



## 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 Florida’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 Florida’s 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 or not each would make AYP under Florida’s system as well as in systems in 27 other states. We used school data estimates from academic year 2005–2006, but applied them against Florida’s AYP rules and cut scores<sup>1</sup> for academic year 2007–2008 (shortened to “2008” in this report).

Here are some key findings:

- We estimate that **15 of 18 elementary schools** and **17 of 18 middle schools** in our sample failed to **make AYP** in 2008 under Florida’s accountability system. (The high failure rate is partly explained by

our sample, which intentionally includes some schools with a relatively large population of low-performing students.)

- Looking across the 28 state accountability systems examined in the study, only 8 states passed fewer of the sample elementary schools than Florida, while 4 states tied with Florida. **In addition, Florida was one of 6 states with a single middle school that made AYP in the sample** (see Figure 1).<sup>2</sup>
- Many of the schools in our sample that failed to make AYP in Florida met expected targets for their overall populations but didn’t make AYP because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.<sup>3</sup>
- **Two sample schools that failed to make AYP in most other states made AYP in Florida. This is**

Only four schools in the study make AYP in **Florida**. This can be attributed to a couple of factors. First, Florida’s cut scores range from the 30th to the 53rd percentile; hence, proficiency standards are relatively hard to achieve. Florida also does not apply a confidence interval (margin of error) to proficiency rate calculations (percentage of students achieving proficient or higher on the state test). This means that in Florida, schools will have greater difficulty achieving their annual targets than they would in states that employ confidence intervals. On the other hand, a couple of schools in the study make AYP in Florida but don’t in most other states. This is likely because the minimum subgroup size in Florida tends to be large, meaning Florida schools will be accountable to fewer groups than schools in many other states.

<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 Florida Comprehensive Assessment Test.

<sup>2</sup> Note that Florida received full approval from the U.S. Department of Education to implement a student growth model for the 2006–2007 school year. The current analysis, which draws on data from 2005–2006, does not in any way use or incorporate student growth model calculations.

<sup>3</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.

