



## Executive Summary

The intent of the No Child Left Behind (NCLB) Act of 2001 is to hold schools accountable for ensuring that all their students achieve mastery in reading and math, with a particular focus on groups that have traditionally been left behind. Under NCLB, states submit accountability plans to the U.S. Department of Education detailing the rules and policies to be used in tracking the adequate yearly progress (AYP) of schools toward these goals.

This report examines Georgia’s NCLB accountability system—particularly how its various rules, criteria, and practices result in schools either making AYP or not making AYP. It also gauges how tough the Georgia system is compared with other states. For this study, we selected 36 schools from various states around the nation, schools that vary by size, achievement, and diversity, among other factors, and determined whether each would make AYP under the Georgia system as well as under the systems of 27 other states. We used school data and proficiency cut score<sup>1</sup> estimates from academic year 2005–2006, but applied them against the Georgia AYP rules for academic year 2007–2008 (shortened to “2008” in this report).

Here are some key findings:

- We estimate that **11 of 18 elementary schools** and **15 of 18 middle schools** in our sample fail to make adequate yearly progress in 2008 under Georgia’s accountability system. (This rate is partly explained by our sample, which intentionally includes some schools with a relatively large population of low-performing students.)

<sup>1</sup> A cut score is the minimum score a student must receive on NWEA’s Measures of Academic Progress (MAP) that is equivalent to performing proficient on the Georgia Criterion-Referenced Competency Tests (CRCT).

<sup>2</sup> It’s important to note that students in subgroups not meeting the minimum *n* sizes are still included for accountability purposes in the overall student calculations; they simply are not treated as their own subgroup.

- Looking across the 28 state accountability systems examined in the study, we find that **eight states exceed Georgia in terms of the number of elementary schools making AYP** (see Figure 1).
- Nearly all of the schools in our sample that fail to make AYP in Georgia are meeting expected targets for their overall populations but failing because of the performance of individual subgroups.<sup>2</sup>
- **Several sample schools made AYP in Georgia that failed to make AYP in most other states. This is likely due to the fact that Georgia’s proficiency standards are relatively easy, compared to other states; these schools also had fewer accountable subgroups.**
- As in other states, schools with fewer subgroups attain AYP more easily in Georgia than schools with more subgroups, even when their average student performance is much lower. In other words, schools with greater diversity and size face greater challenges in making AYP.
- As in other states, Georgia’s middle schools have greater difficulty reaching AYP than do elementary schools, primarily because their student populations

**Georgia’s** AYP rules create a set of circumstances which mean that several schools make AYP in Georgia that do not in most of the 27 other states. This is likely due to the fact that Georgia’s proficiency standards (or cut scores) are relatively easy compared to other states (most are below the 25th percentile). On the other hand, Georgia’s annual 2008 targets for reading are relatively difficult to achieve. In fact, roughly 73 percent of a given population in any school must obtain reading proficiency in order for the school to make AYP. Consequently, every single school with a limited English proficient (LEP) subgroup failed to make AYP.

