



## 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 Minnesota'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 Minnesota'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 Minnesota'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 Minnesota's AYP rules for academic year 2007–2008 (shortened to “2008” in this report).

Here are some key findings:

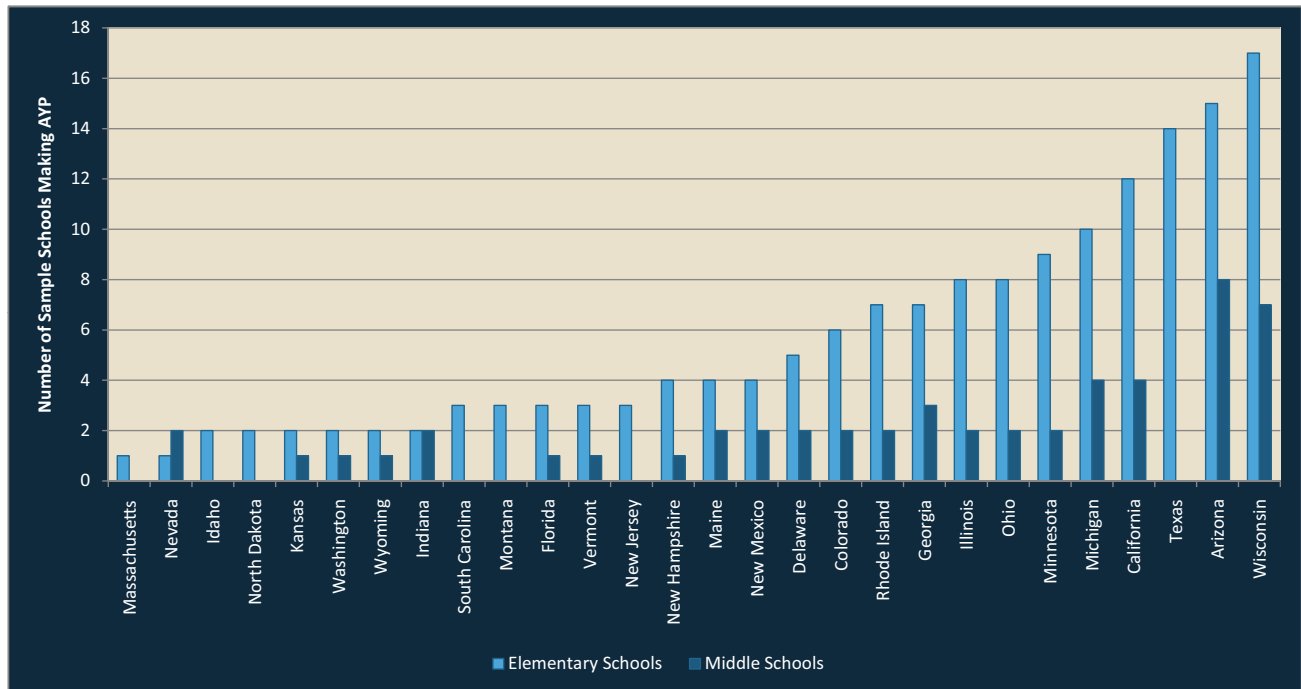
- We estimate that **9 of 18 elementary schools** and **16 of 18 middle schools** in our sample failed to make adequate yearly progress in 2008 under Minnesota's accountability system. (This rate is partly explained by our sample, which intentionally includes some schools with a relatively large population of low-performing students.)
- Looking across the 28 state accountability systems examined in the study, we find that the number of elementary schools making AYP in Minnesota was

exceeded in just 5 other sample states (Michigan, California, Texas, Arizona, and Wisconsin)(see Figure 1).

- Most of the schools in our sample that failed to make AYP in Minnesota are meeting expected targets for

Compared with other states in the study, **Minnesota** is at the high end of the distribution in terms of how many sample schools make AYP. There are several reasons for this. First, while the majority of states examined apply confidence intervals (margins of error) to their measurements of student proficiency rates, Minnesota uses a “sliding” confidence interval range of 95-99 percent, which is highly unusual. These varying confidence intervals make it easier for Minnesota schools to achieve their targets, with schools that have many subgroups receiving more of a “boost” than schools with fewer targets. Second, Minnesota's minimum subgroup size varies by subgroup. Racial, ethnic, and low-income subgroups have lower minimum *n* sizes than do students with disabilities (SWD) and limited English proficient (LEP) subgroups. Because of this, there are fewer subgroups of SWD and LEP students in elementary schools than in middle schools, which tend to be bigger. Therefore, more elementary schools make AYP. Finally, while most states measure school performance by a proficiency rate (or percentage of students achieving “proficient” or higher on the state test), Minnesota employs a performance “index” which gives partial credit to students attaining “partial proficiency.” The resultant score for students in Minnesota is always higher than the actual proficiency percentage (i.e., giving students partial credit for achieving lower proficiency levels is obviously better than no credit, at least for the schools' ratings).