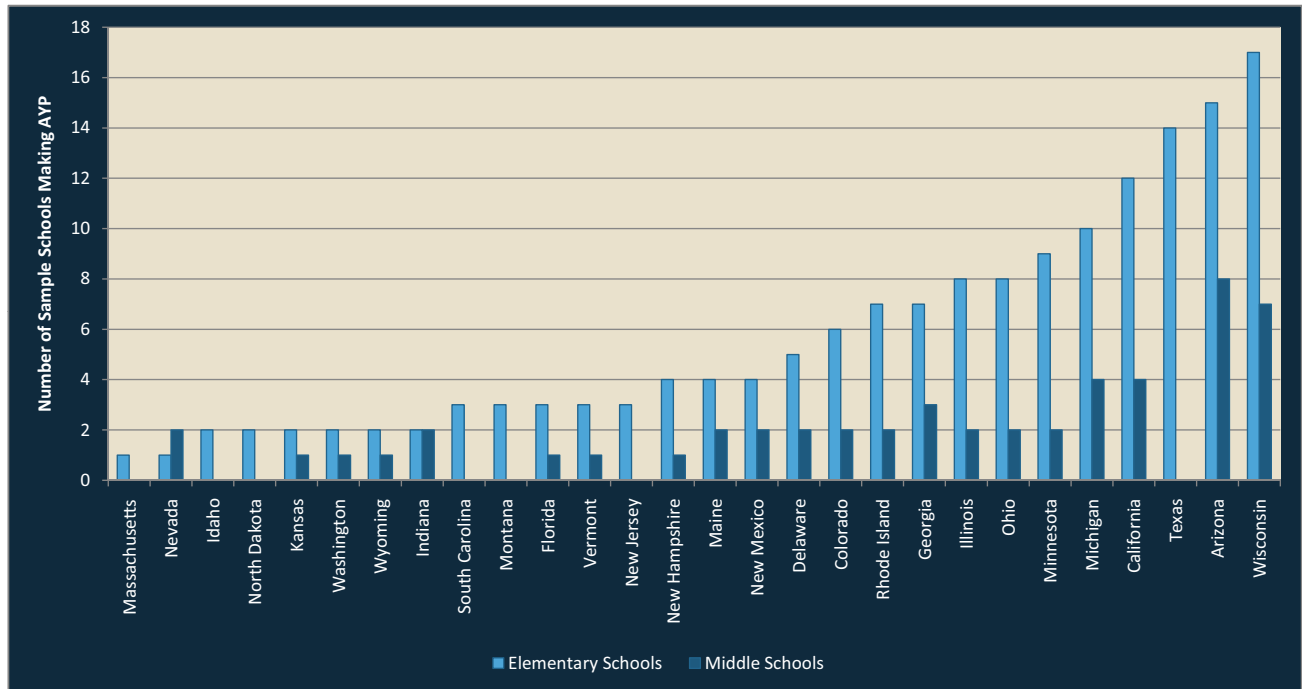


Figure 1. Number of sample schools making AYP by state

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.

probably because these two schools had fewer accountable subgroups under Florida’s AYP rules.

- Schools with fewer subgroups attained AYP more easily in Florida than schools with more subgroups, even when their average student performance was much lower. In other words, schools with greater diversity and size face greater challenges in making AYP. This is true other states as well.
- Middle schools had greater difficulty reaching AYP in Florida than did elementary schools, primarily because some of the middle school proficiency standards are more difficult than at the elementary grades, and because the student populations are larger and therefore the schools have more qualifying subgroups—not because their student achievement was lower than in the elementary schools.

- A strong predictor of whether or not a school would make AYP under Florida’s system is whether it has enough English language learners or SWDs to qualify as separate subgroups. Every school with a limited English proficient (LEP)<sup>4</sup> or SWD subgroup failed to make AYP.<sup>5</sup>

### Introduction

*The Proficiency Illusion* (Cronin et al. 2007a) linked student performance on various standardized tests in 25 states to the Northwest Evaluation Association’s (NWEA’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 (NCLB) Act of 2001. That study revealed

<sup>4</sup> Note that we use “LEP students” and “English language learners” interchangeably to refer to students in the same subgroup.

<sup>5</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 the likely differences between how LEP students and SWDs are treated in MAP, the assessment we used in this study, and in the Florida Comprehensive Assessment Test, the standardized state test. Specifically, the U.S. Department of Education has issued NCLB guidelines in recent years that exclude small percentages of LEP students and SWDs from taking the state test or allow them to take alternative assessments. In this study, however, no valid MAP scores were omitted from consideration.

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* (for which Florida did not participate) by examining other key factors of state NCLB accountability 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 Florida, and 2 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>6</sup> or African American, among others), that must reach the proficient level in order for the school to make AYP in a given year. The 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' con-

fidentiality 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 potential 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.

All of these AYP rules vary by state, which means that a school that makes 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**

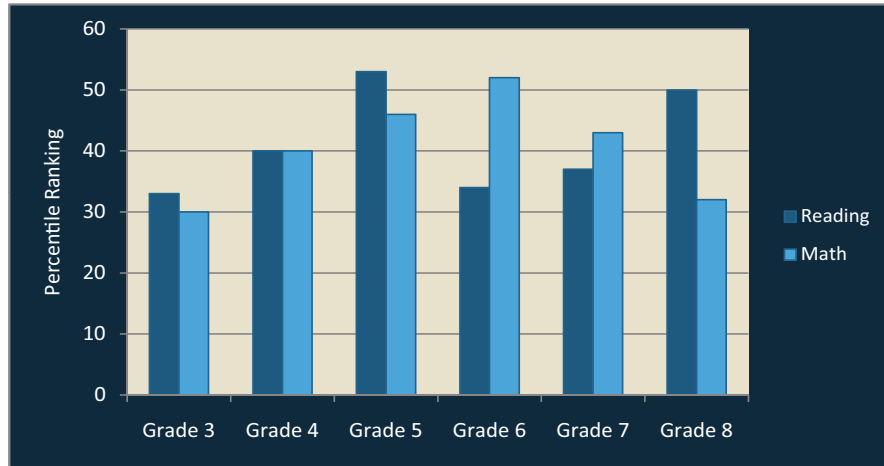
We collected students' MAP test scores from the 2005–2006 academic year from 18 elementary and 18 middle schools around the country. We also collected the NCLB subgroup designations for all students in those schools—in other words, whether they had been classified as members of a minority group or as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

These 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 last 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>7</sup>

Proficiency cut score estimates for the Florida Comprehensive Assessment Test (FCAT) are shown in Figure 2. These cut scores were used to estimate whether students would have scored as proficient or better on the Florida

<sup>6</sup> Low-income students are those who receive a free or reduced-price lunch.

