP as a measure of student performance. As Table 13 displays, these doubts were strongest among superintendents in Georgia, where only 14 percent agreed that their district's AYP status accurately reflected the overall performance of students in the district. Middle school principals in all three states also were considerably less likely to agree about the accuracy of AYP than their elementary school counterparts.

Table 13. Percentage of Superintendents and Principals Agreeing That District/School AYP Status Accurately Reflects Overall Performance of Students

	California	Georgia	Pennsylvania
Superintendents	48 (14)	14 (8)	27 (11)
Middle School Principals	43 (11)	32 (7)	22 (9)
Elementary School Principals	56 (14)	68 (6)	44 (11)

Note: Standard errors are given in parentheses.

These doubts about AYP are in stark contrast with the perceived validity of state tests. As Figure 3 illustrates, in all three states, superintendents were more likely to agree that state tests were an accurate measure of student achievement compared to AYP. These differences were particularly strong in Georgia, and then California, where the vast majority of superintendents perceived state tests to be valid and less than half perceived AYP to be valid. Interestingly, the perceived validity of both test scores and AYP was very low in Pennsylvania. These same state patterns also emerged at the middle school level, where once again principals in Georgia and California viewed test scores as more accurate measures of performance than AYP, but Pennsylvania principals questioned the validity of both.

