New Mexico

Executive Summary

The intent of the No Child Left Behind (NCLB) Act of 2001 is to hold schools accountable for ensuring that all of 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 New Mexico'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 New Mexico'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 each would make AYP under New Mexico's system as well as under the systems of 27 other states. We used school data and proficiency cut score[1] estimates from academic year 2005–2006, but applied them against New Mexico's AYP rules for academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- **We estimate that 14 of 18 elementary schools and 16 of 18 middle schools in our sample failed to make AYP in 2008 under New Mexico's accountability system**. This high failure rate is partly explained by our sample, which intentionally includes some schools with relatively large populations of low-performing students. But it's also partly explained by New Mexico's minimum n size for subgroups, which tends to be smaller than those used in most other states, meaning it holds more subgroups accountable for performance.[2]

- The smaller n size appears to be a factor in the number of schools making AYP in New Mexico, despite the state's low overall cut scores in reading and low annual proficiency targets in math and reading (e.g., the state demands that only 35% of students in grades six through eight reach math proficiency in 2008).

- Looking across the 28 state accountability systems examined in the study, we find that the number of elementary schools making AYP in New Mexico is exceeded in 12 other sample states (New Mexico ties with New Hampshire and Maine, each with 4 elementary schools making AYP). New Mexico is one of 10 states with 2 middle schools each that made AYP in the sample (see Figure 1).

There are some interesting dynamics that place **New Mexico** near the middle of the state distribution in terms of the number of schools making AYP. This is a state which has several rigorous requirements combined with more lenient ones. For example, New Mexico's cut scores in math are close to or above the 50th percentile, while reading cut scores mostly hover around the 30th percentile. So more rigor in math is coupled with less rigor in reading. New Mexico's 99 percent confidence interval provides schools with greater leniency than the more commonly used 95 percent confidence interval found in other states. However, New Mexico's minimum subgroup size is 25, which is smaller than most other states we examined. This means that schools in New Mexico will have more accountable subgroups than would similar schools in other states, making it difficult for large schools with many accountable subgroups to make AYP there.

---

[1] 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 New Mexico Standards Based Assessments.

[2] Keep in mind, however, that school size and n size are related (e.g., small n sizes make sense for small schools).
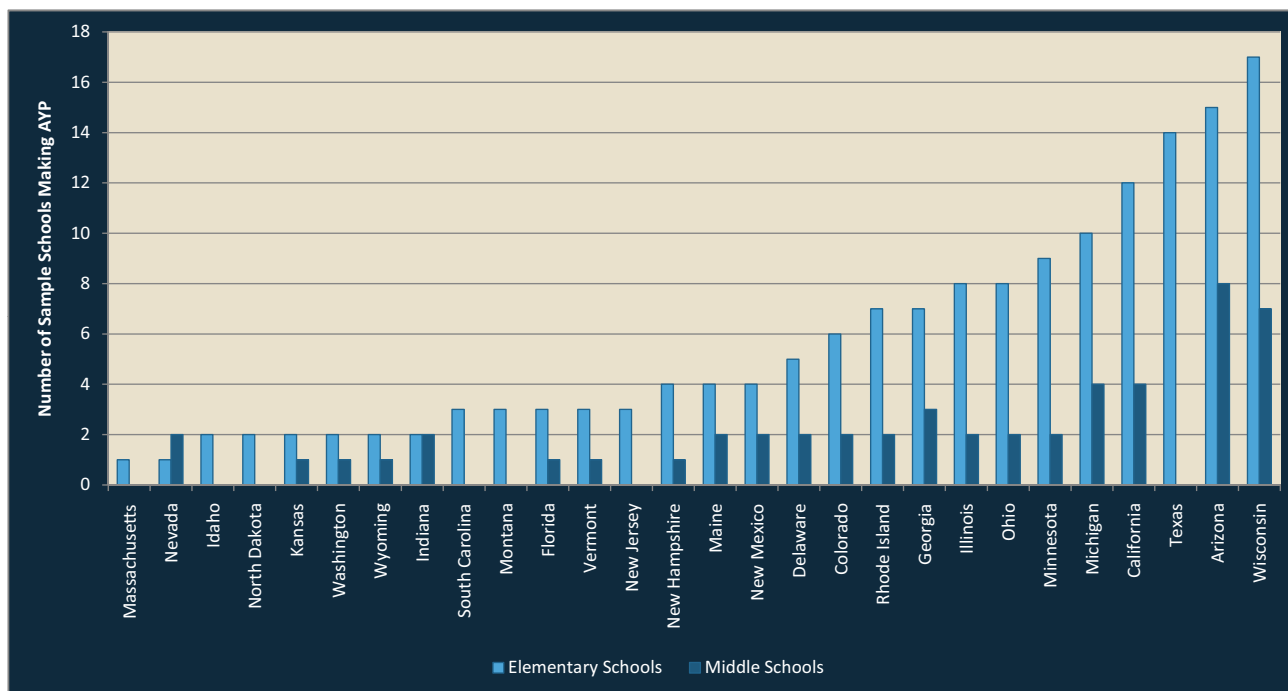
Thomas B. Fordham Institute