<sup>7</sup> We gave all schools in our sample pseudonyms in this report.



**Figure 2.** Florida reading and math cut score estimates, expressed as percentile ranks (2007)

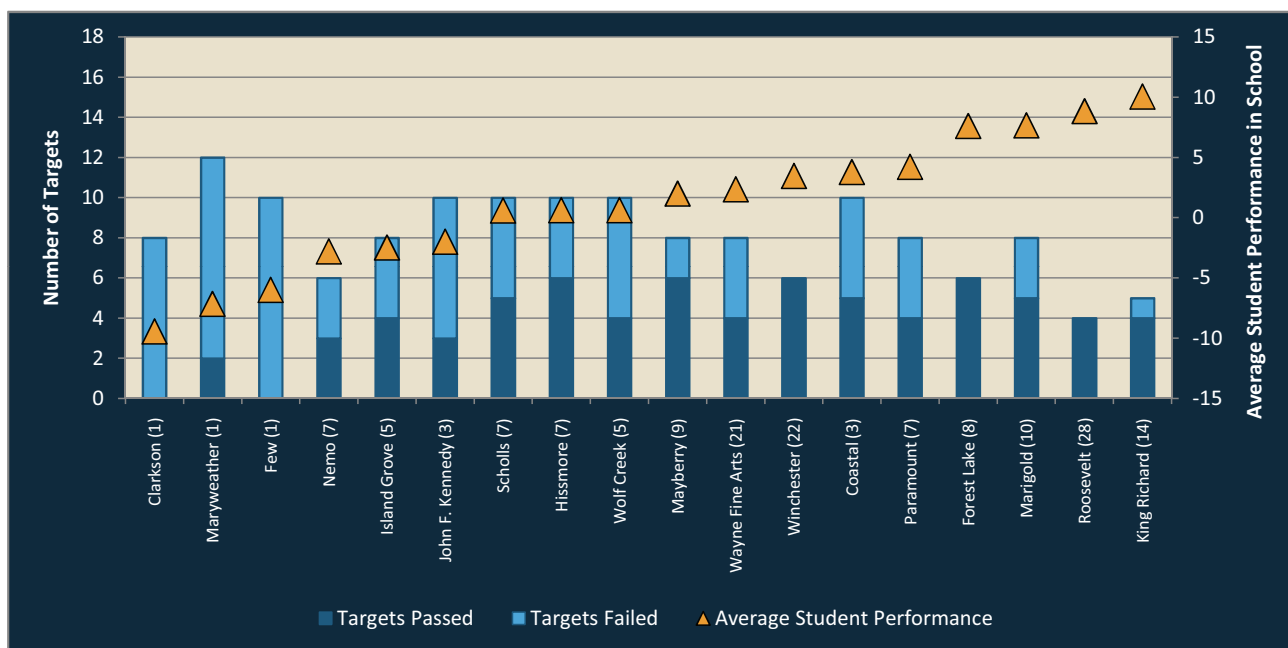
Note: This figure illustrates the difficulty of Florida's cut scores (proficiency passing scores) for its reading and math tests, as percentiles of the NWEA norm groups, in grades three through eight. Percentile ranks denote the percentage of the NWEA norm group that would perform at or below that standard. For example, 70% of the third graders in NWEA's norm group would have exceeded the performance necessary to achieve math proficiency on the FCAT. Higher percentile ranks are more difficult to achieve. Most of Florida's cut scores are near or below the 50th percentile.

