North Dakota

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 towards these goals.

This report examines North Dakota'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 North Dakota'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 North Dakota'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 North Dakota's AYP rules for academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- We estimate that **16 of 18 elementary schools** and **all of the 18 middle schools** in our sample **failed to make adequate yearly progress** in 2008 under North Dakota'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.)
- Looking across the 28 state accountability systems examined in the study, we find that the number of schools making AYP in North Dakota is exceeded in 20 other sample states (five states tie with North Dakota, each with two elementary schools making AYP). In addition, North Dakota is one of five states with *zero* passing middle schools in the sample (see Figure 1).
- Many of the schools in our sample that failed to make AYP in North Dakota are meeting expected targets for their overall populations but failing because of the performance of individual subgroups, particularly students with disabilities and English language learners.[2]
- **Two sample schools failed to make AYP in North Dakota that made AYP in most other states. This is likely due to the fact that North Dakota's minimum subgroup size of 10 is small, compared to other states in the study.[3] In addition, North Dakota's annual targets for proficiency are relatively ambitious.**

Only two of the 36 schools in our sample make AYP in 2008 under **North Dakota's** accountability system. The greatest contributing factor to the high failure rate is that North Dakota's minimum subgroup size is 10, which is considerably smaller than most other states we examined. This means that schools in North Dakota will have more accountable subgroups than would similar schools in other states. On the other hand, North Dakota's proficiency standards are about average when compared to the other states in the study. The state also uses a 99 percent confidence interval which provides schools with greater leniency than the more commonly used 95 percent confidence interval. The latter likely explains why two sample schools were able to make AYP in North Dakota.

[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 North Dakota State Assessment (NDSA).

[2] 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

[3] The state of North Dakota does not have a minimum school size, so it has a large number of very small schools. In addition, the state's population has been declining in recent years. The U.S. Census Bureau (2002) lists North Dakota's population at a little over 642,000, 47th in the United States. Therefore, smaller subgroup sizes are likely warranted.