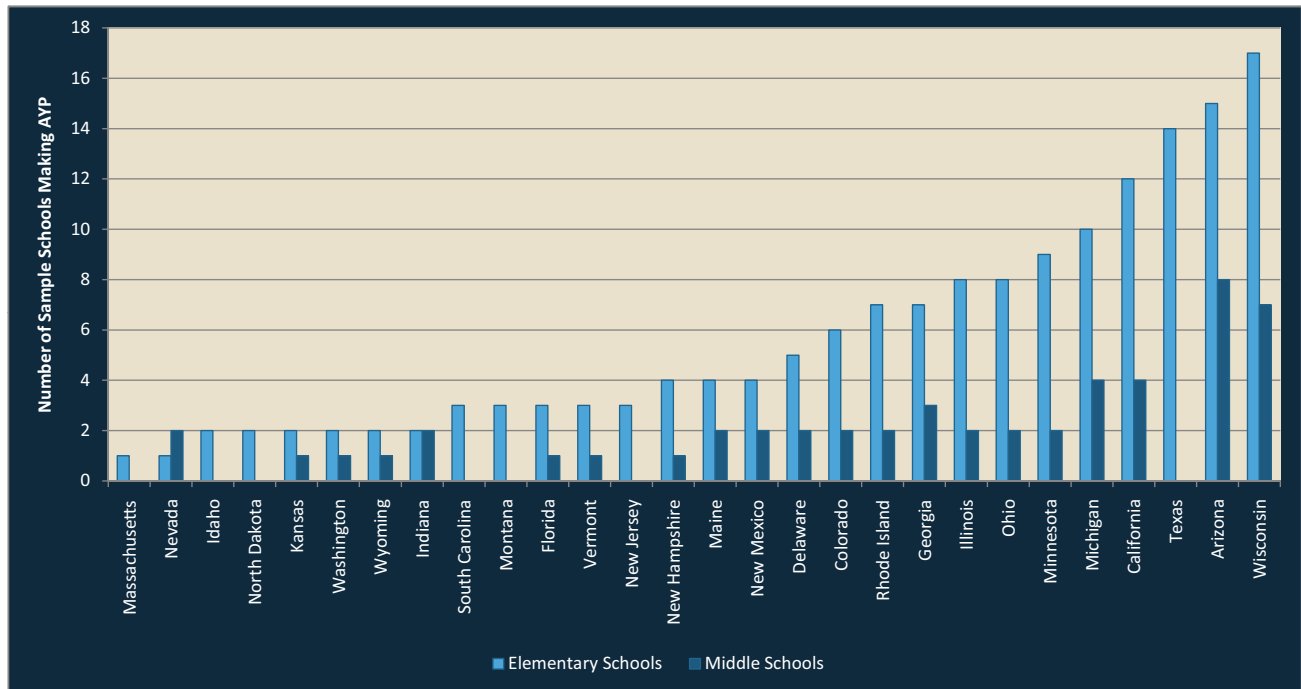


Figure 1. Number of sample schools making AYP by state

Note: Middle schools were not included for Texas and New Jersey; absence of a middle school bar in those states means “not applicable” as opposed to zero. States like Idaho and North Dakota, however, have zero passing middle schools.

are larger and therefore have more qualifying subgroups—not because their student achievement is lower than in the elementary schools.

- A strong predictor of whether or not a school will make AYP under Georgia’s system is whether it has enough students with disabilities (SWD) or English language learners to qualify as a separate subgroup. **Even though Georgia’s proficiency standards (or cut scores) are relatively easy compared to other states, its annual targets (for reading, especially) are relatively difficult to achieve. Consequently, every single school with a limited English proficient (LEP)<sup>3</sup> subgroup failed to make AYP, and almost all schools with enough qualifying SWD subgroups failed to meet their AYP targets.<sup>4</sup>**

## Introduction

*The Proficiency Illusion* (Cronin et al. 2007a) linked student performance on Georgia’s tests and those of 25 other states to the Northwest Evaluation Association’s Measures of Academic Progress (MAP), a computerized adaptive test used in schools nationwide. This single common scale permitted cross-state comparisons of each state’s reading and math proficiency standards to measure school performance under the No Child Left Behind Act (NCLB). That study revealed profound differences in states’ proficiency standards (i.e., how difficult it is to achieve proficiency on the state test), and even across grades within a single state.

Our study expands on *The Proficiency Illusion* by examining other key factors of state NCLB accountability

<sup>3</sup> Note that we use “LEP students” and “English language learners” interchangeably to refer to students in the same subgroup.

<sup>4</sup> SWDs are defined as those students following individualized education plans. We should also note that our subgroup findings for LEP students and SWDs may be more negative than actual findings, mostly because of the likely differences between how LEP students and SWDs are treated in MAP, the assessment we used in this study, and in the Georgia Criterion-Referenced Competency Test, the standardized state test. Specifically, the U.S. Department of Education has issued new NCLB guidelines in recent years that exclude small percentages of LEP students and SWDs from taking the state test or that allow them to take alternative assessments. In this study, however, no valid MAP scores were omitted from consideration.

plans and how they interact with state proficiency standards to determine whether the schools in our sample made adequate yearly progress (AYP) in 2008. Specifically, we estimated how a single set of schools, drawn from around the country, would fare under the differing rules for determining AYP in 28 states (the original 25 in *The Proficiency Illusion* plus 3 others for which we now have cut score estimates). In other words, if we could somehow move these entire schools—with their same mix of characteristics—from state to state, how would they fare in terms of making AYP? Will schools with high-performing students consistently make AYP? Will schools with low-performing students consistently fail to make AYP? If AYP determinations for schools are not consistent across states, what leads to the inconsistencies?

NCLB requires every state, as a condition of receiving Title I funding, to implement an accountability system that aims to get 100% of its students to the proficient level on the state test by academic year 2013–2014. In the intervening years, states set annual measurable objectives (AMOs). This is the percentage of students in each school, and in each subgroup within the school (such as low-income<sup>5</sup> or African American, among others), that must reach the proficient level in order for the school to make AYP in a given year. These AMOs vary by state (as do, of course, the difficulty of the proficiency standards).

States also determine the minimum number of students that must constitute a subgroup in order for its scores to be analyzed separately (also called the minimum *n* [number of students in sample] size). The rationale is that reporting the results of very small subgroups—fewer than ten pupils, for example—could jeopardize students’ confidentiality and risk presenting inaccurate results. (With such small groups, random events, like one student being out sick on test day, could skew the outcome.) Because of this flexibility, states have set widely varying *n* sizes for their subgroups, from as few as 10 youngsters to as many as 100.