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.

- Nearly all of the schools in our sample that failed to make AYP in New Mexico are meeting expected targets for their overall populations[3] but failed because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.

- As in other states, middle schools in New Mexico had greater difficulty reaching AYP than did elementary schools, primarily because their student populations are larger and therefore have more qualifying subgroups—not because their student achievement is lower than in the elementary schools.

- Middle schools with fewer subgroups attained AYP more easily in New Mexico than middle schools with more subgroups, even when their average student performance is lower. In other words, schools with greater diversity and size face greater challenges in making AYP. This is the case in other states as well.

- A strong predictor of whether or not a school makes AYP under New Mexico's system is whether it has enough English language learners to qualify as a separate subgroup. Every single school with a limited English proficient (LEP)[4] subgroup failed to make AYP. Likewise, most of the schools (especially at the middle school level) with enough qualifying SWDs failed to meet their AYP targets.[5]

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance on New Mexico's tests and those of 25

---

[3] 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 are simply not treated as their own subgroup.

[4] Note that we use "LEP students" and "English language learners" interchangeably to refer to students in the same subgroup.

[5] 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 New Mexico Standards Based Assessments, 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.

other 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 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 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[6] 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 10 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 try 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.

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.[7]

---

[6] Low-income students are those who receive a free or reduced-price lunch.

[7] We gave all schools in our sample pseudonyms in this report.
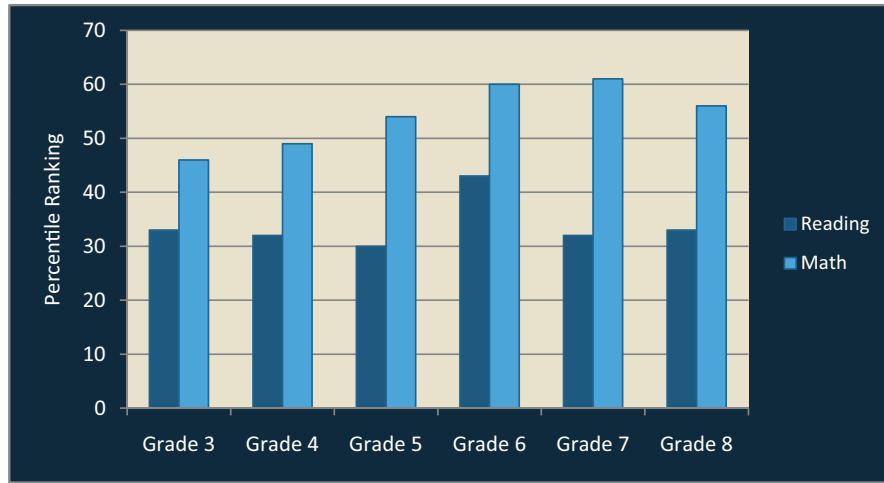
Figure 2. New Mexico reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of New Mexico's cut scores (or proficiency passing scores) for its reading and math tests, as percentiles of the NWEA norm, in grades three through eight. Higher percentile ranks are more difficult to achieve. All of New Mexico's cut scores in reading are below the 50th percentile, but the cut scores in math are close to or above the 50th percentile.