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

Subgroup minimum <i>n</i>	Race/ethnicity: 30 or 15% of school population, up to 100 students	
	SWDs: 30 or 15% of school population, up to 100 students	
	Low-income students: 30 or 15% of school population, up to 100 students	
	LEP students: 30 or 15% of school population, up to 100 students	
CI	Applied to proficiency rate calculations?	
	Not used	
AMOs	Baseline proficiency levels as of 2002 (%)	2008 targets (%)
<b>READING/LANGUAGE ARTS</b>		
Grade 3	31	58
Grade 4	31	58
Grade 5	31	58
Grade 6	31	58
Grade 7	31	58
Grade 8	31	58
<b>MATH</b>		
Grade 3	38	62
Grade 4	38	62
Grade 5	38	62
Grade 6	38	62
Grade 7	38	62
Grade 8	38	62

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 Florida's 2008 AYP rules

Note: This figure indicates how each elementary school within the sample fared under Florida's AYP rules (as described in Table 1). The bars show the number of targets that each school has 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. King Richard, for example, met four of its five 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), which is measured by the average MAP performance of students within the school; this 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.

test, given their performance 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 Florida AYP rules for 2008. (In other words, the school data are from academic year 2005–2006, but we are applying them against Florida's 2007–2008 cut scores and AYP rules.)

Table 1 shows the pertinent Florida AYP rules that were applied to elementary and middle schools in this study. Florida's minimum subgroup size is 30; if 30 does not constitute 15% of the total student population, then the minimum *n* is 15% of the total student population, up to 100 students.<sup>8</sup> **This means that for many schools the actual subgroup size is much larger than 30, meaning that Florida schools will have fewer subgroups for**

**which its held accountable than do schools in many other states.<sup>9</sup>**

Unlike most other states examined in the current study, Florida does not apply confidence intervals (or margins of error) to its measurements of student proficiency rates. This means that in Florida, **schools will have greater difficulty achieving their annual measurable objectives than they would in states that employ confidence intervals.**

**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% rela-

<sup>8</sup> So then, the minimum subgroup size in Florida cannot be less than 30 or more than 100 students. For example, a school with a total population of 1000 would have a minimum subgroup size of 100 since 30 does not constitute 15% and 15% of 1000 (i.e., 150) exceeds the 100-student ceiling.

<sup>9</sup> Keep in mind, however, that school size and *n* size are related (e.g., small *n* sizes make sense for small schools).