Many states have also adopted confidence intervals—basically margins of statistical error—to account for po-

tential measurement error within the state test. In some states, these margins are quite wide, which has the effect of making it easier to achieve an annual target. **In Georgia, however, confidence intervals are only applied to schools that have fewer than 40 students. There were no schools that small in our sample, so confidence intervals were not considered when evaluating the performance of these schools under Georgia AYP rules.**

All of these AYP rules vary by state. This means that a school making AYP in Wisconsin or Ohio, for example, might not make it under South Carolina’s or Idaho’s rules (U.S. Department of Education 2008).

## What We Studied

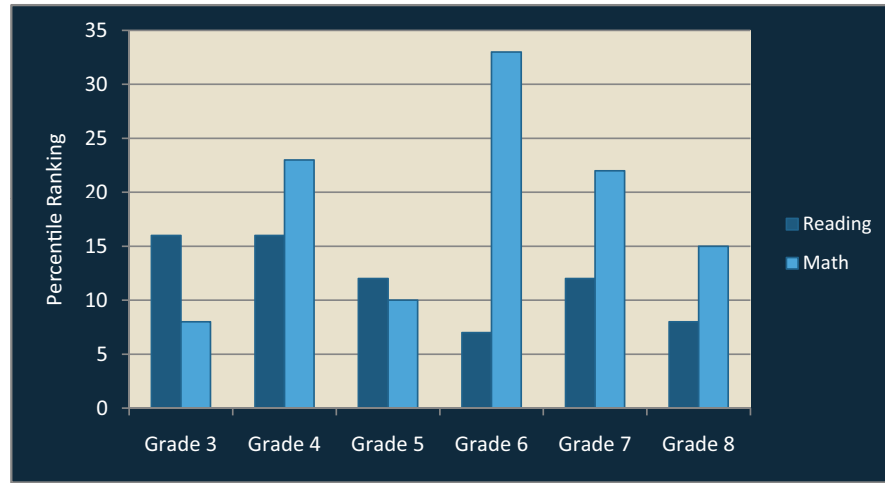
We collected students’ MAP test scores from the 2005–2006 academic year from 18 elementary and 18 middle schools around the country. We also collected the NCLB subgroup designations for all students in those schools—in other words, whether they had been classified as members of a minority group or as English language learners, among other subgroups.

The schools were not selected as a representative sample of the nation’s population. Instead, we selected the schools because they exhibited a range of characteristics on measures such as academic performance, academic growth, and socioeconomic status (the latter calculated by the percentage of students receiving free or reduced-price lunches). Appendix 1 contains a complete discussion of the methodology for this project along with the characteristics of the school sample.<sup>6</sup>

Proficiency cut score estimates for the Georgia Criterion-Referenced Competency Tests (CRCT) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Georgia’s definitions of proficiency ranked below the standards set by the other 25 states examined in that study. These cut scores were used to estimate whether students would have scored as proficient or better on the Georgia test, given their performance

<sup>5</sup> Low-income students are those who receive a free or reduced-price lunch.

<sup>6</sup> We gave all schools in our sample pseudonyms in this report.



**Figure 2.** Georgia reading and math cut score estimates, expressed as percentile ranks (2006)

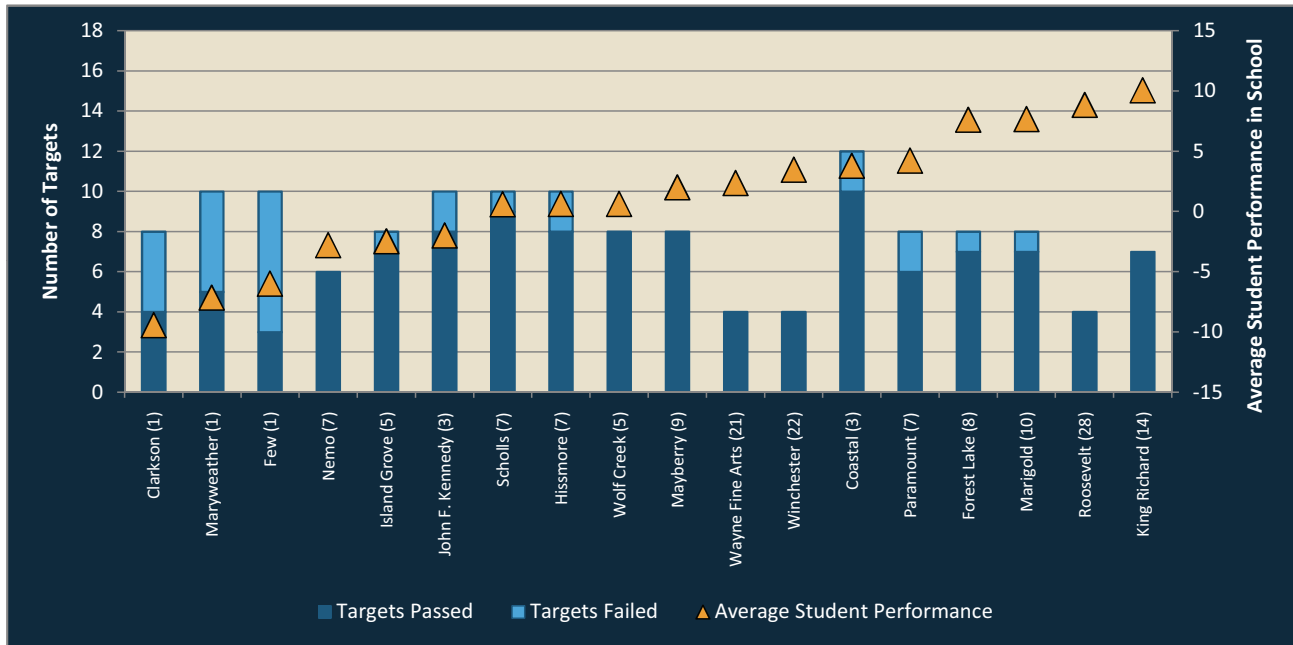
Note: This figure illustrates the difficulty of Georgia's cut scores ("proficiency passing scores") for its reading and mathematics tests, as percentiles of the NWEA norm, in grades three through eight. Higher percentile ranks are more difficult to achieve. All of Georgia's cut scores are below the 35th percentile.