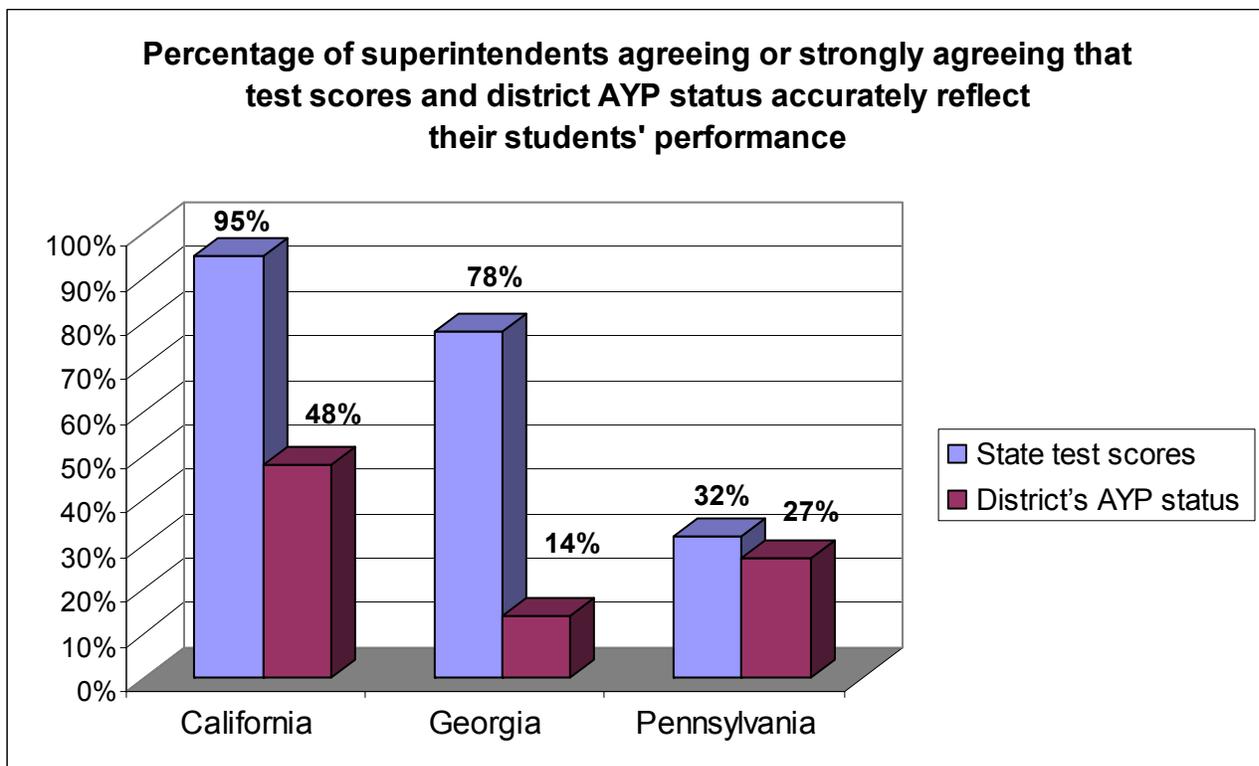


Figure 3. Perceived Validity of State Test Scores versus District AYP Status⁹

Several factors may help explain these differences. One possible explanation for the low perceived validity of tests *and* AYP in Pennsylvania is that, as noted earlier, schools were more likely to miss schoolwide AMOs, which reflect test scores for all students. Thus, administrators in Pennsylvania may have felt the tests were either too demanding or perhaps didn't yet reflect

⁹ Standard errors for items on state test scores: California (5), Georgia (10), and Pennsylvania (11). The standard errors for the items on AYP status are found in Table 13.

the curriculum in place in schools. They may have also had concerns about the state-determined cutoff points for proficiency. In fact, interviews with state policymakers indicated that there was considerable debate within the state regarding whether the proficiency levels on the PSSA were appropriate—many considered them too difficult.

In contrast, Georgia and California schools were more likely to miss AYP due to the subgroup AMOs, particularly students with disabilities in Georgia and ELLs and Latinos in California. Many superintendents interviewed in these states felt that AYP targets were unrealistic for certain subgroups. Specifically, many superintendents believed that not all special education students should be held to the same standards and that including them in AYP calculations presented a skewed measure of school performance.¹⁰ As such, schoolwide test results for regular education students would be viewed as a more accurate indicator of schoolwide performance. Further, unlike the tests alone, AYP calculations factored in a second indicator and participation rates that many considered peripheral and invalid measures of performance. For example, one superintendent complained that a middle school demonstrating the most academic progress in the district failed to make AYP because they did not test enough students. This was viewed as an example of how AYP was less valid than the scores alone. In addition, many superintendents interviewed felt that AYP judgments about performance were inadequate because they failed to account for differences in student populations (e.g., one felt it was unfair to compare a school where most parents had never attended high school to a school with college graduates) and for progress made over time.

This latter concern was strongest among individuals in California. Although, nearly half of superintendents viewed AYP as valid, many of those interviewed voiced a preference for the state's accountability metric, the API, because it factored in growth over time. Further, many individuals in California expressed confusion over the differences in AYP vs. API, noting the cognitive dissonance for staff and parents when their school was deemed to be performing well by meeting the state API goals but then failed to make federal AYP targets. This further undermined the perceived validity of AYP. In fact, the California State Superintendent of Public Instruction articulated this point in an August 2004 statement: “The fact that 317 of our schools grew 30 points or more [on the API], yet failed to make the federal benchmark, illustrates why I

¹⁰ A national study of district officials similarly found widespread concerns about the appropriateness and feasibility of test score requirements for students with disabilities and ELLs (Pinkerton, Kober, Scott, & Buell, 2003; Center for Education Policy, 2005).

believe a growth model of accountability such as we have here in California more accurately reflects actual student learning” (O’Connell, 2004). According to one report, 2,339 schools statewide met API but did not make AYP in 2003-04 (Education Trust-West, 2004).

Reported Impact of AYP. According to surveys and interviews, one of the most significant perceived benefits of AYP, and NCLB more broadly, was the impetus to improve schools’ and teachers’ focus on improving student performance. As displayed in Table 14, more than half of all principals and superintendents in all three states agreed that because of the pressure to meet AYP targets, they and their staffs were focusing more on improving student achievement. However, these views were even more prevalent in Georgia and Pennsylvania compared to California. For example, almost all of Georgia and Pennsylvania superintendents agreed with this statement, compared to two thirds in California.

**Table 14. Percentage of Superintendents and Principals Agreeing That:
*“Because of the pressure to make AYP, my staff and I are focusing more on improving student achievement”***

	California	Georgia	Pennsylvania
Superintendents	67 (14)	91 (7)	93 (6)
Middle School Principals	54 (16)	93 (3)	90 (5)
Elementary School Principals	62 (10)	85 (7)	87 (5)

Note: Standard errors are given in parentheses.

