



## 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 Vermont’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 Vermont’s system is compared with other states. For this study, we selected 36 schools from around the nation, schools that vary by size, achievement, and diversity, among other factors, and determined whether or not each would make AYP under Vermont’s 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 Vermont’s AYP rules 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 fail to make adequate yearly progress in 2008 under Vermont’s accountability system. This high failure rate is partly explained by our sample, which intentionally includes some schools with a relatively large population of low-performing students. But **it’s also partly explained by Vermont’s annual proficiency targets,**

which are fairly rigorous (roughly 87 percent of Vermont’s grade 3-8 students are expected to be proficient in reading in 2008).

- Looking across the 28 state accountability systems examined in the study, we find **Vermont at about the middle of the distribution in terms of the number of elementary sample schools making AYP.** Specifically, it exceeds fifteen states and ties with four others (South Carolina, Montana, Florida and New Jersey) (See Figure 1).
- Some of the schools in our sample that failed to make AYP in Vermont are meeting expected targets for their overall populations but failing because of the performance of individual subgroups.<sup>2</sup>
- In Vermont, as in most states, schools with fewer subgroups attain AYP more easily than schools with more subgroups, even when their average student performance is much lower. In other words, schools

Fifteen of 18 elementary schools and 17 of 18 middle schools in our sample fail to make AYP in 2008 under **Vermont’s** accountability system. This places Vermont at about the middle of the state distribution in terms of the number of schools making AYP. Vermont’s proficiency standards are about average compared to other states, but its annual targets are fairly rigorous (roughly 87 percent of grade 3-8 students are expected to be proficient in reading in 2008). Unlike most states, Vermont measures its student performance with a proficiency index, which gives partial credit for students achieving “partial proficiency.” In the short term, the index makes it easier for Vermont schools to meet their targets, but the effect of the index diminishes as the targets approach the 100 percent proficiency requirement dictated under NCLB for 2014.

<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 New England Common Assessment Program (NECAP).

<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.