<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 Minnesota Comprehensive Assessments – Series II.



**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.

their overall populations but failed because of the performance of individual subgroups, particularly English language learners and students with disabilities (SWDs) in middle schools.<sup>2</sup>

- Schools with fewer subgroups attained AYP more easily in Minnesota than schools with more subgroups, even when their average student performance is much lower. In other words, schools with greater diversity and size face greater challenges in making AYP. This is the case in other states as well.
- Middle schools had greater difficulty reaching AYP in Minnesota than did elementary schools, primarily because their student populations are larger and therefore have more qualifying subgroups—not be-

cause their student achievement is lower than in the elementary schools.

- A strong predictor of whether or not a school would make AYP under Minnesota’s system is whether it has enough English language learners to qualify as a separate subgroup. Every school with a limited English proficient (LEP) subgroup failed to make AYP.<sup>3</sup> Likewise, almost all middle schools with enough qualifying SWDs failed to meet their AYP targets.<sup>4</sup>
- Overall, the application of the confidence interval had some impact on AYP decisions for the sample elementary and middle schools in Minnesota, several of which were assisted in meeting their overall reading and math targets.

<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.

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

<sup>4</sup> SWDs are defined as those students following individualized education plans. We should also note that our subgroup findings for LEP students and SWDs may be more negative than actual findings, mostly because of the likely differences between how LEP students and SWDs are treated in MAP, the assessment we used in this study, and in the Minnesota Comprehensive Assessments – Series II, 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.

## Introduction

*The Proficiency Illusion* (Cronin et al. 2007a) linked student performance on Minnesota'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 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. 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' 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.

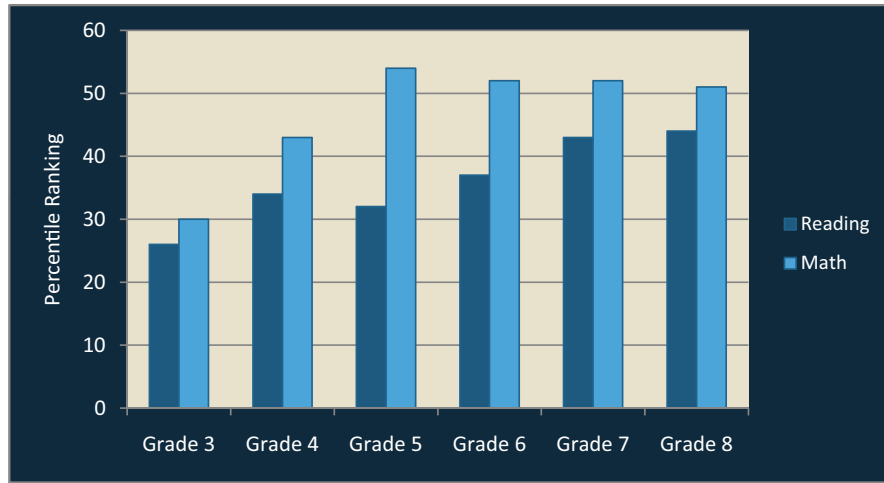
All of these AYP rules vary by state, which means that a school that makes AYP in Wisconsin or Ohio, for example, might not make it under South Carolina's or Idaho's rules (U.S. Department of Education 2008).