**Table 1.** Georgia AYP rules for 2008

Subgroup minimum <i>n</i>	Race/ethnicity: 40 or 15% of school population, up to 75 students	
	SWDs: 40 or 15% of school population, up to 75 students	
	Low-income students: 40 or 15% of school population, up to 75 students	
	LEP students: 40 or 15% of school population, up to 75 students	
CI	Applied to proficiency rate calculations?	
	Yes; 95% CI	Used only when school population is less than 40
AMOs	Baseline proficiency levels as of 2002 (%)	2008 targets (%)
<b>READING/LANGUAGE ARTS</b>		
Grade 3	60.0	73.3
Grade 4	60.0	73.3
Grade 5	60.0	73.3
Grade 6	60.0	73.3
Grade 7	60.0	73.3
Grade 8	60.0	73.3
<b>MATH</b>		
Grade 3	50.0	66.7
Grade 4	50.0	66.7
Grade 5	50.0	66.7
Grade 6	50.0	66.7
Grade 7	50.0	66.7
Grade 8	50.0	66.7

Sources: U.S. Department of Education (2008); Council of Chief State School Officers (2008).

Abbreviations: SWDs = students with disabilities; LEP = limited English proficiency; CI = confidence interval; AMOs = annual measurable objectives



**Figure 3.** AYP performance of the elementary school sample under Georgia's 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under the Georgia AYP rules (as described in Table 1). The bars show the number of targets that each school had to meet in order to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not meet them (light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMO for even a single subgroup didn't make AYP, so any light blue means the school failed. Paramount Elementary, for example, met six of its eight targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). This is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

on MAP. Student test data and subgroup designations were then used to determine how these 18 elementary and 18 middle schools would have fared under Georgia AYP rules for 2008. In other words, the school data are from 2005–06 but we are applying them against Georgia's 2008 AYP rules.

Table 1 shows the pertinent Georgia AYP rules that were applied to elementary and middle schools in the current study. Georgia's minimum subgroup size is 40; if 40 does not constitute 15% of the total student population, then the minimum *n* is 15% of the total student population, up to 75 students. The sliding minimum subgroup number used by Georgia is not used by most other states, but it means that for many schools, the actual minimum number will be larger than 40.<sup>7</sup> Most states examined

also apply confidence intervals (or margins of statistical error) to student proficiency rates. The 95% confidence interval applied by Georgia is the most commonly found confidence interval in the study. So, while schools are supposed to get 73.3% of their students to the proficient level on the state reading test, and 73.3% of their students in each subgroup, applying the confidence interval means that the real target can actually be lower, particularly with smaller groups.

**Note that we were unable to examine the impact of NCLB's "safe harbor" provision.** This provision permits a school to make AYP even if some of its subgroups fail, as long as it reduces the number of nonproficient students within any failing subgroup by at least 10% relative to the previous year's performance. Because we had

<sup>7</sup> So then, the minimum subgroup size in Georgia cannot be less than 40 or more than 75 students. For example, a school with a total population of 1000 would have a minimum subgroup size of 75 since 40 does not constitute 15% and 15% of 1000 (i.e., 150) exceeds the 75-student ceiling.