Table 1. New Mexico AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 25	
	SWDs: 25	
	Low-income students: 25	
	LEP students: 25	
CI	**Applied to proficiency rate calculations?**	
	Yes; 99% CI used	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	n/a	59
Grade 4	30	59
Grade 5	n/a	59
Grade 6	n/a	53
Grade 7	n/a	53
Grade 8	39	53
MATH		
Grade 3	n/a	44
Grade 4	35	44
Grade 5	n/a	44
Grade 6	n/a	35
Grade 7	n/a	35
Grade 8	33	35

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; n/a = not applicable
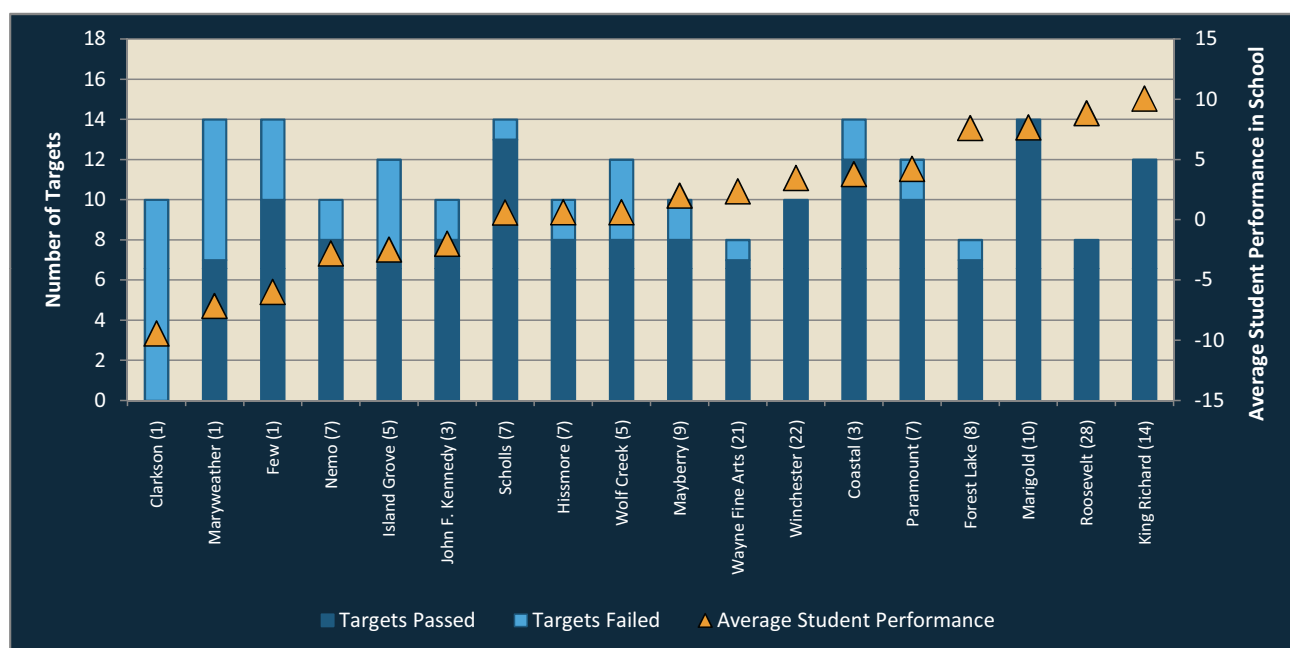
Figure 3. AYP performance of the elementary school sample under New Mexico's 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under New Mexico'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 AMOs for even a single subgroup didn't make AYP, so any light blue means that the school failed. Forest Lake, for example, met 7 of its 8 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.

Proficiency cut score estimates for the New Mexico Standards Based Assessments (NMSBA) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that New Mexico's definitions of proficiency generally ranked below average compared with the standards set by the other 25 states in that study. These cut scores were used to estimate whether students would have scored as proficient or better on the New Mexico 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 New Mexico AYP rules for 2008. In other words, the school data and our proficiency cut score estimates are from academic year 2005–2006, but we are applying them against New Mexico's 2008 AYP rules.

Table 1 shows the pertinent New Mexico AYP rules that we applied to elementary and middle schools in the current study. New Mexico's minimum subgroup size is 25, which is smaller than most other states we examined. This means that schools in New Mexico will have more accountable subgroups than would similar schools in other states.