## 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 discus-

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



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

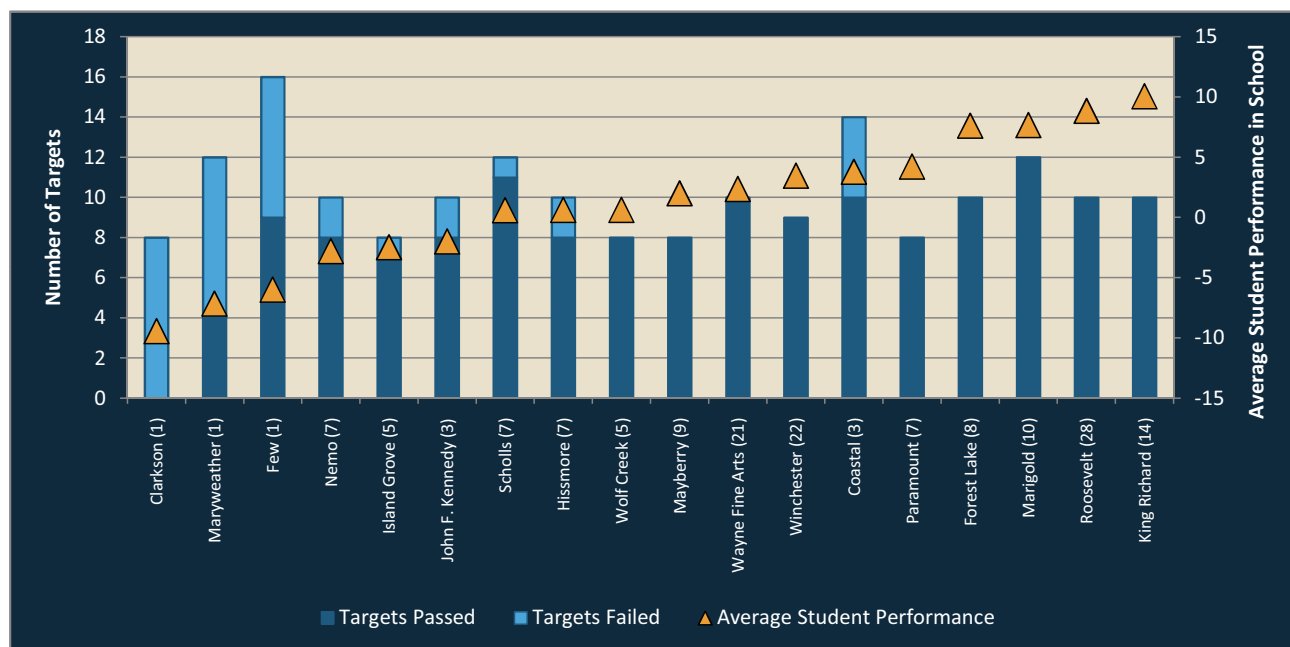
Note: This figure illustrates the difficulty of Minnesota's cut scores (or proficiency passing scores) for its reading and math tests, as percentiles of the NWEA norm, in grades three through eight. Higher percentile ranks are more difficult to achieve. All of Minnesota's reading cut scores are above the 25th percentile and most of the math cut scores are above the 40th percentile.

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

Subgroup minimum <i>n</i>	Race/ethnicity: 20	
	SWDs: 40	
	Low-income students: 20	
	LEP students: 40	
CI	Applied to proficiency rate calculations?	Additional notes:
	Yes, 95%–99% CI, depending on how many subgroups	Confidence interval grows more lenient with more subgroups
AMOs	Baseline proficiency levels as of 2002 (index)	2008 targets (index)
READING/LANGUAGE ARTS		
Grade 3	72.2	76.2
Grade 4	69.5	73.8
Grade 5	71.9	75.9
Grade 6	70.3	74.5
Grade 7	65.6	70.5
Grade 8	64.0	69.2
MATH		
Grade 3	78.9	81.9
Grade 4	69.6	73.9
Grade 5	59.8	65.5
Grade 6	59.9	65.6
Grade 7	58.8	64.7
Grade 8	58.3	64.3

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 Minnesota's 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under Minnesota'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. Hissmore Elementary, for example, met 8 of its 10 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.

sion of the methodology for this project along with the characteristics of the school sample.<sup>6</sup>

Proficiency cut score estimates for the Minnesota Comprehensive Assessments – Series II (MCA-II) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Minnesota's definitions of proficiency generally ranked above the 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 Minnesota 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 Minnesota 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 Minnesota's 2008 AYP rules.

Table 1 shows the pertinent Minnesota AYP rules that we applied to elementary and middle schools in this study. Minnesota's minimum group size varies by subgroup, with race/ethnic groups and low-income groups at 20 students, and SWDs and LEP groups at 40 students. **Forty is about average, compared to most other states, and 20 is smaller than most.**<sup>7</sup> This means that schools in Minnesota may have more accountable subgroups than similar schools in other states. However, because of school size, there are fewer subgroups of SWD and LEP students in elementary schools than in middle schools. This enables more elementary schools to make AYP.

Furthermore, although the majority of states examined in this study apply confidence intervals (margins of sta-

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

<sup>7</sup> Keep in mind, however, that school size and *n* size are related (e.g., small *n* sizes make sense for small schools).