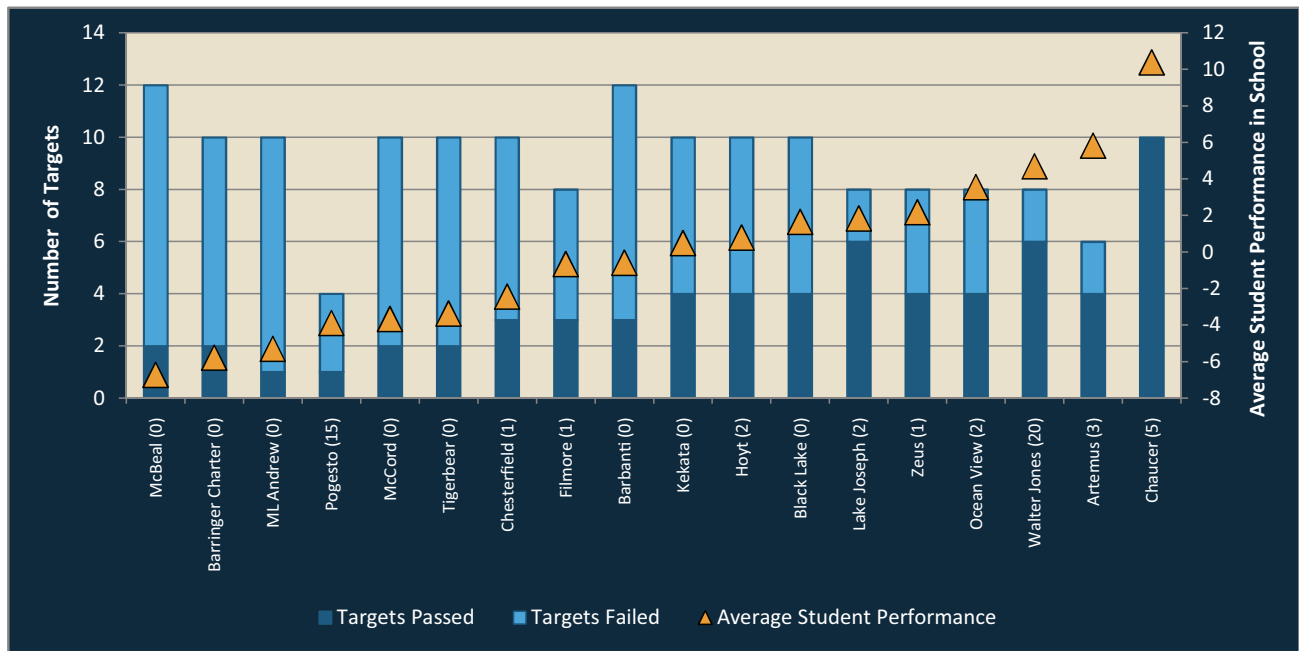


Figure 4. AYP performance of the middle school sample under Florida's 2008 AYP rules

Note: This figure shows how each middle school within the sample fared under Florida's 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 fails to meet the AMOs for even a single subgroup did not make AYP, so any light blue means that the school failed. Lake Joseph, 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), which 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.

tive to the previous year's performance. Because we had 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. Note that most states include attendance rates as an additional indicator in their NCLB accountability system for elementary and middle schools. In addition, 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 Florida's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Florida's 2008 AYP rules. **Only 3 elementary schools (Winchester, Forest Lake, and Roosevelt) out of 18 made AYP.** 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. All schools that made AYP are in the right half of the figure, meaning that the students with the highest average performance were found at these schools.

Yet almost without regard to average student perform-

Table 2. Elementary school subgroup performance of sample schools under the 2008 Florida 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	44.9%	27.9%	N	N			N	N	N	N					N	N					8	0	0%	N	1	
Maryweather	53.0%	41.6%	N	N	N	N	N	N	N	N					N	N			Y	Y	12	2	17%	N	1	
Few	58.4%	42.3%	N	N	N	N	N	N	N	N					N	N					10	0	0%	N	1	
Nemo	60.9%	58.1%	N	Y					N	N									Y	Y	6	3	50%	N	7	
Island Grove	64.3%	59.3%	Y	Y					N	N					N	N			Y	Y	8	4	50%	N	4	
JFK	67.7%	51.9%	Y	N	N	N			N	N	N	N							Y	Y	10	3	30%	N	3	
Scholls	75.0%	59.9%	Y	Y	N	N			Y	N	N	N							Y	Y	10	5	50%	N	7	
Hissmore	75.7%	61.8%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	7	
Wolf Creek	67.0%	61.2%	Y	Y	N	N			N	N					N	N			Y	Y	10	4	40%	N	5	
Alice Mayberry	73.1%	61.9%	Y	Y					Y	N	Y	N							Y	Y	8	6	75%	N	9	
Wayne Fine Arts	71.8%	71.3%	Y	Y					N	N	N	N							Y	Y	8	4	50%	N	21	
Winchester	74.5%	71.1%	Y	Y											Y	Y			Y	Y	6	6	100%	Y	22	
Coastal	76.9%	63.9%	Y	Y	N	N			Y	N	N	N							Y	Y	10	5	50%	N	3	
Paramount	78.1%	68.7%	Y	Y					N	N					N	N			Y	Y	8	4	50%	N	7	
Forest Lake	86.6%	78.8%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	8	
Marigold	88.5%	76.9%	Y	Y	N	N			Y	N									Y	Y	8	5	63%	N	10	
Roosevelt	90.9%	85.4%	Y	Y															Y	Y	4	4	100%	Y	28	
King Richard	86.5%	82.7%	Y	Y	N														Y	Y	5	4	80%	N	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.

ance, the only schools that made AYP were those with relatively few qualifying subgroups—and thus the fewest targets to meet (because each subgroup has its own separate targets). For example, Winchester and Forest Lake passed, but had only six targets each—two in reading and math for their overall populations, two in reading and math for their white population, and two in reading and math for an additional subgroup (Hispanic for Winchester, low income for Forest Lake).