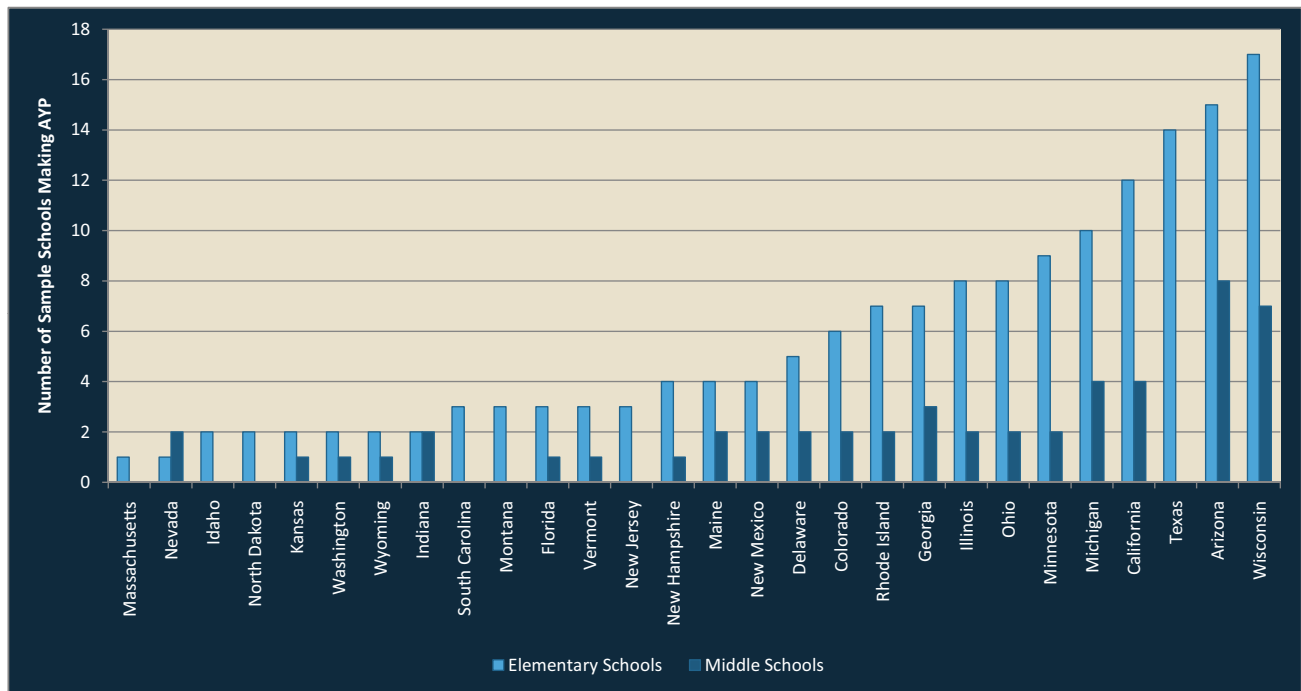


Figure 1. Number of sample schools making AYP by state

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.

with greater diversity and size face greater challenges in making AYP.

- Middle schools have greater difficulty reaching AYP in Vermont than do 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.
- A strong predictor of whether or not a school will make AYP under the Vermont system is whether it has enough students with disabilities (SWDs)<sup>3</sup> or English language learners to qualify as a separate subgroup. In fact, all schools with limited English proficient (LEP)<sup>4</sup> or SWD subgroups failed to make AYP.

## Introduction

*The Proficiency Illusion* (Cronin et al. 2007a) linked student performance on Vermont’s tests and those of 25 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

<sup>3</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 New England Common Assessment Program (NECAP), 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.

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

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 poten-

tial 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**

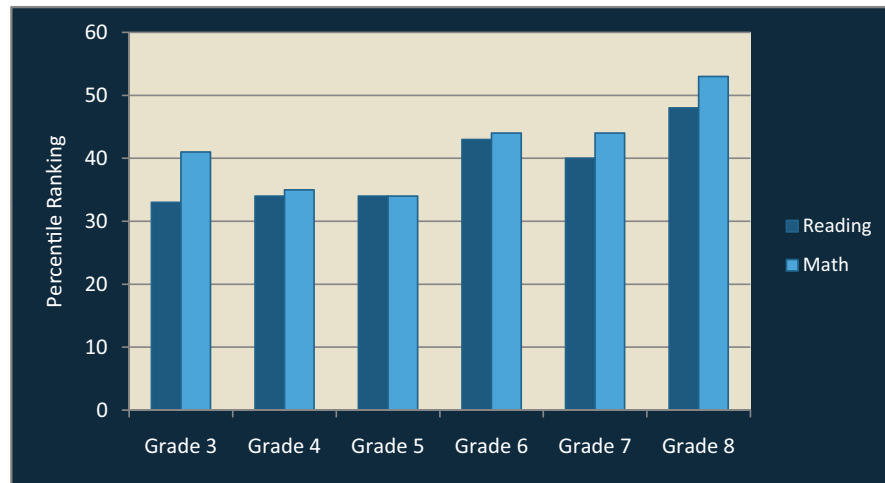
We collected students’ MAP test scores from the 2005–2006 academic year from 18 elementary and 18 middle schools around the country. We also collected the NCLB subgroup designations for all students in those schools—in other words, whether they had been classified as members of a minority group or as 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.<sup>6</sup>

Proficiency cut score estimates for the New England Common Assessment Program (NECAP) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Vermont’s proficiency cut scores were generally ranked about 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 Vermont 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 Vermont AYP rules for 2008. In other words, the school data and proficiency cut score esti-

<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.** Vermont reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Vermont’s cut scores (or proficiency passing scores) for the state’s 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 the state’s cut scores are below the 55th percentile.

mates are from academic year 2005–2006, but we are applying them against Vermont’s 2008 AYP rules.