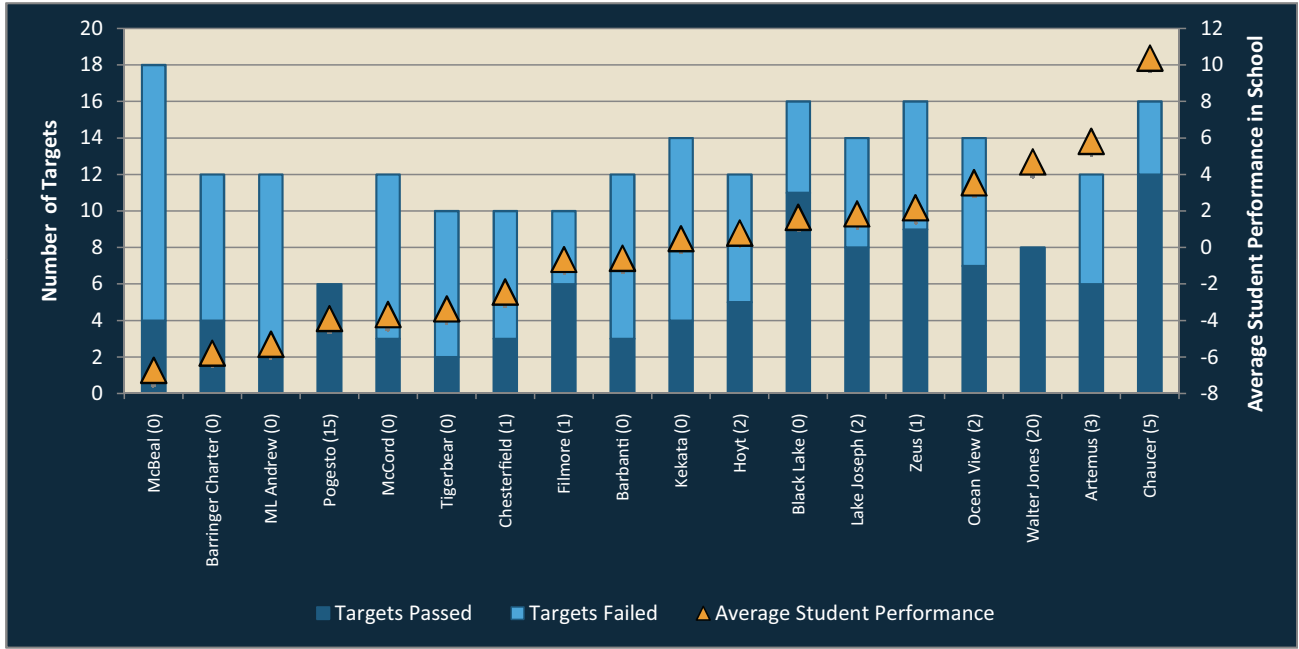


Figure 4. AYP performance of the middle school sample under Minnesota's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under Minnesota'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. Filmore, for example, met 6 of its 10 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.

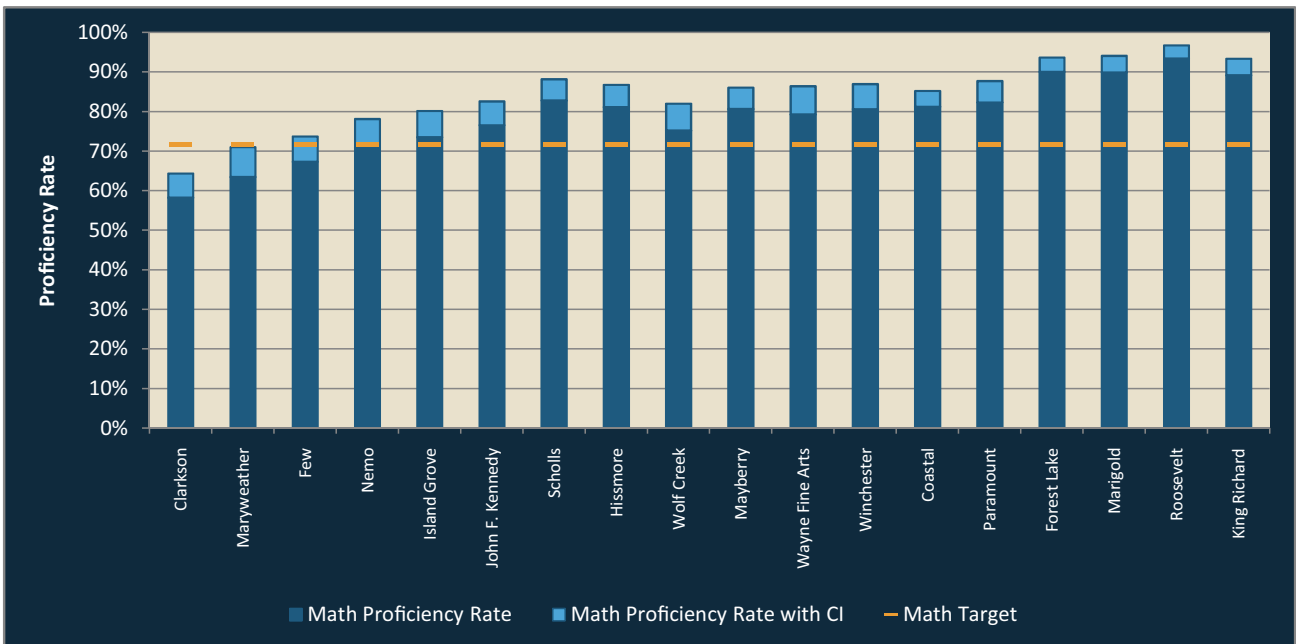
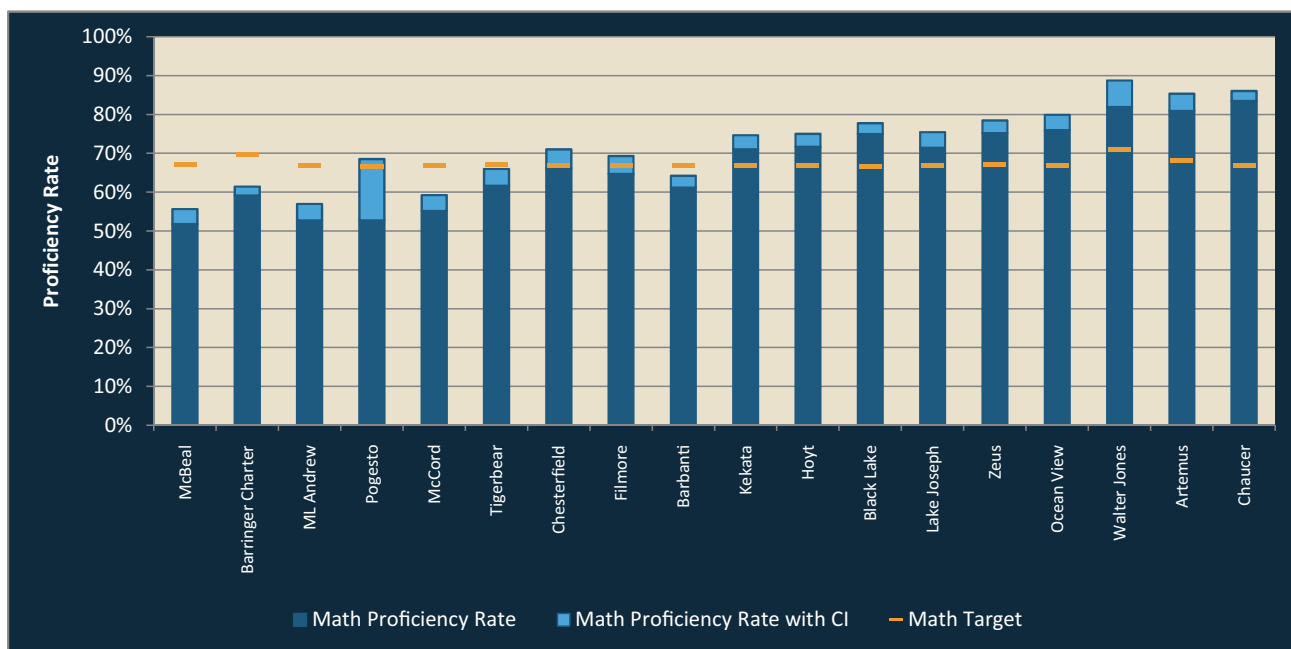