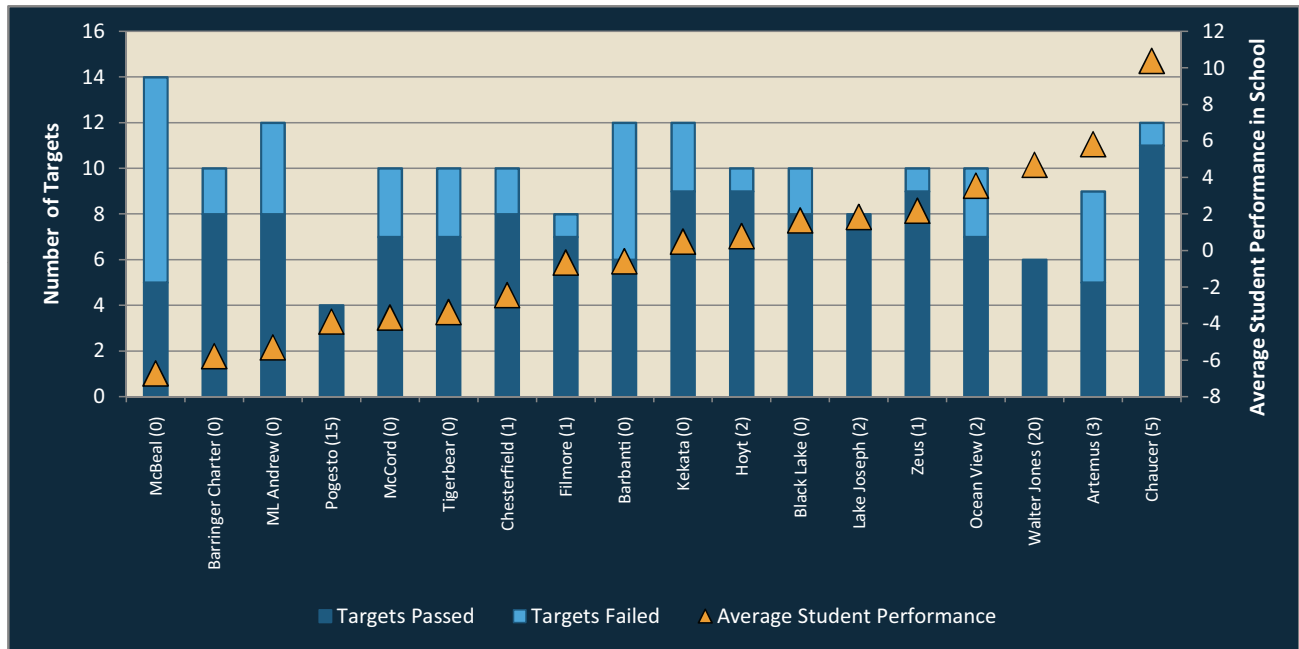


Figure 4. AYP performance of the middle school sample under Georgia's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under the AYP rules in Georgia (as described in Table 1). The bars show the number of targets that each school had to meet in order to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not meet them (light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMO for even a single subgroup didn't make AYP, so any light blue means the school failed. Artemus Middle School, for example, met five of its nine targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). This is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

access to only a single academic year's data (2005–2006), we were not able to include this in our analysis. As a result, it's possible that some of the schools in our sample that failed to make AYP according to our estimates would have made AYP under real conditions.

Furthermore, attendance and test participation rates are beyond the scope of the study. Most states include attendance rates as an additional indicator in their NCLB accountability system for elementary and middle schools. Plus, federal law requires 95% of each school's students—and 95% of the students in each subgroup—to participate in testing.

To reiterate, then, AYP decisions in the current study are modeled solely on test performance data for a single academic year. For each school, we calculated reading and math proficiency rates (along with any confidence intervals) to determine whether the overall school population and any qualifying subgroups achieved the AMOs. We deemed that a school made AYP if its overall student

body and all its qualifying subgroups met or exceeded its AMOs. Again, Appendix 1 supplies further methodological detail.

### How Did the Sample Schools Fare Under Georgia's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Georgia's 2008 AYP rules. **Only seven elementary schools made AYP while eleven failed to make it.** The triangles in Figure 3 show the average academic performance of students within the school, with negative values indicating below-grade-level performance for the average student, and positive values indicating above-grade-level performance. One might expect that schools with higher average student performance might have better AYP outcomes than schools with lower average student performance, but as the triangles in Figure 3 show, this is not universally true. Many of the schools on the right side of the figure (higher per-



Table 2. Elementary subgroup performance of sample schools under the 2008 Georgia AYP rules