Table 1 shows the pertinent Vermont AYP rules that were applied to elementary and middle schools in the current study. Vermont’s minimum subgroup size is 40, which is comparable to most other states we examined. Most states examined also apply confidence intervals (or margins of error) to their measurements of student proficiency rates. However, Vermont’s 99% confidence interval provides schools with greater leniency than the more commonly used 95% confidence interval. This means that while schools are supposed to get 87% of their grade 3-8 students to the proficient level on the state reading test, as well as 87% of the students in each subgroup, applying the confidence interval means that the real target can be lower, particularly with smaller groups.

Unlike most states, Vermont measures its student performance with a proficiency index, which gives partial credit for students achieving “partial proficiency.” In the short term, the index makes it easier for Vermont schools to meet their targets, although the effect of the index diminishes as the targets approach the 100% proficiency requirement dictated under NCLB for 2014.<sup>7</sup>

**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. (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

<sup>7</sup> In six of the states studied (Massachusetts, Minnesota, Rhode Island, New Hampshire, and Wisconsin, as well as Vermont), an index is used that gives full credit to students who achieve proficient (or better) and partial credit to students performing at lower levels. Consequently, the resultant score in states using this “hybrid” model is always higher than the actual proficiency percentage (giving students partial credit for achieving lower proficiency levels is obviously better than no credit, at least for the schools’ ratings). The index provides a fair amount of help when annual targets are below 50%; however, once targets rise above 75%, the index has far less impact.

Table 1. Vermont AYP rules for 2008

Subgroup minimum <i>n</i>	Race/ethnicity: 40	
	SWDs: 40	
	Low-income students: 40	
	LEP students: 40	
CI	Applied to proficiency rate calculations?	
	Yes; 99% CI used	
AMOs	Baseline proficiency levels as of 2002 (index)	2008 targets (index)
READING/LANGUAGE ARTS		
Grade 3	n/a	87.0
Grade 4	n/a	87.0
Grade 5	n/a	87.0
Grade 6	n/a	87.0
Grade 7	n/a	87.0
Grade 8	n/a	87.0
MATH		
Grade 3	n/a	85.4
Grade 4	n/a	85.4
Grade 5	n/a	85.4
Grade 6	n/a	85.4
Grade 7	n/a	85.4
Grade 8	n/a	85.4

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

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 Vermont's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Vermont's 2008 AYP rules. **Only three elementary schools made AYP while fifteen 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. All schools making AYP are in the right half of the figure, meaning that they are among the schools which contain the highest average performing students.

Yet among these schools with high average performing students, the only schools actually to make AYP are those with relatively few qualifying subgroups—and thus the fewest targets to meet (since each subgroup has its own separate targets). For example, Wayne Fine Arts, Winchester and Roosevelt made it, but have only four targets each—two in reading and math for their overall populations, and two in reading and math for the only subgroup that exceeds Vermont's minimum “*n* size”: white students.