Figure 5. Impact of the confidence interval on elementary school math proficiency rates

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 two of the sample elementary schools (Few and Nemo) 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.



**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 two of the sample middle schools (Pogesto and Filmore) 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.

tistical error) to their measurements of student proficiency rates, **Minnesota’s sliding confidence interval range of 95%–99% is unusual. The confidence intervals make it easier for Minnesota schools to achieve their targets, with schools that have many subgroups receiving more of a “boost” than schools with fewer targets.**<sup>8</sup>

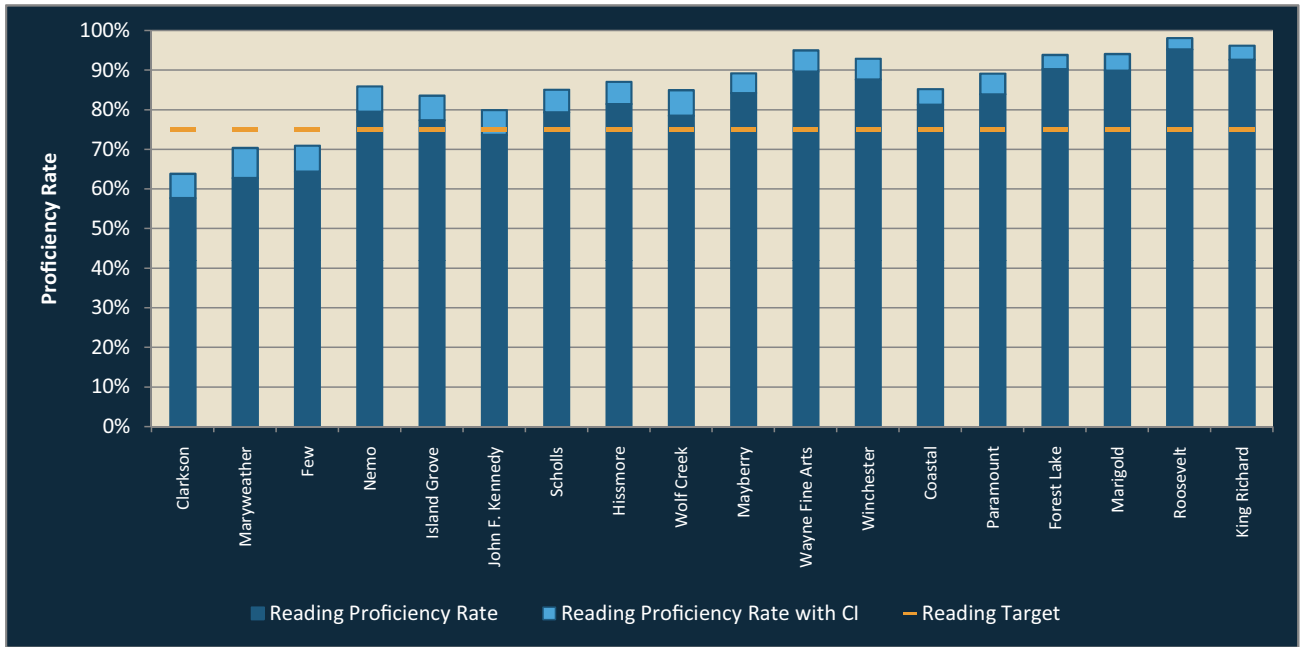
Finally, while most states measure school performance by a proficiency rate (or percentage of students achieving a proficient or higher level of performance on the state test), **Minnesota employs a performance index that gives partial credit to students attaining partial proficiency. In the short term, the index makes it easier for schools to meet their targets, although this benefit decreases as the targets approach 100%.**<sup>9</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. 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