Further, although most states also apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates, New Mexico's 99% confidence interval gives schools greater leniency than the more commonly used 95% confidence interval. So, for instance, although schools are supposed to get 59% of their grade 3 students (and 59% of grade 3 students in each subgroup) to the proficient level on the state reading test, applying the confidence interval means that the real target can be lower, particularly with smaller groups.[8]

---

[8] We also conducted an analysis to show the effect of confidence intervals on the reading and math proficiency rates for elementary and middle schools. We describe those results later in the report.
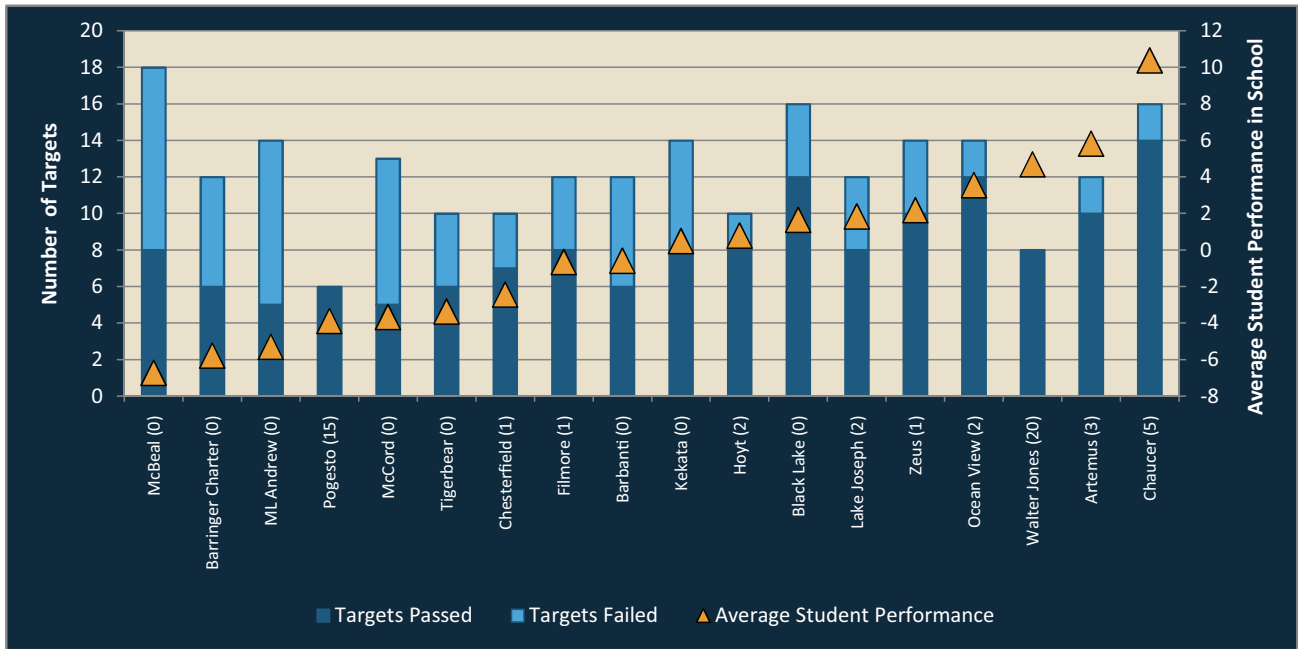
Figure 4. AYP performance of the middle school sample under New Mexico's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under New Mexico'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 failed to meet the AMOs for even a single subgroup did not make AYP, so any light blue means that the school failed. Artemus, for example, met 10 of its 12 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, and 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.