One possible explanation for the different pattern observed in California is the fact that the state’s accountability system implemented prior to NCLB already motivated staff to focus on student achievement. For several years prior to NCLB, schools faced pressure to raise test scores of all students and students in various subgroups, and potential rewards and consequences for their ability to do so. As a result, NCLB did not impose an entirely new set of incentives, as they may have in Georgia and Pennsylvania, which for the first time under NCLB were being asked to account for the performance of student sub-populations. These views were expressed fairly consistently in our superintendent and case study interviews. In high-performing California schools in particular, AYP was seen as less important or motivating because these schools were able to meet the academic AYP requirements fairly easily. In contrast, staff in these same schools found the state’s API metric more motivating because it required demonstration of growth.

Despite the slight differences noted above, the majority of administrators across the states did report that they focused more on improvement achievement as a result of AYP. Even if their

school or district did not make AYP and/or they did not buy-in to the validity of AYP, they generally reported that it motivated them to focus on student achievement. Herein lies a potential conflict with the underlying theory of action embedded in NCLB. Under NCLB, it is believed that the threat of sanctions and the ability of the public to regularly view the status of student performance disaggregated by subpopulations will motivate staff to focus on and improve the performance of all students. While our survey data do indicate signs of motivation, one might question how meaningful and sustainable this motivation is if individuals doubt the validity of the measures and targets on which these determinations are based. In fact, several superintendents interviewed expressed concerns about their ability to motivate staff in the future given the widespread perception that AYP targets over time are unrealistic. They doubted their ability to continue to sustain momentum once 70 to 80 percent of students had reached proficiency, fearing that complete mastery of content and skills was simply impossible for all subgroups. As one superintendent explained, “reaching 100 percent proficiency by 2014 is nearly impossible for this district and therefore it is difficult to motivate the staff to try. I'm concerned that the bar is set to high.” Another superintendent expressed concerns about motivating staff working with certain subpopulations: “I'm very worried that schools that are working hard with poorer children are going to look like they are failing and schools that are working not very hard with rich children are going to look like successes. And that's nothing new, but it's demoralizing for those working with poor children.”

Early AYP Sanctions: School Choice and Supplemental Educational Services

NCLB requires states to impose a set of escalating sanctions on Title I schools that fail to make AYP targets. Schools failing to attain the statewide target for two consecutive years (i.e., NI schools) must implement an improvement plan and offer parents the option of transferring their children to better-performing schools and use up to 20 percent of their Title I funds to provide transportation. After a third year of failure, schools also must offer supplemental educational services (SES) from approved providers (which can include private organizations) to students of low-income families. NCLB defines supplemental educational services as “additional academic instruction designed to increase the academic achievement of students in low-performing schools.” In addition, SES must be provided outside of the regular school day. After the fourth year, districts can take “corrective actions” against schools that continue to fail such as instituting new curricula, replacing staff, or changing governance. Schools that do not

meet the target level of achievement for five years are restructured. Title I schools that were identified for improvement under the prior law (the 1994 reauthorization of ESEA) carry their status with them under NCLB. NCLB also establishes a corrective action structure for districts and states that fail to meet AYP targets. This portion of the paper focuses on school and district experiences with the NCLB transfer and supplemental service provisions.

School Choice

This section examines the utilization of the transfer option in California, Georgia and Pennsylvania, possible reasons for the underutilization of transfers, and the impact of transfers on schools and districts. Overall, very few students took advantage of the transfer option in the three states, though schools and districts in Georgia had more experience with transfers than schools in California and Pennsylvania. Thus, the findings related to the impact of transfers over-represent the experience of Georgia educators.

Utilization. Overall, superintendents and principals in Georgia report more experience with the transfer component of NCLB than those in the two other states: superintendents and principals in Pennsylvania in particular reported very little experience with student transfers under NCLB. As Table 15 illustrates, the vast majority of Georgia superintendents reported having schools in their district that were required to allow students to transfer to another school under NCLB. This compared to 60 percent of superintendents in California and 5 percent of superintendents in Pennsylvania. A similar proportion of Georgia elementary school principals and California elementary principals reported having students eligible for transfer, but this was far greater than the percent eligible in Pennsylvania. However, a greater proportion of middle school principals in California report having students eligible for transfer than middle school principals in Georgia, and far more than in Pennsylvania.