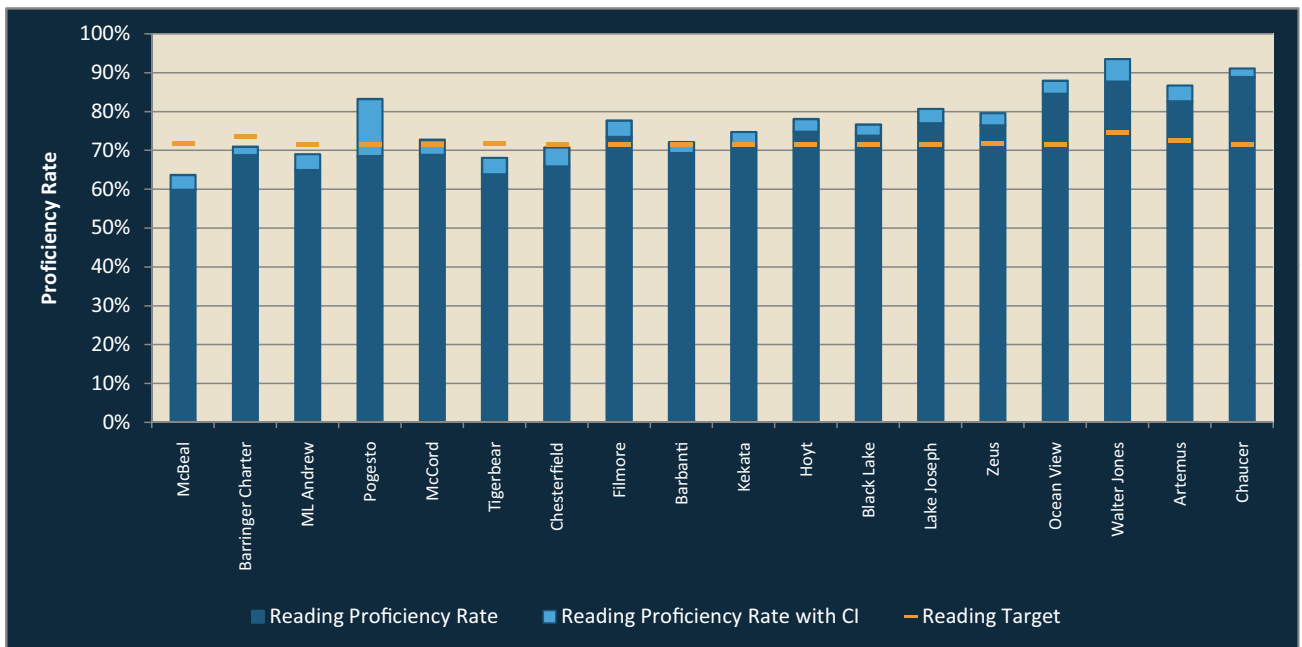
<sup>8</sup> We also conducted an analysis to show the effect of confidence intervals on the reading and math proficiency rates for elementary and middle schools. We describe those results later in the report.

<sup>9</sup> Minnesota is one of six states (Massachusetts, New Hampshire, Rhode Island, Vermont, and Wisconsin are the others) in our 28-state sample to use an index 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.



**Figure 7.** Impact of the confidence interval on elementary school reading 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 8.** Impact of the confidence interval on middle school reading 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 middle schools (Pogesto) 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.