Figure 5. Impact of the confidence interval on elementary school math proficiency rates under New Mexico's 2008 AYP rules

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that one of the sample elementary schools (Maryweather) was assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.
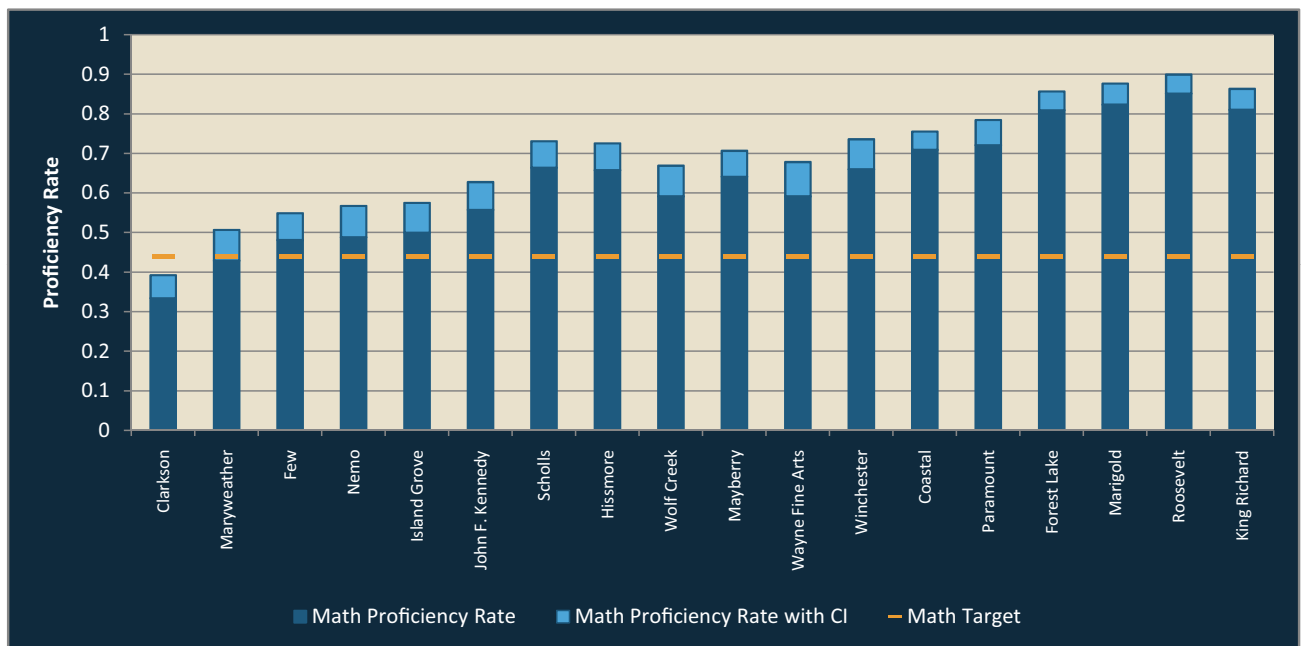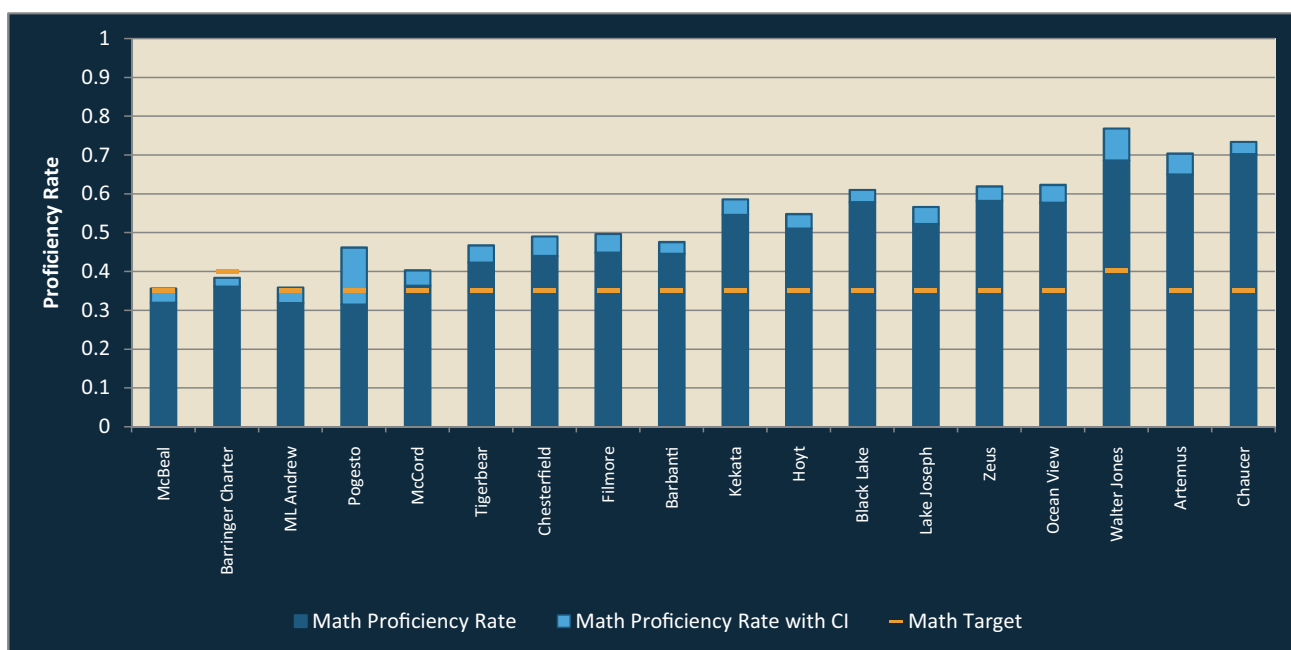
Figure 6. Impact of the confidence interval on middle school math proficiency rates under New Mexico's 2008 AYP rules