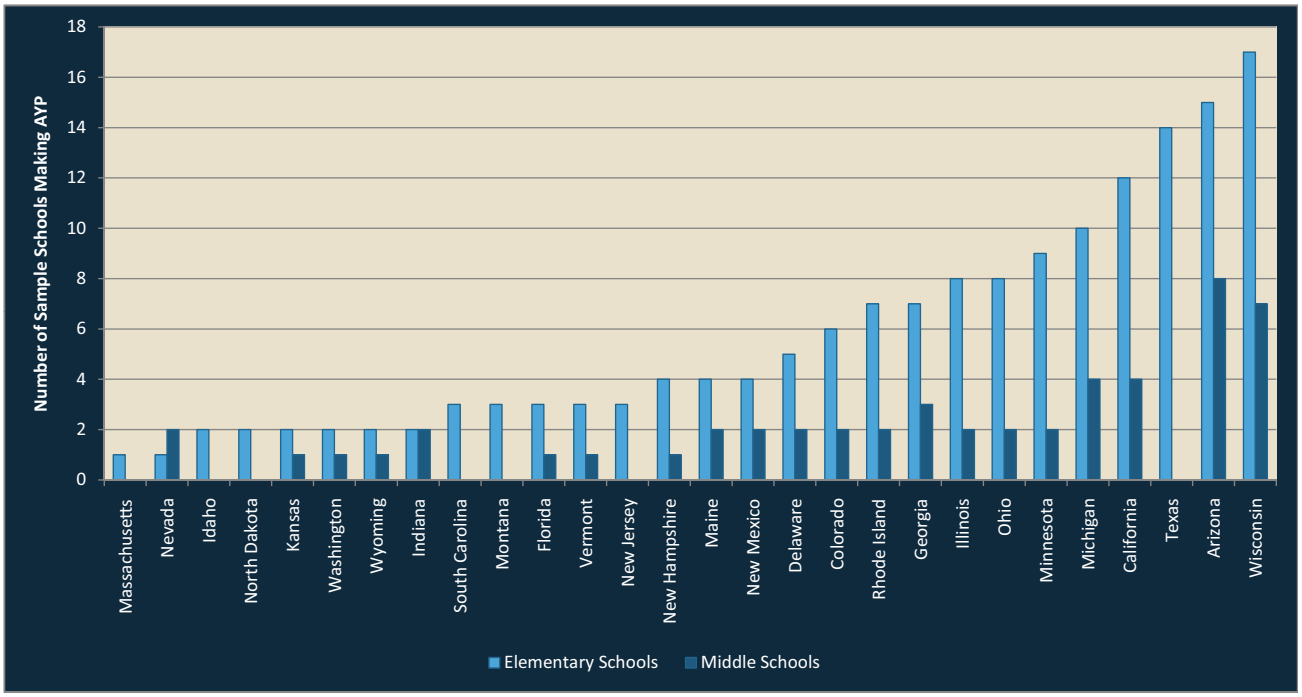

Figure 1. Number of sample schools making AYP by state

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.

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

- Part of the reason all middle schools failed to make AYP in North Dakota is that its schools have enough low-income, disabled, or limited English proficiency (LEP)[4] students to qualify as separate subgroups. Each of our sample middle schools in North Dakota has one or more of these subgroups and each failed to make AYP. Likewise, many elementary schools with enough students qualifying for these subgroups also failed, though they tended to reach their math targets more often than their reading targets.[5]

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance North Dakota's tests and those of 25 other states to the Northwest Evaluation Association's Measures of Academic Progress (MAP), a computerized adaptive test used in schools nationwide. This single common scale permitted cross-state comparisons of each state's reading and math proficiency standards to measure school performance under the No Child Left Behind (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 stan-

[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 North Dakota State Assessment (NDSA), 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. Our 2005–2006 MAP data do not capture these subgroup nuances. In this study, however, no valid MAP scores were omitted from consideration.

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

What We Studied

We collected students' MAP test scores from the 2005-06 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, such 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]

Proficiency cut score estimates for the North Dakota State Assessment (NDSA) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that North Dakota's definitions of proficiency 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 North Dakota 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 North Dakota AYP rules for 2008. In other words, the school data and our proficiency cut score es-

[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.

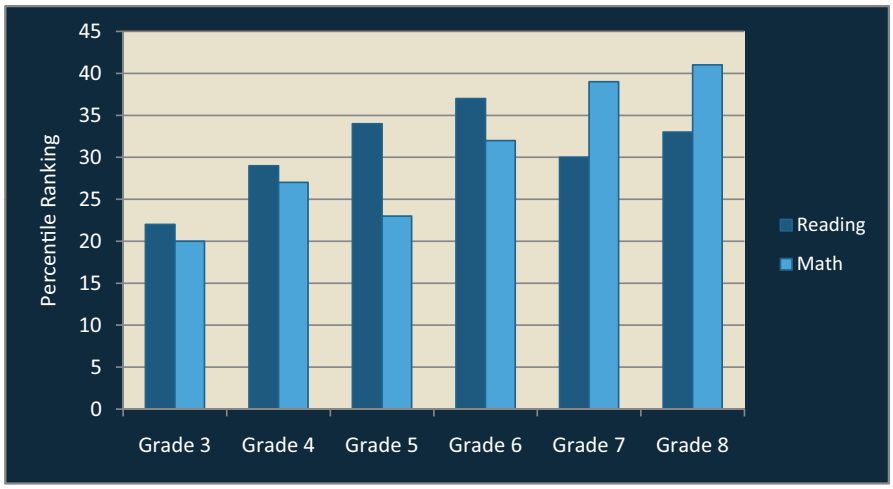

Figure 2. North Dakota reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of North Dakota'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 North Dakota's cut scores are below the 45th percentile.

timates are from academic year 2005–2006, but we are applying them against North Dakota's 2008 AYP rules.

Table 1 shows the pertinent North Dakota AYP rules that were applied to elementary and middle schools in this study. North Dakota's minimum subgroup size is 10, which is considerably smaller than most other states we examined.[8] This means that schools in North Dakota will have more accountable subgroups than would similar schools in other states. North Dakota's annual targets also differ by grade and subject. For example, 66.7% of grade 8 math students are expected to be proficient in 2008; the percentage for grade 3 reading students is 82.6%.