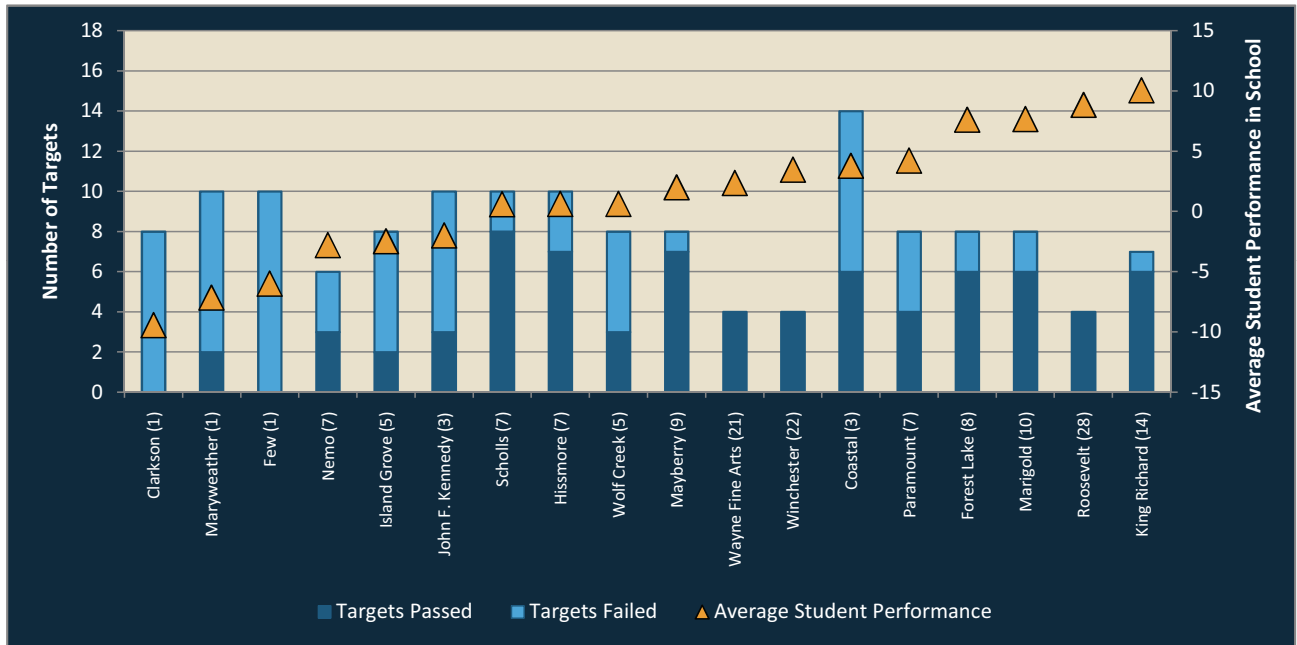


Figure 3. AYP performance of the elementary school sample under Vermont's 2008 AYP rules

Note: This figure shows how each of the elementary schools within the sample fared under the Vermont 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. Marigold Elementary, for example, meets 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.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Vermont AYP rules. **Out of eighteen in our sample, only one middle school made AYP**—Walter Jones—a high-performing school with relatively few qualifying subgroups.

Figures 5 and 6 indicate the degree to which schools' math proficiency rates are aided by the confidence interval for elementary and middle schools, respectively. On these figures, the darker portions of the bars show the actual proficiency rates at each school and the lighter portions of the bars show the degree to which these proficiency rates were increased by applying the confidence interval. The orange lines show the AMOs needed to meet AYP. The figures show that one elementary (JFK) and no middle schools are assisted in meeting their over-

all math targets by the confidence intervals. However, JFK still failed to make AYP due to the performance of multiple subgroups (see Figure 3).

The effect of the confidence intervals on reading proficiency rates at the elementary and middle school levels is similar (not shown). In reading, two elementary schools (Hissmore and Paramount) and two middle schools (Pogesto and Artemus) were able to meet the overall target with the confidence interval, although we know from Figures 3 and 4 that these schools still failed to meet targets for their subgroups. **In short, applying the confidence interval (even a generous one like the 99% confidence interval used in Vermont) has little or no effect on whether schools meet their overall reading and math targets in Vermont.**<sup>8</sup>

<sup>8</sup> In the current analyses, confidence intervals were applied to both the overall school population and to all eligible subgroups in our sample schools. Thus, the ultimate impact of the confidence interval is likely 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.