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that three sample middle schools (McBeal, ML Andrew, and Pogesto) were assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.

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 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 inter-

vals) 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 New Mexico's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under New Mexico's 2008 AYP rules. **Only 4 of 18 elementary schools (Winchester, Marigold, Roosevelt, and King Richard) 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 passing schools are in the right half of the figure, meaning that the highest average performing students were found in these schools.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 New Mexico AYP rules.

Table 2. Elementary school subgroup performance of sample schools under the 2008 New Mexico 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					
Clarkson	33.4%	42.3%	N	N	N	N	N	N	N	N					N	N					10	0	0%	N	1
Maryweather	42.9%	51.1%	Y	N	N	N	N	N	Y	N	Y	Y			Y	N			Y	Y	14	7	50%	N	1
Few	48.1%	54.3%	Y	Y	N	N	N	N	Y	Y	Y	Y			Y	Y			Y	Y	14	10	71%	N	1
Nemo	48.8%	67.9%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Island Grove	50.0%	67.5%	Y	Y	N	N	N	N	Y	Y					N	Y			Y	Y	12	7	58%	N	4
JFK	55.8%	61.2%	Y	Y	Y	N			Y	Y	Y	N							Y	Y	10	8	80%	N	3
Scholls	66.4%	69.5%	Y	Y	Y	N	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	13	93%	N	7
Hissmore	65.8%	73.3%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Wolf Creek	59.2%	67.6%	Y	Y	N	N	N	N	Y	Y					Y	Y			Y	Y	12	8	67%	N	5
Alice Mayberry	64.1%	75.4%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	9
Wayne Fine Arts	59.2%	83.3%	Y	Y					Y	Y	N	Y							Y	Y	8	7	88%	N	21
Winchester	66.0%	79.1%	Y	Y	Y	Y			Y	Y					Y	Y			Y	Y	10	10	100%	Y	22
Coastal	70.9%	76.0%	Y	Y	Y	N	Y	N	Y	Y	Y	Y			Y	Y			Y	Y	14	12	86%	N	3
Paramount	72.1%	76.1%	Y	Y	Y	Y	N	N	Y	Y					Y	Y			Y	Y	12	10	83%	N	7
Forest Lake	81.0%	84.9%	Y	Y	Y	N			Y	Y									Y	Y	8	7	88%	N	8
Marigold	82.4%	87.0%	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	Y	Y			Y	Y	14	14	100%	Y	10
Roosevelt	85.2%	92.2%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	28
King Richard	81.1%	89.5%	Y	Y	Y	Y	Y	Y	Y	Y					Y	Y			Y	Y	12	12	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.

Of 18 middle schools in our sample, only 2 made AYP—one low-performance school (Pogesto) and one high-performance school (Walter Jones), both of which have relatively few qualifying subgroups.