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Florida AYP rules. Of 18

middle schools in our sample, only a single school passed—Chaucer—the school with the highest average student performance.

### Where Do Schools Fail?

Figures 3 and 4 illustrate how the elementary and middle schools, respectively, within the sample fared under the Florida rules, but do not identify which subgroups failed or passed in which school. Tables 2 and 3 list information on individual subgroup performance for elementary and middle schools, respectively.

Table 3. Middle school subgroup performance of sample schools under the 2008 Florida 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	49.3%	48.5%	N	N	N	N	N	N	N	N					N	N				Y	Y	12	2	17%	N	0
Barringer Charter	49.9%	48.4%	N	N	N	N			N	N	N	N			Y	Y						10	2	20%	N	0
ML Andrew	48.3%	51.4%	N	N					N	N	N	N			N	N				N	Y	10	1	10%	N	0
Pogesto	53.7%	53.7%	N	N																N	Y	4	1	25%	N	15
McCord Charter	48.9%	57.4%	N	N					N	N	N	N			N	N				Y	Y	10	2	20%	N	0
Tigerbear	58.4%	50.9%	N	N	N	N			N	N	N	N								Y	Y	10	2	20%	N	0
Chesterfield	63.7%	52.4%	Y	N	N	N			N	N	N	N								Y	Y	10	3	30%	N	1
Filmore	60.6%	61.7%	N	Y					N	N					N	N				Y	Y	8	3	38%	N	1
Barbanti	59.2%	58.3%	N	Y	N	N	N	N	N	N					N	N				Y	Y	12	3	25%	N	0
Kekata	67.7%	60.8%	Y	Y	N	N			N	N	N	N								Y	Y	10	4	40%	N	0
Hoyt	68.8%	64.5%	Y	Y	N	N			N	N	N	N								Y	Y	10	4	40%	N	2
Black Lake	73.1%	61.4%	Y	Y	N	N			N	N	N	N								Y	Y	10	4	40%	N	0
Lake Joseph	70.2%	66.9%	Y	Y					Y	Y					N	N				Y	Y	8	6	75%	N	2
Zeus	72.4%	67.4%	Y	Y	N	N			N	N										Y	Y	8	4	50%	N	1
Ocean View	72.9%	77.2%	Y	Y					N	N					N	N				Y	Y	8	4	50%	N	2
Walter Jones	74.4%	77.7%	Y	Y					N	Y					N	Y				Y	Y	8	6	75%	N	20
Artemus	76.1%	77.7%	Y	Y					N	N										Y	Y	6	4	67%	N	3
Chaucer	82.8%	83.3%	Y	Y					Y	Y				Y	Y	Y	Y			Y	Y	10	10	100%	Y	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.

Tables 2 and 3 show which subgroups qualified for evaluation at each school (i.e., whether the number of students 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 are 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 Florida 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:

- Three elementary schools (Clarkson, Maryweather, and Few) failed to meet overall population targets for both reading and math. One additional school (JFK) failed to meet the overall target in reading, and one other school (Nemo) failed to meet the overall target in mathematics.
- Six middle schools (McBeal, Barringer, ML Andrew,



**Table 4.** Summary of subgroup performance of sample elementary schools under the 2008 Florida 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	9	9	8
Students with limited English proficiency	3	3	3
Low-income students	15	9	14
African-American students	6	4	6
Asian/Pacific Islander students	0	0	0
Hispanic students	7	6	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 Florida 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	9	9	9
Students with limited English proficiency	2	2	2
Low-income students	17	15	14
African-American students	8	8	8
Asian/Pacific Islander students	1	0	0
Hispanic students	10	8	7
American Indian/Alaska Native students	0	0	0
White students	17	2	0

Pogesto, McCord, and Tigerbear) failed to meet overall targets in both reading and math. An additional school (Chesterfield) failed its overall target in reading, and two more (Filmore and Barbanti) failed overall targets in mathematics.

- One of the 15 elementary schools that didn't make AYP (King Richard) missed only for the SWD subgroup.