Table 2. Elementary subgroup performance of sample schools under the 2008 Minnesota 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	58.3%	57.7%	N	N			N	N	N	N					N	N					8	0	0%	N	1
Maryweather	63.5%	62.8%	N	N			N	N	N	N	Y	Y			N	N			Y	Y	12	4	33%	N	1
Few	67.4%	64.4%	Y	N	N	N	N	N	Y	N	Y	Y			Y	N	Y	Y	Y	Y	16	9	56%	N	1
Nemo	70.9%	79.5%	Y	Y					N	Y	N	Y			Y	Y			Y	Y	10	8	80%	N	7
Island Grove	73.5%	77.4%	Y	Y					Y	Y					Y	N			Y	Y	8	7	88%	N	4
JFK	76.6%	73.7%	Y	Y	Y	N			Y	Y	Y	N							Y	Y	10	8	80%	N	3
Scholls	82.8%	79.4%	Y	Y	Y	N			Y	Y	Y	Y			Y	Y			Y	Y	12	11	92%	N	7
Hissmore	81.2%	81.5%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Wolf Creek	75.2%	78.5%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	5
Alice Mayberry	80.7%	84.3%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	9
Wayne Fine Arts	79.3%	89.7%	Y	Y					Y	Y	Y	Y			Y	Y			Y	Y	10	10	100%	Y	21
Winchester	80.7%	87.7%	Y	Y					Y	Y				Y	Y	Y			Y	Y	9	9	100%	Y	22
Coastal	81.3%	81.3%	Y	Y	N	N	N	N	Y	Y	Y	Y			Y	Y			Y	Y	14	10	71%	N	3
Paramount	82.3%	84.0%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	7
Forest Lake	90.1%	90.3%	Y	Y	Y	Y			Y	Y	Y	Y							Y	Y	10	10	100%	Y	8
Marigold	89.9%	89.9%	Y	Y	Y	Y			Y	Y			Y	Y	Y	Y			Y	Y	12	12	100%	Y	10
Roosevelt	93.4%	95.3%	Y	Y					Y	Y	Y	Y			Y	Y			Y	Y	10	10	100%	Y	28
King Richard	89.2%	92.7%	Y	Y	Y	Y			Y	Y					Y	Y			Y	Y	10	10	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.

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 Minnesota's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Minnesota's 2008 AYP rules. **Nine schools 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. Nearly all of the passing schools are in the right half of the figure, meaning that the highest performing students were found at these schools.

Table 3. Middle school subgroup performance of sample schools under the 2008 Minnesota 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	51.8%	59.8%	N	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N	N	Y	Y	18	4	22%	N	0
Barringer Charter	59.2%	68.8%	N	N	N	N			N	N	N	N			Y	Y			Y	Y	12	4	33%	N	0
ML Andrew	52.7%	64.9%	N	N	N	N			N	N	N	N			N	N			Y	Y	12	2	17%	N	0
Pogesto	52.8%	68.5%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	15
McCord Charter	55.2%	68.9%	N	Y	N	N			N	N	N	N			N	N			Y	Y	12	3	25%	N	0
Tigerbear	61.7%	63.8%	N	N	N	N			N	N	N	N							Y	Y	10	2	20%	N	0
Chesterfield	66.3%	65.9%	Y	N	N	N			N	N	N	N							Y	Y	10	3	30%	N	1
Filmore	64.7%	73.4%	Y	Y	N	N			Y	Y					N	N			Y	Y	10	6	60%	N	1
Barbanti	61.3%	69.3%	N	Y	N	N	N	N	N	N					N	N			Y	Y	12	3	25%	N	0
Kekata	71.1%	71.1%	Y	Y	N	N	N	N	N	N	N	N			N	N			Y	Y	14	4	29%	N	0
Hoyt	71.7%	74.8%	Y	Y	N	N			N	N	N	N			Y	N			Y	Y	12	5	42%	N	2
Black Lake	75.0%	73.8%	Y	Y	N	N			Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	16	11	69%	N	0
Lake Joseph	71.5%	77.0%	Y	Y	N	N	N	N	Y	Y	Y	Y			N	N			Y	Y	14	8	57%	N	2
Zeus	75.2%	76.4%	Y	Y	N	N	N	N	Y	N	Y	Y	Y	Y	N	N			Y	Y	16	9	56%	N	1
Ocean View	76.0%	84.6%	Y	Y	N	Y	N	N	N	N			Y	Y	N	N			Y	Y	14	7	50%	N	2
Walter Jones	82.0%	87.7%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	20
Artemus	81.0%	82.6%	Y	Y	N	N			N	N			Y	Y	N	N			Y	Y	12	6	50%	N	3
Chaucer	83.5%	88.9%	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	16	12	75%	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.