Most states examined also apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates. However, North Dakota's 99% confidence interval provides schools with greater leniency than the more commonly used 95% confidence interval. So, for instance, while schools are supposed to get 82.6% of their students in grade 3 to the "proficient" level on the state reading test, and 82.6% of the students in each subgroup, applying the confidence interval means that the real target can be lower.

Note that we were unable to examine the effect 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 is 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 school's 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

[8] The state of North Dakota does not have a minimum school size, so it has a large number of very small schools. In addition, the state's population has been declining in recent years. The U.S. Census Bureau (2002) lists North Dakota's population at a little over 642,000, 47th in the United States. Therefore, smaller subgroup sizes are likely warranted.

Table 1. North Dakota AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 10	
	SWDs: 10	
	Low-income students: 10	
	LEP students: 10	
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	82.6
Grade 4	65.1	82.6
Grade 5	n/a	82.6
Grade 6	n/a	80.7
Grade 7	n/a	80.7
Grade 8	61.4	80.7
MATH		
Grade 3	n/a	72.9
Grade 4	45.7	72.9
Grade 5	n/a	72.9
Grade 6	n/a	66.7
Grade 7	n/a	66.7
Grade 8	33.3	66.7

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

Abbreviations: SWDs = students with disabilities; LEP = limited English proficiency; CI = confidence interval; AMOs = annual measurable objectives; 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 North Dakota's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under North Dakota's 2008 AYP rules. **Only 2 elementary schools (Wayne Fine Arts and Roosevelt) made AYP while 16 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. The two schools making AYP are in the right half of the figure, meaning that they are among the schools that contain the higher average performing students. Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 North Dakota AYP rules. Not a single middle school in the sample makes AYP under the North Dakota rules.

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 dark blue bars show the actual proficiency rates at each school, and the light blue bars show the degree to