SCHOOL PSEUDONYM	Overall Proficiency Rate		Overall		SWDs		LEP Students		Low-income Students		AA		Asian		Hispanic		AI/AN		White		AYP Targets Required		Targets MET	% of Targets Met	School Met AYP?	Number of states in which school met AYP?
	Math	Reading	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	AYP	Targets				
Clarkson	77.8%	65.6%	Y	N			Y	N	Y	N					Y	N					8	4	50%	N	1	
Maryweather	78.1%	70.3%	Y	N			N	N	Y	N					Y	N			Y	Y	10	5	50%	N	1	
Few	83.2%	72.5%	Y	N	N	N	N	N	Y	N					Y	N					10	3	30%	N	1	
Nemo	85.6%	85.1%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	7	
Island Grove	86.6%	84.0%	Y	Y					Y	Y					Y	N			Y	Y	8	7	88%	N	4	
JFK	91.8%	81.3%	Y	Y	Y	N			Y	Y	Y	N							Y	Y	10	8	80%	N	3	
Scholls	93.3%	85.5%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	7	
Hissmore	90.9%	88.5%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7	
Wolf Creek	88.1%	84.5%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	5	
Alice Mayberry	94.1%	90.7%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	9	
Wayne Fine Arts	96.0%	95.4%	Y	Y															Y	Y	4	4	100%	Y	21	
Winchester	91.5%	91.0%	Y	Y															Y	Y	4	4	100%	Y	22	
Coastal	90.2%	87.8%	Y	Y	Y	N			Y	Y	Y	Y			Y	N			Y	Y	12	10	83%	N	3	
Paramount	89.6%	87.3%	Y	Y					Y	N					Y	N			Y	Y	8	6	75%	N	7	
Forest Lake	96.2%	94.6%	Y	Y	Y	N			Y	Y									Y	Y	8	7	88%	N	8	
Marigold	95.7%	93.9%	Y	Y	Y	N			Y	Y									Y	Y	8	7	88%	N	10	
Roosevelt	99.0%	98.0%	Y	Y															Y	Y	4	4	100%	Y	28	
King Richard	96.6%	96.3%	Y	Y	Y	Y			Y										Y	Y	7	7	100%	Y	14	

Abbreviations: M = math; R = reading; N = no; Y = yes; SWDs = students with disabilities; AA = African American; Asian/Pacific Islander = Asian; Hispanic/Latino = Hispanic; American Indian/Alaska Native = AI/AN.

Note: Schools are ordered from lowest (Clarkson) to highest (King Richard) average student performance as measured by combined and weighted math and reading performance on the MAP assessment (not shown in table). A blank space underneath a subgroup means that subgroup contained fewer than the minimum number of students required for evaluation, so it wasn't counted. A "Y" in blue means that the group met the AMOs and an "N" in peach means that the group did not meet the AMOs. The two rightmost columns show (1) whether that school met AYP (i.e., it met the targets for its overall population and all required subgroups); and (2) the total number of states in the study for which that school met AYP.

forming) made AYP, while most of the lower performing schools did not. However, one low performing school with few subgroup targets made it (Nemo), while four higher performing schools (Coastal, Paramount, Forest Lake, and Marigold) with higher numbers of subgroup targets did not.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Georgia AYP rules. **Out of 18 in our sample, only 3 made AYP**—1 low-performance school and 2 high-performance schools, but all three have relatively few qualifying subgroups.

### Where do schools fail?

Figures 3 and 4 illustrate that schools with low or mid-dling performance can still make AYP when the school has fewer targets to meet, thanks to fewer subgroups. These figures do not, however, indicate which subgroups failed in which school. Information on individual subgroup performance appears in Tables 2 and 3 for elementary and middle schools, respectively.

Tables 2 and 3 show which subgroups qualified for evaluation at each school (i.e., whether the number of stu-

**Table 3.** Middle school subgroup performance of sample schools under the 2008 Georgia AYP rules

SCHOOL PSEUDONYM	Overall Proficiency Rate		Overall		SWDs		LEP Students		Low-income Students		AA		Asian		Hispanic		AI/AN		White		AYP Targets Required	Targets MET	% of Targets Met	School Met AYP?	Number of states in which school met AYP?
	Math	Reading	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R	M	R					
McBeal	67.6%	81.0%	Y	Y	N	N	N	N	N	N	N	Y			N	N			Y	Y	14	5	36%	N	0
Barringer Charter	76.0%	84.8%	Y	Y	N	N			Y	Y	Y	Y			Y	Y					10	8	80%	N	0
ML Andrew	70.4%	90.0%	Y	Y	N	N			N	Y	N	Y			Y	Y			Y	Y	12	8	67%	N	0
Pogesto	77.8%	98.1%	Y	Y															Y	Y	4	4	100%	Y	15
McCord Charter	69.5%	91.9%	Y	Y					N	Y	N	Y			N	Y			Y	Y	10	7	70%	N	0
Tigerbear	77.8%	88.0%	Y	Y	N	N			Y	Y	N	Y							Y	Y	10	7	70%	N	0
Chesterfield	80.3%	91.9%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	1
Filmore	77.7%	93.8%	Y	Y					Y	Y					N	Y			Y	Y	8	7	88%	N	1
Barbanti	73.6%	89.5%	Y	Y	N	N	N	N	N	Y					N	Y			Y	Y	12	6	50%	N	0
Kekata	84.9%	91.6%	Y	Y	N	Y			Y	Y	Y	Y			N	N			Y	Y	12	9	75%	N	0
Hoyt	83.7%	92.6%	Y	Y	N	Y			Y	Y	Y	Y							Y	Y	10	9	90%	N	2
Black Lake	87.2%	92.1%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	0
Lake Joseph	84.6%	93.8%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	2
Zeus	87.6%	93.3%	Y	Y	N	Y			Y	Y	Y	Y							Y	Y	10	9	90%	N	1
Ocean View	87.7%	96.5%	Y	Y			N	Y	N	Y					N	Y			Y	Y	10	7	70%	N	2
Walter Jones	90.1%	93.1%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	20
Artemus	87.1%	94.2%	Y	Y	N				N	Y					N	N			Y	Y	9	5	56%	N	3
Chaucer	92.8%	98.5%	Y	Y	N	Y			Y	Y			Y	Y	Y	Y			Y	Y	12	11	92%	N	5