Yet almost without regard to average student performance, the schools that made AYP were those with relatively few qualifying subgroups—and thus the fewest targets to meet. For example, three out of five schools with the fewest (eight) targets made AYP.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Minnesota AYP rules. Of 18 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' 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 the application of the confidence interval. The orange lines show the annual measurable objective needed to meet AYP. These figures show that two elementary schools (Few and Nemo) and two middle schools (Pogesto and Filmore) were assisted by the peach confidence intervals, though all of these except Pogesto still failed to make AYP because of

**Table 4.** Summary of subgroup performance of sample elementary schools under the 2008 Minnesota 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	3	5
Students with limited English proficiency	4	4	4
Low-income students	18	3	3
African-American students	11	1	1
Asian/Pacific Islander students	2	0	0
Hispanic students	14	2	4
American Indian/Alaska Native students	1	0	0
White students	17	0	0

**Table 5.** Summary of subgroup performance of sample middle schools under the 2008 Minnesota 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	15
Students with limited English proficiency	7	7	7
Low-income students	18	11	13
African-American students	12	9	9
Asian/Pacific Islander students	6	0	0
Hispanic students	15	10	11
American Indian/Alaska Native students	2	1	1
White students	18	0	0

poor subgroup performance (see Figures 3 and 4).

Figures 7 and 8 show the effect of confidence intervals on the reading proficiency rates for elementary and middle schools, respectively. Only one elementary school (JFK) and one middle school (Pogesto) met the overall

reading target with the assistance of the confidence interval, but JFK failed to meet all its subgroup targets (see Figure 3). **Overall, the application of the confidence interval provides moderate assistance in helping Minnesota schools achieve their overall math and reading targets.**<sup>10</sup>

<sup>10</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 through 8. However, we chose not to show how the confidence interval impacted subgroup performance because it would have added greatly to this report's length and complexity.

## Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low or mid-level performance can still make 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 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 Minnesota 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:

- Only two elementary schools (Clarkson and Maryweather) failed to meet both math and reading targets for their overall school populations, and one additional school (Few) failed to meet its reading target for its overall population.
- Six middle schools failed to achieve their overall math targets and five missed their overall reading targets.
- Two (Scholls and Hissmore) of the nine failing elementary schools missed AYP only because of the SWD subgroup.
- One elementary school (Island Grove) passed in every subgroup except for Hispanic students.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively. First, the performance of SWDs is proving to

be very challenging for schools under Minnesota's system, particularly in middle schools, where this subgroup tends to have enough students to meet the state's minimum  $n$  of 40. In fact, nearly all middle schools in the study with qualifying SWD subgroups failed to make AYP. Students with LEP are also struggling 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.

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

A close look at Figures 3 and 4 indicates that Minnesota's NCLB accountability system is, in some respects, behaving like 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 Minnesota, too. Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make in Minnesota.

But Minnesota is home to a few anomalies. First, consider Wolf Creek Elementary (see Figure 3). It failed to make AYP in 23 of the 28 states in our sample, yet made AYP in Minnesota. In examining Table 2, we can see that Wolf Creek did not meet the minimum numbers for the LEP or SWD subgroups, which create difficulty for so many other schools within the sample. With fewer accountable subgroups, Wolf Creek made AYP, even when other schools with higher average performance (like Coastal) failed. Second, look at Pogesto Middle School (Figure 4). Even with its relatively low average performance it made AYP in Minnesota, but not in 13 out of 28 states. Like Wolf Creek, its AYP success in Minnesota is most likely attributable to the relatively small number of targets (six) it has to meet.

This is consistent with the patterns shown in Table 6, which compares schools that did and didn't make AYP on a number of academic and demographic dimensions. Within the sample, schools that made AYP did indeed show higher average student performance, but they also

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	9	9	2	16
Average student body size	275	335	124	951
Average % low income	26	67	42	45
Average % nonwhite	30	52	27	46
Average performance†	5.21	-2.76	0.40	-0.11
Average % growth‡	124	106	109	97
Average number of targets to meet	9	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.

differed in the following ways: they had smaller student populations, fewer subgroups (and thus fewer targets to meet), and much lower percentages of low income and nonwhite 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 Minnesota's AYP rules (and AMOs) for 2008. We found that 9 elementary schools and 2 middle schools—11 in all from a sample of 36—would have made AYP in Minnesota. Looking across the 28 state accountability systems examined in the study, this puts Minnesota at the high end of the sample distribution in terms of the number of schools making AYP (as shown in Figure 1). In addition, **Minnesota's minimum subgroup size varies by particular subgroup, meaning that schools in Minnesota may have more accountable subgroups than similar schools in other states. The application of the confidence interval in Minnesota also provides moderate assistance in helping Minnesota schools achieve their overall math and reading targets.**

Because the overriding goal of NCLB is to eliminate educational disparities within and across states, it's important to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, Minnesota'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 low-income or Hispanic students. The majority of the sample schools met the Minnesota math and reading targets for their student populations as a whole, i.e., without considering subgroup performance. 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 weren'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 the student population has so much influence over making AYP? Does it make sense that having fewer subgroups enhances the likelihood of

making AYP? Doesn't the failure of English language learners, SWDs, low-income students, and other minority groups to meet Minnesota's targets (especially at the middle school level) 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 various subgroups of students, as for other students, but when half or more of schools is not 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 re-authorization 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.