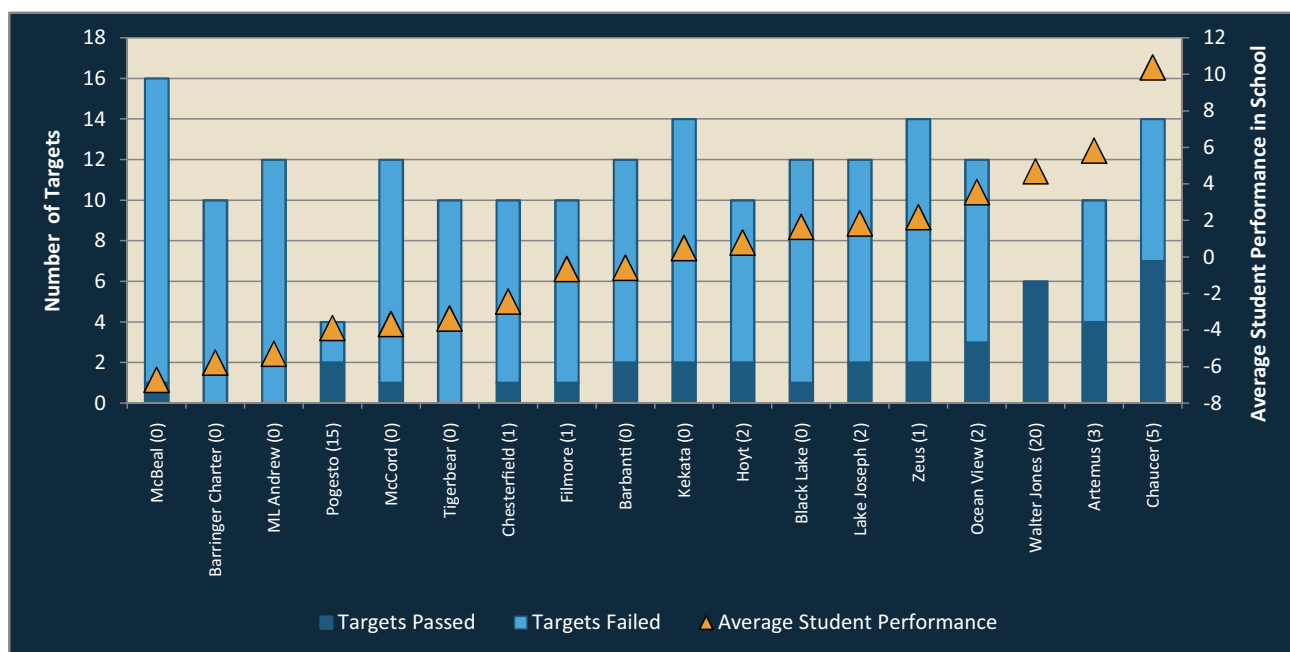


Figure 4. AYP performance of the middle school sample under Vermont's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under the AYP rules in Vermont (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. Chaucer, for example, meets seven of its fourteen 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.