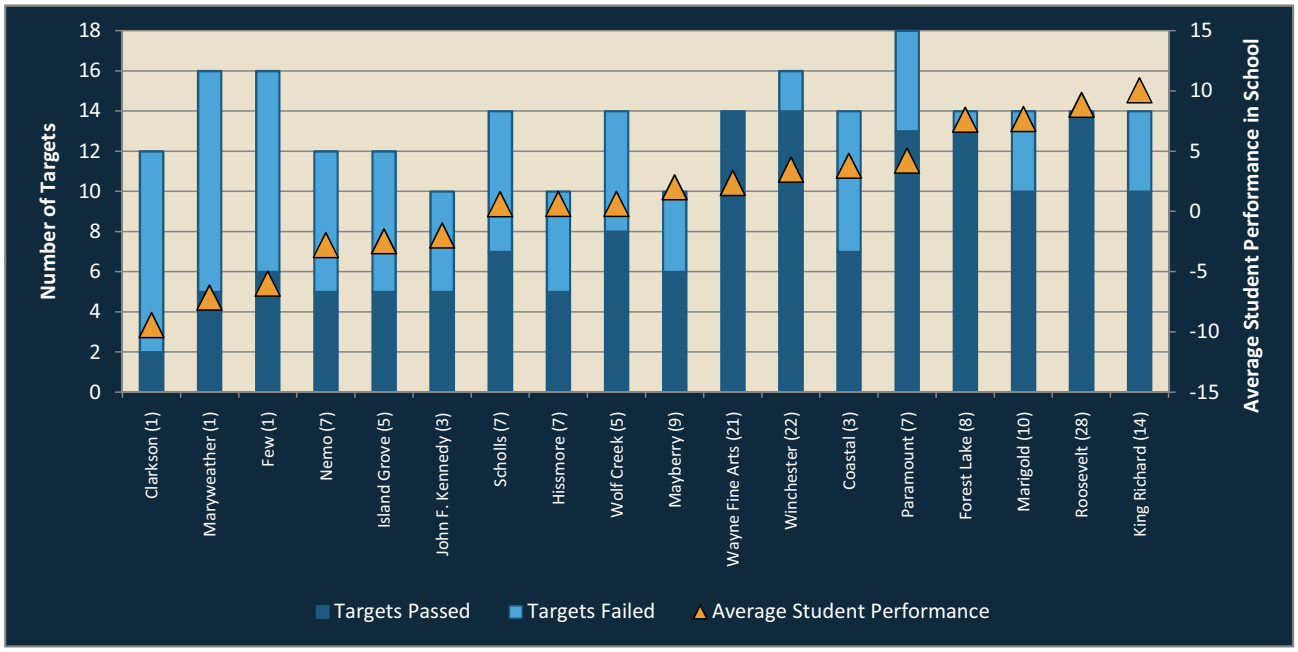

Figure 3. AYP performance of the elementary school sample under North Dakota 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under North Dakota'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 (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, met ten 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 makes AYP in the study.

which these proficiency rates are increased by the application of the confidence interval. The orange lines show the annual measurable objective (or annual target) needed to meet AYP. The figures show that only one of the sample elementary schools (Maryweather) and three of the middle schools (Tigerbear, Chesterfield, and Filmore) were assisted by the confidence intervals (note how the orange lines fall within the light blue bands). However, we know that all of these schools still failed to make AYP because of low subgroup performance (see Figures 3 and 4). Tigerbear, for instance, didn't meet nine of its twelve targets.

The effect of confidence intervals on the reading proficiency rates for elementary and middle schools shows largely the same pattern (not shown). In reading, two elementary schools (Mayberry and Paramount) and one middle school (Pogesto) met the overall target with the confidence interval, but we know from Figures 3 and 4 that these schools still fail to meet targets for subgroups. In short, **the application of the confidence interval had little or no impact on whether schools achieved their overall math and reading targets in North Dakota (or whether they made AYP).**[9]

Where Do Schools fail?

Figures 3 and 4 illustrate the number of subgroup targets at each of the sample schools, but these figures do not 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.

[9] 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 complexity and length.

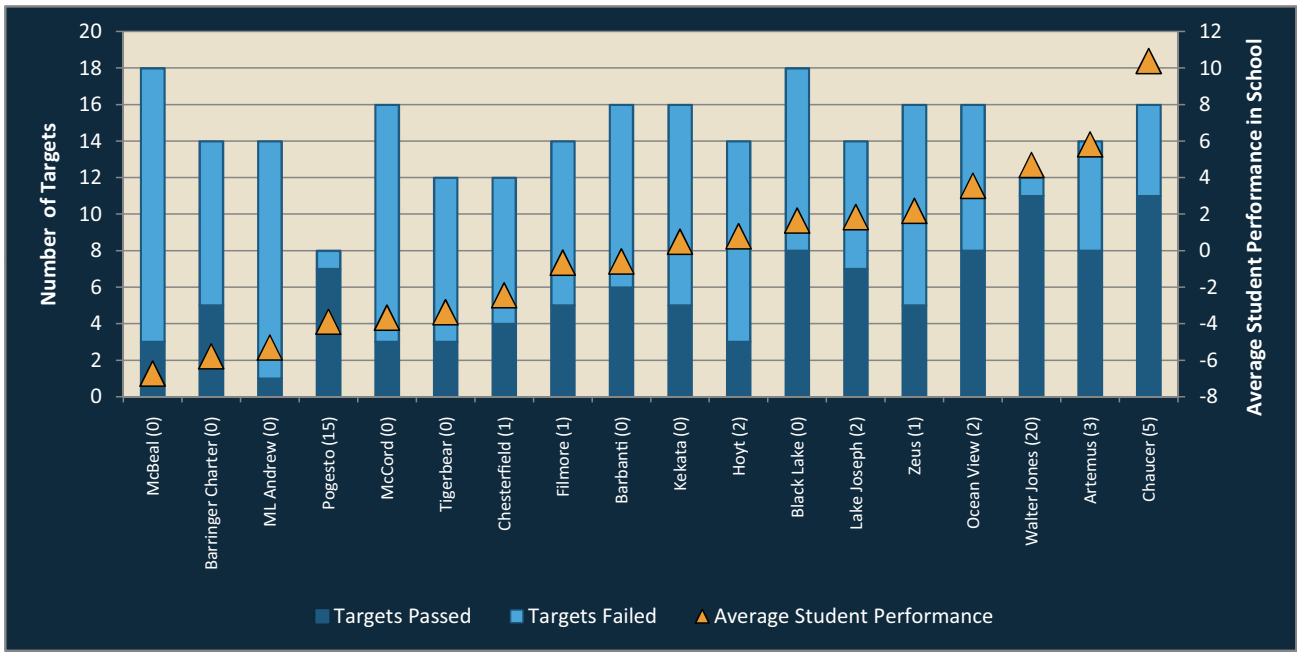

Figure 4. AYP performance of the middle school sample under North Dakota 2008 AYP rules