Table 15: Superintendent and Principal Reports on Eligibility for Student Transfers

	California	Georgia	Pennsylvania
Percentage of districts/schools with eligible students			
Districts	60 (14)	90 (8)	5 (3)
Middle Schools	36 (12)	23 (7)	7 (3)
Elementary Schools	27 (13)	38 (7)	5 (3)

Note: Standard errors are given in parentheses.

Of the districts reporting students eligible for school choice, most superintendents reported experience with at least some students transferring: 4 out of 5 in California, 10 out of 12 in Georgia, and 2 out of 3 in Pennsylvania. However, as Table 16 shows, almost all of the principals surveyed in California and Pennsylvania reported that no students transferred in or out of their school under NCLB. In contrast, at least some eligible students exercised their transfer options in Georgia, where 40 percent (standard error: 8) of elementary school principals and 32 percent (standard error: 9) of middle school principals reported having experience with at least some students transferring in or out of their schools. Of the superintendents interviewed, none reported receiving students through inter-district transfer suggesting that the transfers taking place in Georgia are happening within district boundaries.

Table 16: Principal Reports on Experience with Student Transfers

	California	Georgia	Pennsylvania
Schools in which no students transferred in or out			
Middle Schools	99 (1)	68 (9)	89 (6)
Elementary Schools	92 (7)	60 (8)	90 (5)

Note: Standard errors are given in parentheses.

Despite the fact that a sizeable number of Georgia principals and superintendents in all three states report some experience with transfers, overall a very small proportion of eligible students are taking advantage of the transfer option in any of the three states. Data submitted to the U.S. Department of Education by each state and recently reported in *Education Week* suggest very low utilization rates (Olson, 2005). Nationwide, an average of one percent of eligible students transferred to a higher performing school in 2003-04: 0.3 percent in California, 0.6 percent in Pennsylvania, and 0.9 percent in Georgia. These findings are consistent with those of other studies, which found underutilization of the NCLB transfer provision (Buell et al., 2004; Brown, Lopez-Medina & Reed, 2004; Century Foundation, 2004; Neil et al, 2004).

Reasons for Underutilization. This begs the question, why are more students not taking advantage of the transfer option? While we cannot definitively answer this question due to the limitations of the study design, we can provide some speculative data drawing on interviews with superintendents, interviews and focus groups with school staff and parents in case study schools, as well as findings from prior studies. Of the 20 superintendents interviewed who were required to offer transfers—a disproportionate number of whom were from Georgia (11)—all reported

underutilization of the transfer option. The most common explanation for underutilization was that parents and students were satisfied or happy with the quality of their existing schools and did not see the need to transfer. The next most common explanations, in order, were:

- a lack of choices of schools to which students can transfer;
- a preference for the location of their current school and a desire not to travel;
- the fact that many parents had already taken advantage of existing choice programs in the district (e.g., magnet schools); and
- strong marketing that convinced parents/students to remain in their existing school.

Interviewees also commented on the lack of timeliness in presenting the transfer option to parents. For example, one superintendent reported that they do not know which schools make AYP until late August, leaving little time to notify parents of their options and to figure out transportation. In fact, surveys indicate that more than 75 percent of superintendents in Georgia and Pennsylvania (standard errors: 14 and 18 respectively) and about half in California (standard error: 24) found that it was difficult to inform parents of their transfer choices because they did not receive information from the state in a timely manner.

Further, interviews with school staff and parents at case study schools confirm that the transfer option is rarely exercised. Reasons given for the underutilization of transfers included parents wanting their children to stay in their neighborhood school and schools making an effort to keep their students. In addition, the small number of parents who were interviewed suggested that they judge their children's schools on their personal experiences with the school, not test scores published in the newspaper.

These findings echo those of other researchers who have investigated school choice under NCLB. In a study of transfer policies in 10 urban districts, researchers at the Harvard Civil Rights Project found that parents whose transfer requests were approved often chose to keep their children in neighborhood schools although the study did not identify why parents made this choice (Kim & Sunderman, 2004). Finally, the Center for Education Policy's study of NCLB again revealed underutilization of the transfer option and analysis of case study districts suggested parents preferred to keep their children in their home school where they were more comfortable.