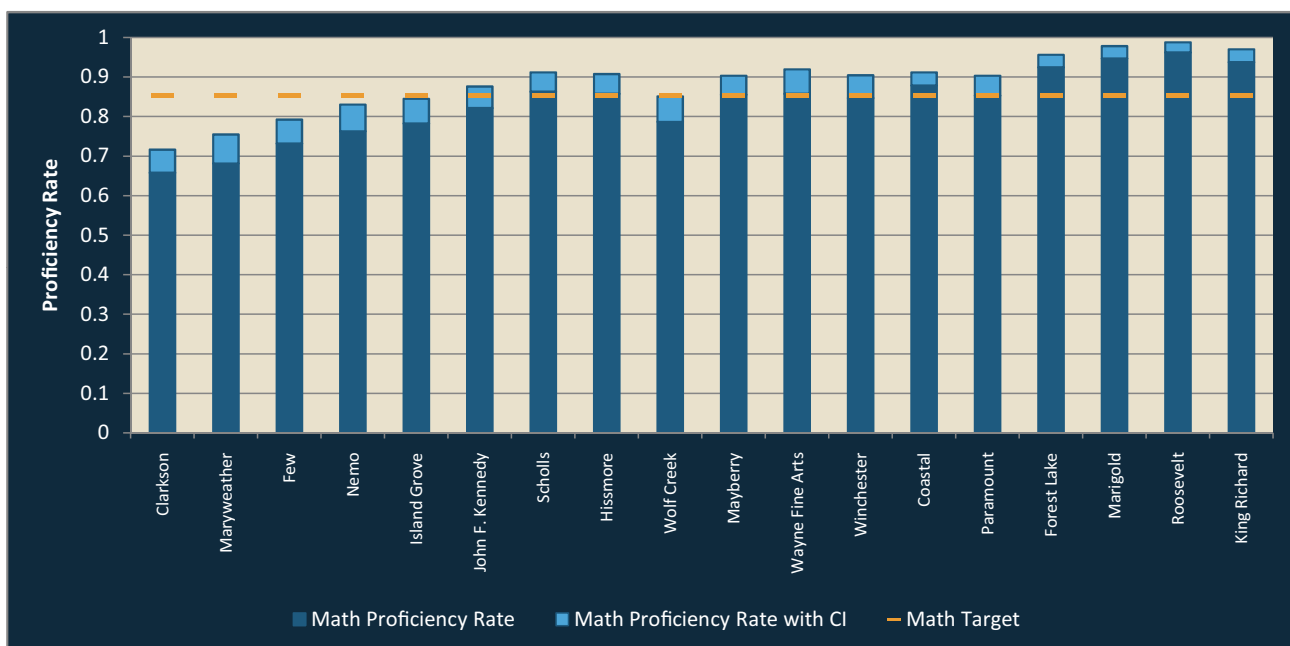
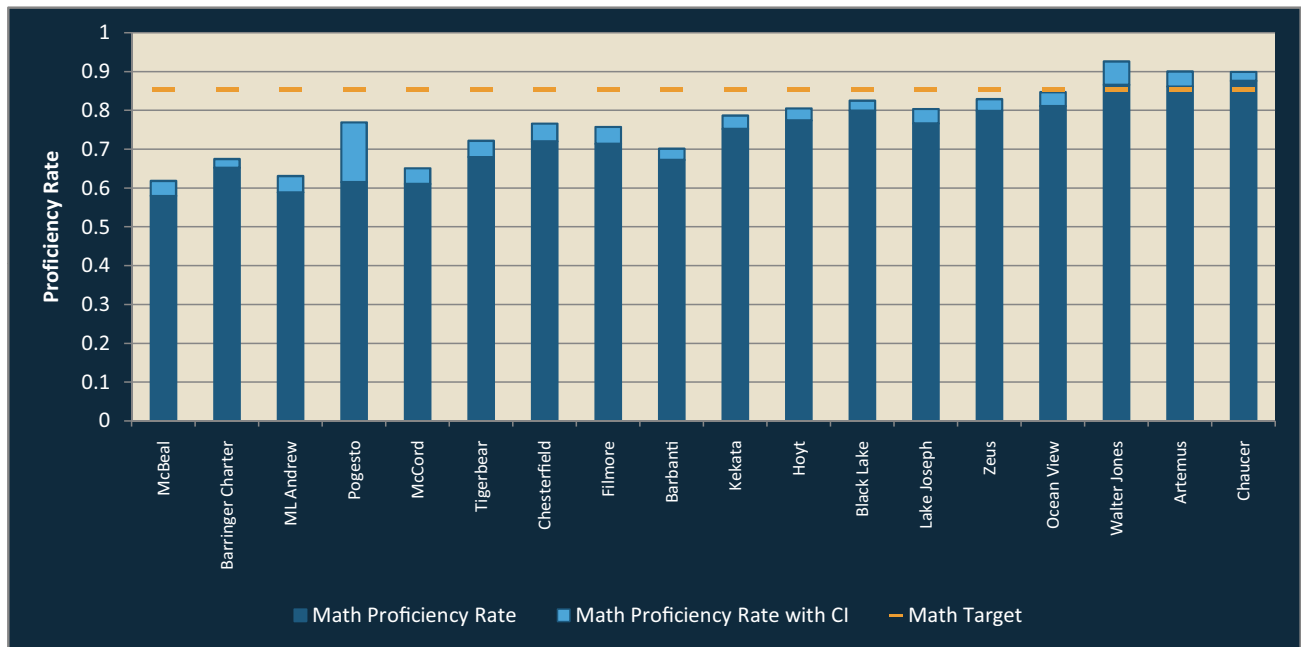


Figure 5. Impact of the confidence interval on elementary school math proficiency rates

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 (JFK) 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

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 none of the sample middle schools 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.