Abbreviations: M = math; R = reading; N = no; Y = yes; SWDs = students with disabilities; AA = African American; Asian/Pacific Islander = Asian; Hispanic/Latino = Hispanic; American Indian/Alaska Native = AI/AN.

Note: Schools are ordered from lowest (McBeal) to highest (Chaucer) average student performance as measured by combined and weighted math and reading performance on the MAP assessment (not shown in table). A blank space underneath a subgroup means that subgroup contained fewer than the minimum number of students required for evaluation, so it wasn't counted. A "Y" in blue means that the group met the AMOs and an "N" in peach means that the group did not meet the AMOs. The two rightmost columns show (1) whether that school met AYP (i.e., it met the targets for its overall population and all required subgroups); and (2) the total number of states in the study for which that school met AYP.

dents within that subgroup exceeded the state's minimum *n*), and whether that subgroup passed or failed. Although all schools are evaluated on the proficiency rate of their overall population, potential subgroups that are separately evaluated for AYP include SWDs, students with LEP, low-income students, and the following race/ethnic categories: African American, Asian/Pacific Islander, Hispanic/Latino, American Indian/Alaska Native, and white. Tables 2 and 3 also show whether a school met AYP under the 2008 Georgia rules, and the total number of states within the study in which that school met AYP.

The school-by-school findings in Tables 2 and 3 show that:

- Every elementary and middle school in the sample met overall math targets.
- Seven out of eight elementary schools and all twelve middle schools with qualifying SWD subgroups failed to meet their targets for this group of students.
- All three elementary schools (Clarkson, Maryweather, and Few) and all three middle schools (McBeal, Barbanti, and Ocean View) with qualifying LEP subgroups failed to meet their targets for this group of students.



**Table 4.** Summary of subgroup performance of sample elementary schools under the 2008 Georgia AYP rules

SUBGROUP	Number of schools with qualifying subgroups	Number of schools where subgroup failed to meet math target	Number of schools where subgroup failed to meet reading target
Students with disabilities	8	2	7
Students with limited English proficiency	3	2	3
Low-income students	15	0	4
African-American students	5	0	1
Asian/Pacific Islander students	0	0	0
Hispanic students	7	0	6
American Indian/Alaska Native students	0	0	0
White students	16	0	0

**Table 5.** Summary of subgroup performance of sample middle schools under the 2008 Georgia AYP rules

SUBGROUP	Number of schools with qualifying subgroups	Number of schools where subgroup failed to meet math target	Number of schools where subgroup failed to meet reading target
Students with disabilities	12	12	7
Students with limited English proficiency	3	3	2
Low-income students	17	6	1
African-American students	10	4	0
Asian/Pacific Islander students	1	0	0
Hispanic students	11	7	3
American Indian/Alaska Native students	0	0	0
White students	17	0	0

- Four elementary schools and six middle schools failed to meet their targets for their low-income subgroups.
- Seven elementary schools and nine middle schools failed to make AYP due to one or more racial/ethnic subgroups.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. The perform-

ance of SWDs and LEP students are proving most challenging for schools under Georgia's system. In fact, all but one elementary and more than half of the middle schools in the study with qualifying SWD subgroups failed to make AYP in reading. Every middle school in the sample with a minimum *n* of 40 in the SWD subgroup failed to make AYP in math. All but one school with a large enough LEP population to qualify as a separate subgroup failed to meet its reading targets for these students.

Table 6. Comparisons between schools that did and didn't make AYP in Georgia, 2008

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	7	11	3	15
Average student body size	286	316	350	961
Average % low income	26	59	39	46
Average % nonwhite	29	48	51	42
Average performance <sup>†</sup>	3.51	-0.23	0.88	-0.24
Average % growth <sup>‡</sup>	113	116	110	96
Average number of targets to meet	6	9	6	11

<sup>†</sup> Student performance is measured by NWEA's MAP assessment and is expressed as an index of grade level normative performance. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance.

<sup>‡</sup> Average growth refers to improvement from fall to spring on the NWEA MAP assessments, averaged across all students within the school. Growth is expressed as an index value relative to NWEA norms and is scaled as a percentage. Thus, 100% means that students at the school are achieving normative levels of growth for their age and grade. Less than 100% growth means that the average student is increasing *by less* than normative amounts, while percentages over 100 mean that the average student is *exceeding* normative growth expectations.

## Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Georgia's NCLB accountability system is, in some respects, behaving similarly as those in other states. For example, among the elementary schools in our sample, Roosevelt, Winchester, and Wayne Fine Arts all make AYP in the greatest number of states—28, 22, and 21, respectively. And these schools all made AYP in Georgia, too. Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make AYP in Georgia.

But Georgia is also home to a few anomalies. First, consider Mayberry Elementary (see Figure 3). It failed to make AYP in 19 of the 28 states in our sample, yet made AYP in Georgia. In examining Table 2, one can see that Mayberry does not meet the minimum numbers for the LEP or SWD subgroups, which creates difficulty for so many other schools within the sample. Similarly, Wolf Creek and Nemo Elementary Schools also had no SWD or LEP subgroups. With fewer accountable subgroups, and with relatively easy proficiency standards (Figure 1),

these schools are able to meet AYP, even when other schools with higher average performance fail.

Second, look at Pogesto Middle School (Figure 4). Even with its relatively low average performance it made AYP in Georgia, but failed to do so in 13 of 28 states. Like Mayberry, its AYP success in Georgia is likely attributable to the relatively small number of targets (four) it has to meet (as shown in Table 3), along with the relatively easy proficiency standards in Georgia, compared to other states.

This is consistent with the patterns shown in Table 6, which compares the sample schools that did and didn't make AYP on a number of academic and demographic dimensions. Within the sample, elementary schools that made AYP do indeed show higher average student performance, but they also differ in the following ways: they have much smaller student populations, and have fewer subgroups (meaning fewer targets to meet).

## Concluding Observations

This study examined the test performance data of students from 18 elementary and 18 middle schools across the

country to see how these schools would have fared under Georgia's AYP rules and (and AMOs) for 2008. We found that only 7 elementary schools and 3 middle schools—10 in all, from a sample of 36—would have made AYP in Georgia. Looking across the 28 state accountability systems examined in the study, this puts Georgia in the upper middle of the distribution in terms of the number of elementary schools making AYP (see Figure 1).

Because the overriding goal of NCLB is to eliminate educational disparities within and across states, it's important to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, the NCLB accountability system in Georgia is working exactly as Congress intended: identifying as needing attention those schools with relatively high test score averages that mask low performance for particular groups of students, such as low-income or Hispanic students. Almost all the sample schools met the Georgia AMO targets for their student populations as a whole, i.e., not considering subgroup results. In the pre-NCLB era, such schools might have been considered effective or at

least not in need of improvement, even though sizable numbers of their pupils were not meeting state standards. Disaggregating data by race, income, and so on has made those students visible. That is surely a positive step.

Yet NCLB's design flaws are also readily apparent. Does it make sense that the size of a school's enrollment has so much influence over making AYP? Does it make sense that having fewer subgroups enhances the likelihood of making AYP? Even if actual participation guidelines for English language learners and SWDs are more generous under the current state assessment system,<sup>8</sup> doesn't the massive failure of middle school students to meet Georgia's targets indicate that a new approach is needed for holding schools accountable for the performance of these students? Yes, schools should redouble their efforts to boost achievement for ELL students and students with disabilities, as for other students, but when so few schools are able to meet the goal, perhaps that indicates that the goal is unrealistic. These will be critical considerations for Congress as it takes up NCLB reauthorization in the future.

## Limitations

Although the purpose of our study was to explore how various elements of accountability systems in different states jointly affect a school's AYP status, the study will not precisely replicate the AYP outcome for every single school for several reasons. Because we projected students' state test performance from their MAP scores, and because MAP assessments—unlike state tests—are not required of all students within a school, it's possible that sampling or measurement error (or both) affected school AYP outcomes within our model. Nevertheless, for all but two of the sampled schools, our projections matched NCLB-reported proficiency ratings (in each respective state) to within 5 percentage points.

An additional limitation of the study was that it was not possible to consider NCLB's safe harbor provisions, which might have allowed some schools to make AYP even though they failed to meet their state's required AMOs. A few schools would have also passed under the new growth-model pilots currently under way in a handful of states, such as Ohio and Arizona. Others identified as making AYP in our study might actually have failed to make it because they did not meet their state's average daily attendance requirement or because they did not test 95% of some subgroup within their overall student population. At the end of the day, then, it's important to keep in mind that the number of schools that did or did not make AYP in our study do

<sup>8</sup> See footnote 4.

not by themselves measure the effectiveness of the entire state accountability system, of which there are many parts.

Despite these limitations, we believe that the study illuminates the inconsistency of proficiency standards and some of the rules across states. It's also useful for illustrating the challenges that states face as the requirements for AYP continue to ratchet up. The national report contains additional discussion of the study methodology and its limitations.