While our findings and these studies suggest underutilization is primarily due to lack of parent and student interest in changing schools, other studies stress the role of state and district

practices. Brown, Lopez-Medina and Reed (2004) reported that many school districts only partially complied with NCLB requirements to notify parents of their options. In addition, parents lacked viable within-district options and other districts were not required to accept transfer students. One superintendent interviewed for our study explained that the district had no incentive to accept children from outside of the district: the community pays taxes to educate their own children and wants to spend whatever limited money they have on them. This is consistent with other interviews indicating that no districts reported receiving students from other districts through the NCLB transfer provision.

Reported Impact of Transfers. Our survey asked principals and superintendents about the impact of student transfers on their school or district. Caution should be taken when interpreting these findings since the number of respondents is small (especially in California and Pennsylvania), because only superintendents and principals who had experiences with transfers were asked about their impact. Overall, the California superintendents and principals interviewed were more positive about NCLB transfer requirements. A smaller proportion of superintendents and principals reported that transfers made it difficult to plan for next year, or increased pressure on staff to improve student performance. California principals and superintendents were also more likely than their counterparts in the other two states to report that the transfer policy had no noticeable effect on their school/district. Superintendents in California less frequently reported that it was difficult to inform parents of choices. In contrast, principals in Georgia middle schools were the most likely to report that the transfer policy had a noticeable effect on their school, and specifically that it increased pressure on staff to improve student performance. Surprisingly, all of the districts in Pennsylvania with schools offering transfers reported that the provision created a large administrative burden. Superintendents in California and Georgia were half as likely (standard error: 24 and 16 respectively) to report an administrative burden. Finally, principals reported very few actions taken at the school level in response to the transfer policy with one exception: more than half of the middle school principals (standard error: 16) in Georgia reported increasing class size due to student transfers.

While our sample districts and schools reported limited experiences with transfers as of the 2003-04 school year, this may change over time as an increasing number of schools advance in years under NI status. Given this possibility, some interview respondents who were not yet

required to offer transfers expressed concerns about the future. Three superintendents reported concerns about providing transportation should the district be required to offer transfers in the future. This was seen as especially problematic in districts that did not have buses. In addition, superintendents in two small districts feared not having enough options available for students should the district be required to offer transfers in the future.

Supplemental Education Services

Similar to the findings related to student transfers, more schools and districts in Georgia were required to offer students supplemental educational services (SES) than in California or Pennsylvania. This section discusses eligibility for and utilization of supplemental educational services for all three states, but draws more heavily on the comments of Georgia respondents to examine the perceived quality of and attitudes toward the SES provision of NCLB.

Utilization. As Table 18 shows, a much larger proportion of Georgia superintendents reported having students eligible for supplemental educational services: almost all, compared to less than half of superintendents in California and only a few in Pennsylvania. Superintendents and principals in Pennsylvania were far less likely to report having eligible students.

Table 18: Percentage of Superintendents and Principals Reporting Students Eligible for SES

	California	Georgia	Pennsylvania
Superintendents	39 (16)	83 (10)	4 (2)
Middle School Principals	38 (10)	28 (9)	7 (3)
Elementary School Principals	10 (9)	31 (7)	10 (5)

Note: Standard errors are given in parentheses.

Again, data from the U.S. Department of Education, recently reported in *Education Week*, suggest that underutilization of SES is a nation-wide trend (Olson, 2005). The average percent of eligible students who received supplemental services in 2003-04 was 11 percent. Georgia reported 14 percent of eligible students receiving SES compared to 7.0 percent in California (data was not available for Pennsylvania). However, utilization rates for SES are notably higher than for transfers: an average of 11 percent nation-wide compared to 1 percent.

Adding to the underutilization of SES due to low enrollment, across all three states more than 25 percent of principals reported that many students enrolled in SES do not regularly attend. This reported low attendance was stronger and more consistent in Georgia, where more than 60

percent (standard error for elementary and middle school principals respectively: 14 and 17) of Georgia principals interviewed at both the elementary and middle school levels reported that many enrolled students did not attend. This is similar to the results for California elementary schools, where 74 percent (standard error: 0) of the principals interviewed reported a lack of regular attendance.