### 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 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 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 Vermont 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

- Four elementary schools failed to meet both their overall reading and math targets.
- Thirteen middle schools failed to meet both their reading and math targets for their overall populations.
- Three elementary schools (Scholls, Forest Lake, and King Richard) failed for their SWD subgroup only.
- One elementary school (Alice Mayberry) met targets for every subgroup except for its low income students.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. First, the performance of SWDs and LEP students were particularly challenging for Vermont schools. Every single school with enough students to comprise a SWD or LEP subgroup failed to make AYP, in part due to these groups’ performances. Traditionally academically disadvantaged subgroups, such as low income and Hispanic students, also had difficulty under Vermont’s accountability system, especially at the middle school level.



Table 2. Elementary school subgroup performance of sample schools under the 2008 Vermont 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	65.8%	62.0%	N	N			N	N	N	N					N	N						8	0	0%	N	1
Maryweather	68.2%	66.4%	N	N			N	N	N	N					N	N				Y	Y	10	2	20%	N	1
Few	73.2%	69.7%	N	N	N	N	N	N	N	N					N	N						10	0	0%	N	1
Nemo	76.3%	81.5%	N	Y					N	N									Y	Y	6	3	50%	N	7	
Island Grove	78.3%	79.9%	N	N					N	N					N	N				Y	Y	8	2	25%	N	4
JFK	82.2%	78.0%	Y	N	N	N			N	N	N	N							Y	Y	10	3	30%	N	3	
Scholls	86.3%	81.7%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7	
Hissmore	85.7%	84.3%	Y	Y	N	N			Y	Y	Y	N							Y	Y	10	7	70%	N	7	
Wolf Creek	78.7%	81.2%	N	Y					N	N					N	N				Y	Y	8	3	38%	N	5
Alice Mayberry	85.5%	86.2%	Y	Y					Y	N	Y	Y							Y	Y	8	7	88%	N	9	
Wayne Fine Arts	85.8%	92.4%	Y	Y															Y	Y	4	4	100%	Y	21	
Winchester	84.7%	88.3%	Y	Y															Y	Y	4	4	100%	Y	22	
Coastal	87.9%	83.4%	Y	Y	N	N	N	N	Y	N	Y	N			N	N			Y	Y	14	6	43%	N	3	
Paramount	85.3%	84.6%	Y	Y					N	N					N	N			Y	Y	8	4	50%	N	7	
Forest Lake	92.4%	92.0%	Y	Y	N	N			Y	Y									Y	Y	8	6	75%	N	8	
Marigold	94.7%	91.2%	Y	Y	Y	N			Y	N									Y	Y	8	6	75%	N	10	
Roosevelt	96.2%	96.2%	Y	Y															Y	Y	4	4	100%	Y	28	
King Richard	93.8%	93.5%	Y	Y	Y	N			Y										Y	Y	7	6	86%	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.

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

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

in Vermont. A striking difference between schools that consistently make and don't make AYP, appears to be the number of subgroups for which each is held accountable—and hence, the number of academic targets for which each must demonstrate proficiency.

This is consistent with the patterns shown in Table 6, which compares the schools that did and didn't make AYP on several academic and demographic dimensions. Within the sample, elementary schools that make AYP do indeed show higher average student performance, but they also differ in the following ways: they have much smaller stu-