Note: This figure indicates how each of the middle schools within the sample fared under North Dakota'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 (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. Pogesto, for example, met seven 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 make AYP in the study.

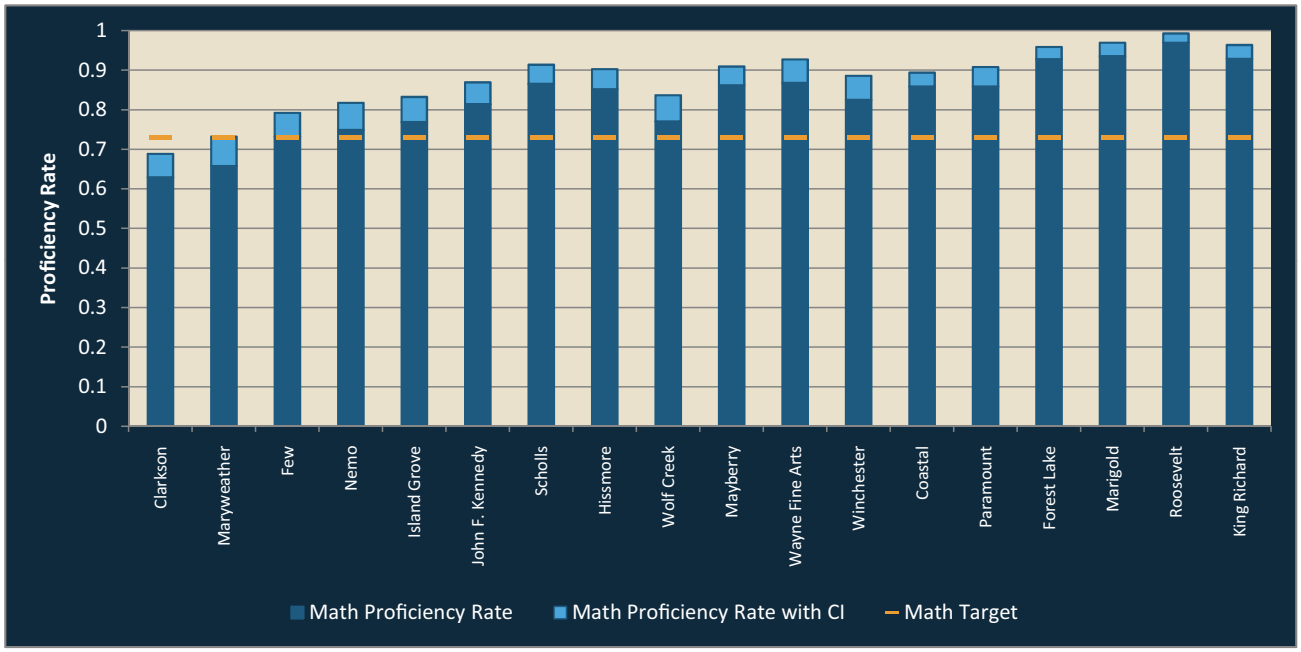

Figure 5. Impact of the confidence Interval on elementary school math proficiency rates under North Dakota 2008 AYP rules

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.