Reasons for Underutilization. While the surveys did not ask superintendents and principals to speculate as to why all eligible students did not take advantage of SES, interviews with superintendents, as well as staff interviews and parents focus groups in case study schools provided some potential insight. About two-thirds of superintendents interviewed reported that they were not required to offer SES to any students in their district. The most common explanation for underutilization given by superintendents with eligible students was that parents and students preferred to utilize existing tutoring programs on campus: about half reported this. This is similar to comments made by staff interviewed at the case study schools, suggesting schools and districts were already providing tutoring services to students in need of extra assistance. Closely tied to this first explanation is a belief by some that parents chose not to take advantage of SES because they were not in a convenient location and would require travel. Other explanations for underutilization of SES included: lack of transportation to SES, inconvenient time, inadequate options, student/parent resistance or lack of interest, and lack of parent understanding.

Perceived Quality of SES. Our surveys asked superintendents and principals to report on their impressions of the quality of SES available in their district. Due to low response rates to these questions in California and Pennsylvania, coupled with limited experience with SES in these states, we report findings only for Georgia. While 63 percent of Georgia superintendents agreed that they were general satisfied with the quality of SES available, they expressed specific concerns about SES effectiveness: more than half (53 percent) reported that SES were not conveniently located and even more (72 percent) believed SES were not effective in meeting the needs of English Language Learners and students eligible for Special Education services (66 percent). A few superintendents interviewed reported concerns about coordination between SES and the district's regular curriculum/instruction. One superintendent reported that the content delivery models used in SES programs and in students' regular classrooms were inconsistent. Another superintendent reported concerns about the alignment of supplemental services with the

district's standards-based curriculum and continuity with students' regular instruction. This is consistent with the findings of a study conducted by the Harvard Civil Rights Project that concluded that most supplemental services are not well coordinated with the classroom curriculum and that there are few mechanisms in place for providers to communicate with classroom teachers (Sunderman and Kim, 2004).

Conclusion

In conclusion, the data analyzed for this paper point to a number of issues about the implementation of AYP provisions and sanctions for schools failing to meet AYP that warrant further attention and ongoing consideration. First, technical aspects of the law are making a significant impact on whether or not schools make AYP. For example, state choices about which additional indicator to include for AYP calculation seem to impact schools' success in meeting AYP. California schools met their additional indicator (an API of 560 or growth of 1 point) with relative ease compared to Pennsylvania and Georgia which selected attendance rates or test scores. In addition, the number of tested grades in each state impacts the number of subgroup targets schools face. Consequently, the variations in the technical components of state accountability systems under NCLB make cross-state comparison particularly problematic. While not ruling out real differences between the achievement of students in different states, differences in the technical aspects of state systems – including potential differences in the rigor of state tests and annual AMO targets – should be kept in mind as national attention to standards-based accountability naturally results in comparisons of state performance.

Second, our data reveal that the subgroup AMO for students with disabilities present significant challenges for schools. A national study of district officials found widespread concern about the appropriateness and feasibility of test score requirements for students with disabilities and ELLs (Pinkerton et al., 2003). This situation may be exacerbated by the fact that some educators in our study expressed concern that supplemental educational services were not likely to meet the needs of special education students. Further research should examine the specialized needs of students with disabilities and the supports and resources made available to schools and districts to help them achieve proficiency among these subgroups.

Third, Georgia's experience with AYP and related sanctions may provide insight into the issues that California, Pennsylvania and other states may face in the future due to the fact that Georgia implemented their Needs Improvement program earlier under the 1994 authorization of

ESEA and as a result has more schools in advanced stages of NI. This study found that educators in Georgia were more likely to have experience with sanctions such as school choice and SES. There are signs that schools and districts in other states may soon have more schools facing sanctions. Our survey data reveal that superintendents and principals across the three states were skeptical about their ability to continue to meet rising AYP standards. Pennsylvania schools may face more difficulty meeting AYP as they begin testing students in more grades and as a result have to meet more subgroup AMO targets. And national projections indicate that many more schools and districts will be identified for NI status as NCLB targets continue to escalate (Linn, Baker & Betebenner, 2002). Therefore, continuing to monitor the experiences of educators in Georgia may provide insight into what educators in other states will face in the future.