- One middle school (Artemus) failed to make AYP only because of its low-income subgroup.
- One middle school (Lake Joseph) passed in every subgroup except for Hispanic students.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. First, the performance of SWDs proved most challenging for schools

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	3	15	1	17
Average student body size	243	317	1083	846
Average % low income	20	52	10	47
Average % nonwhite	21	45	29	45
Average performance†	6.65	0.14	10.38	-0.67
Average % growth‡	131	112	175	94
Average number of targets to meet	5	9	10	9

† 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.

‡ 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.

under Florida's system. In fact, every elementary and middle school in the sample with qualifying SWD subgroups failed to meet its targets for that population. Students with LEP also struggled to meet the state's targets; every school with a large enough LEP population to qualify as a separate subgroup failed to meet its reading and math targets for these students. It is also clear that students belonging to traditionally academically disadvantaged subgroups (low income, Hispanic, and African American, among others) also struggled under the strict Florida AYP rules—many elementary and middle schools within the sample for which these subgroups were accountable failed to meet AYP.

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

A close look at Figures 3 and 4 indicates that Florida's NCLB accountability system is, in many respects, behaving like systems in other states. For example, among the elementary schools in our sample, Roosevelt and Winchester both made AYP in the greatest number of states—28 and 22, respectively. And these schools made

AYP in Florida, too. Likewise, most of the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make AYP in Florida.

But Florida is also home to a few anomalies. First, consider Wayne Fine Arts (see Figure 3). It made AYP in 21 of the 28 states in our sample, but failed to make AYP in Florida. In examining Table 2, we can see that Wayne Fine Arts failed for its low-income and African American populations. **The fact that it didn't make AYP in Florida but made AYP in most other states is likely because Florida schools report no confidence interval around their proficiency rates, making it more difficult to achieve their AMOs compared to states that do use confidence intervals.**

A second anomaly is Forest Lake, which didn't make AYP in 20 of 28 states, but made AYP in Florida. Table 2 shows that this school has a relatively homogeneous student body with no accountable subgroups other than its low-income and white populations. Florida's sliding minimum subgroup rule prevents Forest Lake from having to account for most traditionally disadvantaged populations.

Two middle school anomalies are seen in Table 3 as well. Walter Jones Middle School made AYP in 20 of 28 states but failed to make AYP in Florida, because of the math performance of the Hispanic and low-income populations. As with Wayne Fine Arts Elementary, this may be attributable to Florida's lack of use of confidence intervals, making it more difficult to achieve their AMOs than it is for states that do use them. On the other hand, Chaucer Middle School made AYP in Florida but failed to make AYP in 22 of the 27 other states. This is most likely attributable to the sliding minimum  $n$  policy in Florida, which means that Chaucer does not have to account for either its students with LEP population or its SWDs, two subgroups that present the greatest challenges in Florida.

These observations are consistent with the patterns shown in Table 6, which compares schools that make and do not make AYP on several academic and demographic dimensions. Within the sample, schools that make AYP do indeed show higher average student performance, but they also differ in the following ways: they have much smaller student populations, fewer subgroups (and thus fewer targets to meet)—at least at the elementary school level—and much lower percentages of low-income students.

## **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 fare under Florida's AYP rules and AMOs for 2008. We found that only 3 elementary schools and 1 middle school— 4 in all, from a sample of 36—would have made AYP in Florida. Looking across the 28 state accountability systems examined in the study, this puts Florida roughly in the middle of the sample distribution as shown in Figure 1. In addition, Florida is 1 of 6 states with a single middle school that made AYP in the sample.

There are several other factors of note about Florida: First, it does not apply confidence intervals (or margins of error) to its measurement of student proficiency rates. This means that schools will have greater difficulty achieving their AMOs than they would in states that employ confidence intervals. Second, the manner in which the state defines minimum  $n$  sizes means that Florida schools will have fewer subgroups for which it is held accountable than do schools in many other states.

The overriding goal of the federal NCLB is to eliminate educational disparities within and across states; it is important to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, Florida's No Child Left Behind accountability system is working exactly as Congress intended: it is 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. Most of the elementary schools and about half of the sample middle schools made AYP in Florida for their student populations as a whole, that is, without 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? Yes, schools should redouble their efforts to boost achievement for LEP students and SWDs, as for other students, but when almost no school is 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 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.