Figures 5 and 6 indicate the degree to which schools' overall math proficiency rates are aided by the confidence interval for elementary and middle schools, respectively. On these figures, the dark blue bars show the actual proficiency rates at each school, and the light blue bars show the degree to which these proficiency rates are "increased" by the application of the confidence interval. The orange lines show the AMO needed to meet AYP. These figures show that one of the sample elementary schools (Maryweather) and three middle schools (McBeal, ML Andrew, and Pogesto) are assisted by the confidence intervals. However, of the latter three, only Pogesto also meets all of its subgroup targets in order to make AYP (see Figure 4).

The effect of confidence intervals on schools' proficiency rates in reading is much the same (not shown). In reading, just one elementary school (Few) and one middle school (McBeal) met the overall target with the confidence interval, but we know from Figures 3 and 4 that both schools

Table 3. Middle school subgroup performance of sample schools under the 2008 New Mexico 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	32.0%	52.7%	Y	Y	N	N	N	N	N	N	N	Y	Y	Y	N	N	N	Y	Y	Y	18	8	44%	N	0
Barringer Charter	36.1%	57.1%	N	Y	N	N			N	Y	N	N			Y	Y			Y	Y	12	6	50%	N	0
ML Andrew	31.9%	55.9%	Y	Y	N	N	N	N	N	N	N	N			N	Y			Y	Y	14	5	36%	N	0
Pogesto	31.5%	66.7%	Y	Y							Y	Y							Y	Y	6	6	100%	Y	15
McCord Charter	36.3%	59.2%	Y	Y	N	N	N		N	N	N	N			N	Y			Y	Y	13	5	38%	N	0
Tigerbear	42.3%	56.9%	Y	Y	N	N			Y	Y	N	N							Y	Y	10	6	60%	N	0
Chesterfield	44.0%	58.6%	Y	Y	N	N			Y	Y	Y	N							Y	Y	10	7	70%	N	1
Filmore	44.9%	67.4%	Y	Y	N	N	N	N	Y	Y					Y	Y			Y	Y	12	8	67%	N	1
Barbanti	44.5%	62.8%	Y	Y	N	N	N	N	N	N					Y	Y			Y	Y	12	6	50%	N	0
Kekata	54.6%	66.7%	Y	Y	N	N	N	N	Y	Y	Y	N			Y	N			Y	Y	14	8	57%	N	0
Hoyt	51.1%	69.2%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	2
Black Lake	57.9%	69.2%	Y	Y	N	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y			Y	Y	16	12	75%	N	0
Lake Joseph	52.2%	74.3%	Y	Y	N	N	N	N	Y	Y					Y	Y			Y	Y	12	8	67%	N	2
Zeus	58.2%	70.5%	Y	Y	N	N	N	N	Y	Y	Y	Y			Y	Y			Y	Y	14	10	71%	N	1
Ocean View	57.7%	80.9%	Y	Y	Y	Y	N	N	Y	Y			Y	Y	Y	Y			Y	Y	14	12	86%	N	2
Walter Jones	68.6%	80.6%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	20
Artemus	65.0%	79.2%	Y	Y	Y	N			Y	Y			Y	Y	Y	N			Y	Y	12	10	83%	N	3
Chaucer	70.2%	85.3%	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	16	14	88%	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.

still failed to meet targets for some of their subgroups. Overall, **the application of the confidence interval had only modest impact on final AYP decisions for the sample elementary and middle schools in New Mexico**.[9]

Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low or middling performance can still pass AYP when the school has fewer targets to meet because it has fewer subgroups. These figures do not, however, indicate which subgroups failed or passed 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 students within that subgroup exceeded the state's

[9] In the current analyses, confidence intervals were applied to both the overall student population and to all eligible subgroups in our sample schools. Thus, the ultimate impact of the confidence interval may be larger than the impact depicted in Figures 5 and 6. However, we chose not to show how the confidence interval impacted subgroup performance because it would have added greatly to the report's length and complexity.

Table 4. Summary of subgroup performance of sample elementary schools under 2008 New Mexico 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	16	8	12
Students with limited English proficiency	10	6	7
Low-income students	18	1	2
African-American students	9	1	1
Asian/Pacific Islander students	1	0	0
Hispanic students	12	2	2
American Indian/Alaska Native students	0	0	0
White students	17	0	0

Table 5. Summary of subgroup performance of sample middle schools under 2008 New Mexico 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	16	14	14
Students with limited English proficiency	11	9	10
Low-income students	18	5	4
African-American students	11	5	7
Asian/Pacific Islander students	5	0	0
Hispanic students	14	3	3
American Indian/Alaska Native students	1	1	0
White students	18	0	0

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 New Mexico

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:

- Almost all schools met their reading and math targets for their overall school population.