Finally, underutilization of transfers and supplemental educational services may be symptomatic of lack of public knowledge about NCLB. This suggests that as parents become increasingly aware of their rights related to school choice and SES, that schools and districts may face increased demands for these services and increasingly severe administrative and financial burdens. However, the data also suggest an alternate scenario: specifically that some underutilization may be due to the fact that transfers and/or supplemental services are not desirable to parents. If parents are not motivated to seek transfers and/or supplemental services this may remove a critical leverage point or incentive embedded in NCLB's underlying theory of action. Therefore, utilization of transfers and supplemental educational services should continue to be monitored.

Our findings suggest several areas for further research. For example, research should examine the extent to which supplemental educational services complement the curriculum and instruction students receive during the regular school day. More generally, there is a need to examine whether school choice and supplemental service provisions relate to trends in student achievement. In other words, do students receiving SES or transferring to higher performing schools perform better academically? Future research on NCLB more broadly might also examine the long-term effects of educators' doubts about the validity of AYP on their motivation to improve teaching and learning for all students. Finally, as more districts become identified as needing improvement over time, research will be need to understand the effects of these consequences and the role that states play in bringing about improvement.

References

- Brown, C., Lopez-Medina, L., & Reed, R.A. (2004). *Choosing better schools: A report on student transfers under the No Child Left Behind Act*. Washington, DC: Citizens' Commission on Civil Rights.
- Buell, B., Kober, N., Pinkerton, E., & Scott, C. (2004). *From the capital to the classroom: Year 2 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.
- Center on Education Policy. (2005) *From the capital to the classroom: Year 3 of the No Child Left Behind Act*. Washington, DC.
- The Century Foundation. (2004). *Can separate be equal? The overlooked flaw at the center of No Child Left Behind*. New York: The Century Foundation.
- Consortium for Policy Research in Education (CPRE). (2000). "Georgia Assessment and Accountability Profile." *Assessment and Accountability in the Fifty States: 1999-2000*. Philadelphia: Author.
- Education Trust-West. (2004). *Achievement in California: How is our progress?* Oakland, CA: Author.
- Erpenbach, W., Forte-Fast, E., & Potts, A. (2003). *Statewide educational accountability under NCLB: Central issues arising from an examination of state accountability workbooks and U.S. Department of Education reviews under the No Child Left Behind Act of 2001*. Washington, D.C.: Council of Chief State School Officers.
- Kim, J., & Sunderman, G. L. (2004). *Does NCLB provide good choices for students in low-performing schools?* Cambridge, MA: The Civil Rights Project at Harvard University.
- Linn, R.L, Baker, E. L., & Betebenner, D.W. (2002). "Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001." *Educational Researcher*, 31(6), 3-6.
- Neill, M., Guisbond, L., Schaeffer, B., Madden, J., & Legeros, L. (2004). *Failing our children: How "No Child Left Behind" undermines quality and equity in education*. Cambridge, MA: FairTest.
- Novak, J. R., & Fuller, B. (2003). *Penalizing Diverse Schools? Similar test scores, but different students, bring federal sanctions*. Berkeley, CA: Policy Analysis for California Education.
- O'Connell, J. (2004, August 31) "O'Connell releases data showing most California schools improve API scores; meet federal AYP criteria," press release. Sacramento, CA: California Department of Education.

Pinkerton, E., Kober, N., Scott, C. & Buell, B. (2003). *Implementing the No Child Left Behind Act: A first look inside 15 school districts in 2002-03*. Washington, DC: Center on Education Policy.

Robelen, E. W. (2004). "Schools Seek Participation on Test Days." *Education Week*, March 31 2004.

Socular, P. (2004). "More schools making AYP: What does it mean?" *Philadelphia Public School Notebook*, Fall 2004 Edition (www.thenotebook.org/editions/2004/fall/ayp.htm)

Sunderman, G. L., Kim, J. (2004). *Increasing bureaucracy or increasing opportunities? School district experience with supplemental educational services*. Cambridge, MA: The Civil Rights Project at Harvard University.

Ziebarth, T. (2004, December). *StateNotes: State policies for school restructuring*. Denver, CO: Education Commission of the States.