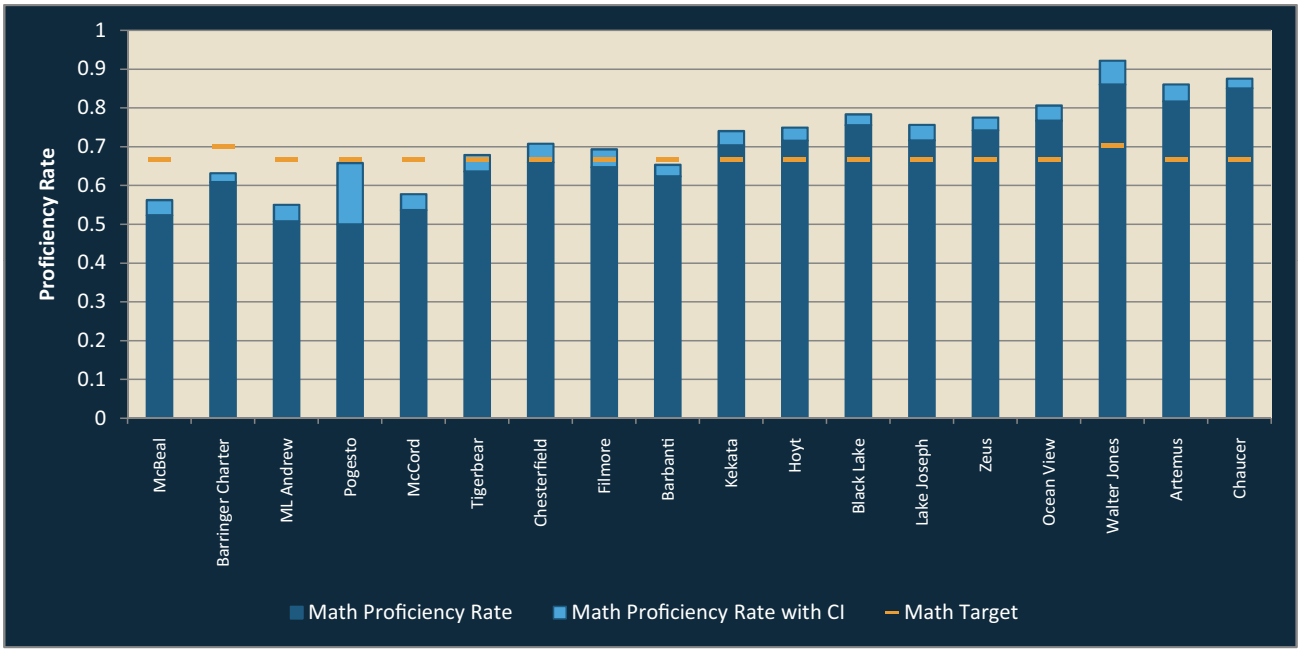

Figure 6. Impact of the confidence interval on middle school math proficiency rates under North Dakota 2008 AYP rules

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 of the sample middle schools (Tigerbear, Chesterfield, 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.

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. While all schools are evaluated on the proficiency rate of their overall population, potential subgroups that are separately evaluated for AYP purposes include SWDs, students with LEP, low-income students, and the following race/ethnic categories: African American (AA), Asian/Pacific Islander (Asian), Hispanic/Latino (Hispanic), American Indian/Alaska Native (AI/AN), and White. Tables 2 and 3 also show whether a school made AYP under the North Dakota rules, and the total number of states within the study in which that school makes AYP.

The school-by-school findings in Tables 2 and 3 are summarized as shown:

- One of the 18 elementary schools (Clarkson) failed to meet both reading and math targets for its overall student population, while nine others (Maryweather, Few, Nemo, Island Grove, JFK, Scholls, Hissmore, Wolf Creek, and Coastal) failed to meet their overall reading targets.
- Twelve middle schools (McBeal, Barringer, ML Andrew, McCord, Tigerbear, Chesterfield, Filmore, Barbanti, Kekata, Hoyt, Black Lake, and Zeus) failed to meet their overall reading targets, and six (McBeal, Barringer, ML Andrew, Pogesto, McCord, and Barbanti) failed in math.
- One elementary school (Forest Lake) met every target except for its reading target for students with disabilities.
- One middle school (Walter Jones) met all targets for every subgroup except for its African American population.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively.[10] The performance of SWDs is proving challenging

[10] Recall that elementary students generally do better on North Dakota's math test than middle school students, partly because North Dakota's cut scores are lower in math than in reading at the elementary level (see Figure 2).