- Just two elementary schools (Clarkson and Mary-

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	4	14	2	16
Average student body size	225	328	124	951
Average % low income	14	56	42	45
Average % nonwhite	25	45	27	46
Average performance†	7.51	-0.57	0.40	-0.11
Average % growth‡	126	112	109	97
Average number of targets to meet	11	11	7	13

† 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.

weather) failed to meet the reading targets for their overall school population. One failed to meet its math target for the overall population.

- Only one middle school (Barringer) failed to meet its overall math target, and none failed to meet overall reading targets.

- Other subgroups (low income, Hispanic, and African American, among others) performed fairly well at the elementary level.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively. First, the performance of SWDs is proving challenging for schools under New Mexico's system, where this subgroup tends to have enough students to meet the state's minimum *n* of 25. In fact, all but one middle school in the study with qualifying SWD subgroups failed to make AYP (Ocean View Middle missed because of its students with LEP subgroup). Students with LEP and African American students are also struggling to meet the state's middle school targets (which are not as problematic for Hispanic or low-income students).

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that New Mexico's NCLB accountability system is, in most respects, behaving like those in other states. For example, Roosevelt, Winchester, and King Richard are among the schools that made AYP in the greatest number of states—28, 22, and 14, respectively. And these schools all made AYP in New Mexico, too. Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make AYP in New Mexico.

But New Mexico is also home to a few anomalies. First, consider Wayne Fine Arts (see Table 2). It made AYP in 21 of the 28 states in our sample, yet failed to make AYP in New Mexico. In examining Table 2, we can see that the subgroup of African American students failed to meet its target in math. Second, look at Pogesto Middle School (Table 3). Even with its relatively low average performance, it made AYP in New Mexico, but failed to do so in 13 of 28 states. Like Wayne Fine Arts, its AYP success in New Mexico is most likely attributable to the relatively small number of targets (six) it has to meet, as shown in Figure 4.

This is consistent with the patterns shown in Table 6, which compares schools that do and don't make AYP on a number of academic and demographic dimensions. Within the sample, schools that make AYP do indeed show higher average student performance, but they also have much smaller student populations and much lower percentages of nonwhite students. Surprisingly, though, the elementary schools that make AYP have the same number of subgroups (and thus same targets to meet). Middle schools that make AYP have slightly higher performing students, on average, than middle schools that don't, but have drastically smaller total enrollments, smaller nonwhite populations, and fewer subgroups (and thus targets to meet).

Concluding Observations

This study examined evaluated the test performance data of students from 18 elementary and 18 middle schools across the country to see how these schools would fare under New Mexico's AYP rules (and AMOs for 2008). Among this sample, only 4 elementary schools and 2 middle schools—6 in all from a total of 36—would have made AYP in New Mexico. Looking across the 28 state accountability systems examined in the study, this puts New Mexico roughly in the middle of the sample distribution, as shown in Figure 1. The fairly high failure rate in New Mexico is perhaps partly explained by the state's minimum *n* size for subgroups, which tends to be smaller than those used in most other states, meaning it holds more subgroups accountable for performance (this despite the state's low overall cut scores in reading and low annual proficiency targets in math and reading).

Because the overriding goal of NCLB is to eliminate education 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, New Mexico's NCLB accountability system is working exactly as Congress intended: identifying as "needing attention" schools with relatively high test score averages that mask low performance for particular groups of students, such as SWD, LEP, or African American students. Almost all of the sample schools made AYP in New Mexico for their student populations as a whole. In the pre-NCLB era, such schools might have been considered to be effective or at least not in need of improvement, even though sizable numbers of their pupils aren't 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 the participation guidelines for English language learners and students with disabilities are more generous under the current state assessment system,[10] doesn't the massive failure of these students (particularly in middle school) to meet New Mexico'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 LEP students and students with disabilities, 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

---

[10] See footnote 5.

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.