**Table 3.** Middle school subgroup performance of sample schools under the 2008 Vermont 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	58.0%	65.0%	N	N	N	N	N	N	N	N	N	N			N	N	N	N	N	N	Y	16	1	6%	N	0	
Barringer Charter	65.3%	73.1%	N	N	N	N			N	N	N	N			N	N						10	0	0%	N	0	
ML Andrew	58.9%	71.2%	N	N	N	N			N	N	N	N			N	N					N	N	12	0	0%	N	0
Pogesto	61.6%	74.5%	N	Y																	N	Y	4	2	50%	N	15
McCord Charter	61.1%	73.9%	N	N	N	N			N	N	N	N			N	N					N	Y	12	1	8%	N	0
Tigerbear	68.0%	68.8%	N	N	N	N			N	N	N	N									N	N	10	0	0%	N	0
Chesterfield	72.1%	72.9%	N	N	N	N			N	N	N	N									Y	N	10	1	10%	N	1
Filmore	71.4%	78.9%	N	N	N	N			N	N					N	N					N	Y	10	1	10%	N	1
Barbanti	67.3%	74.1%	N	N	N	N	N	N	N	N					N	N					Y	Y	12	2	17%	N	0
Kekata	75.3%	76.5%	N	N	N	N	N	N	N	N	N	N			N	N					Y	Y	14	2	14%	N	0
Hoyt	77.5%	79.3%	N	N	N	N			N	N	N	N									Y	Y	10	2	20%	N	2
Black Lake	80.0%	78.7%	N	N	N	N			N	N	N	N			N	N					Y	N	12	1	8%	N	0
Lake Joseph	76.7%	82.4%	N	N	N	N	N	N	N	N					N	N					Y	Y	12	2	17%	N	2
Zeus	79.9%	80.5%	N	N	N	N	N	N	N	N	N	N			N	N					Y	Y	14	2	14%	N	1
Ocean View	81.1%	87.3%	N	Y	N	N	N	N	N	N					N	N					Y	Y	12	3	25%	N	2
Walter Jones	86.6%	88.9%	Y	Y					Y	Y											Y	Y	6	6	100%	Y	20
Artemus	86.2%	85.9%	Y	Y	N	N			N	N					N	N					Y	Y	10	4	40%	N	3
Chaucer	87.7%	91.0%	Y	Y	N	N	N	N	N	N				Y	Y	N	Y				Y	Y	14	7	50%	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.

dent populations, fewer subgroups (and thus fewer targets to meet), and lower percentages of low-income students. Similarly, middle schools that make AYP have slightly higher performing students, on average, than middle schools that failed to make it, but have dramatically smaller total enrollments, smaller nonwhite populations, and fewer subgroups (and thus 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 Vermont's AYP rules and annual measurable objectives 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 Vermont. Looking across the 28 state accountability systems examined in this study, this puts Vermont at about the middle of the distribution in terms of the number of elementary sample schools making AYP (as shown in Figure 1).

Because the overriding goal of NCLB is to eliminate educational disparities within and across states, it's impor-

**Table 4.** Summary of subgroup performance of sample elementary schools under the 2008 Vermont 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	6	8
Students with limited English proficiency	4	4	4
Low-income students	15	8	11
African-American students	5	1	3
Asian/Pacific Islander students	0	0	0
Hispanic students	7	7	7
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 Vermont 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	16	16
Students with limited English proficiency	7	7	7
Low-income students	17	16	16
African-American students	10	10	10
Asian/Pacific Islander students	1	0	0
Hispanic students	13	13	12
American Indian/Alaska Native students	1	1	1
White students	17	6	4

tant to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, Vermont's No Child Left Behind 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 low-income or Hispanic students. Some of the sample schools made AYP in Vermont for their student popula-

tions 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 weren't meeting state standards. Disaggregating data by race, income, etc. has made those students visible. That is surely a good thing.

Yet NCLB's design flaws are also readily apparent. Does it make sense that the size of a school's enrollment has

Table 6. Comparisons between schools that did and didn't make AYP in Vermont, 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	225	321	165	900
Average % low income	16	52	38	45
Average % nonwhite	27	44	33	45
Average performance <sup>†</sup>	4.89	0.49	4.69	-0.33
Average % growth <sup>‡</sup>	113	115	111	97
Average number of targets to meet	4	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.

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>9</sup> doesn't the massive failure of middle school students to meet Vermont's targets indicate that a new approach is needed for holding schools accountable for the performance of these students? Is it "fair" that, in

Vermont and in a handful of other states, students are awarded "partial" credit even though they do not achieve proficiency? 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

<sup>9</sup> See footnote 3.

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.