Table 2. Elementary subgroup performance of sample schools under the 2008 North Dakota 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	62.9%	43.1%	N	N	N	N	N	N	N	N					N	N			Y	Y	12	2	17%	N	1
Maryweather	65.8%	52.5%	Y	N	N	N	N	N	N	N	Y	N			N	N	Y	N	Y	Y	16	5	31%	N	1
Few	73.2%	55.3%	Y	N	N	N	N	N	Y	N	Y	N			Y	N	Y	N	Y	N	16	6	38%	N	1
Nemo	74.9%	71.6%	Y	N	N	N			Y	N	N	N			Y	N			Y	Y	12	5	42%	N	7
Island Grove	76.9%	68.3%	Y	N	N	N	N	N	Y	N					Y	N			Y	Y	12	5	42%	N	4
JFK	81.4%	66.0%	Y	N	Y	N			Y	N	Y	N							Y	N	10	5	50%	N	3
Scholls	86.6%	71.7%	Y	N	N	N	Y	N	Y	N	Y	N			Y	N			Y	Y	14	7	50%	N	7
Hissmore	85.2%	74.8%	Y	N	N	N			Y	N	Y	N							Y	Y	10	5	50%	N	7
Wolf Creek	77.1%	73.1%	Y	N	N	N	Y	N	Y	N			Y	Y	Y	N			Y	Y	14	8	57%	N	5
Alice Mayberry	86.2%	78.9%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	9
Wayne Fine Arts	86.8%	83.3%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	14	100%	Y	21
Winchester	82.5%	82.0%	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	16	14	88%	N	22
Coastal	85.9%	75.2%	Y	N	N	N	Y	N	Y	N	Y	N			Y	N			Y	Y	14	7	50%	N	3
Paramount	85.9%	78.7%	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	18	13	72%	N	7
Forest Lake	92.8%	86.6%	Y	Y	Y	N			Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	14	13	93%	N	8
Marigold	93.5%	84.8%	Y	Y	Y	N	Y	N	Y	N			Y	Y	Y	N			Y	Y	14	10	71%	N	10
Roosevelt	97.0%	92.9%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	14	100%	Y	28
King Richard	92.9%	87.8%	Y	Y	Y	N	Y	N	Y	N			Y	Y	Y	N			Y	Y	14	10	71%	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.

for schools under North Dakota's system where this subgroup tends to have enough students to meet the state's minimum *n* of 10. In fact, all but two elementary schools and all but one middle school in the study with qualifying SWDs reading subgroups failed to make AYP. Students with LEP and low-income students are also struggling to meet the state's targets; almost every single school with a large enough population to qualify as a separate subgroup failed to meet its targets for these students (though they tend to do better in math at the elementary level).

Other state reports contain a section comparing some of the characteristics of the sample schools that made AYP versus those that did not. In North Dakota, none of the sample middle schools made AYP, and among elementary schools, there were no striking differences among schools that did and didn't make AYP. The one exception (rather expected) was that schools that made AYP had students with higher average performance than did schools that didn't make it, as measured by NWEA reading and math tests.

Table 3. Middle school subgroup performance of sample schools under the 2008 North Dakota 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	52.4%	54.2%	N	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N	N	Y	N	18	3	17%	N	0
Barringer Charter	60.9%	60.2%	N	N	N	N	Y	Y	N	N	N	N			Y	N			Y	Y	14	5	36%	N	0
ML Andrew	50.8%	59.8%	N	N	N	N	N	N	N	N	N	N			N	N			Y	N	14	1	7%	N	0
Pogesto	50.0%	68.5%	N	Y					Y	Y					Y	Y			Y	Y	8	7	88%	N	15
McCord Charter	53.7%	61.4%	N	N	N	N	N	N	N	N	N	N	Y	Y	N	N			Y	N	16	3	19%	N	0
Tigerbear	63.6%	59.6%	Y	N	N	N			N	N	N	N			Y	N			Y	N	12	3	25%	N	0
Chesterfield	66.0%	60.9%	Y	N	N	N			N	N	N	N			Y	Y			Y	N	12	4	33%	N	1
Filmore	64.7%	69.9%	Y	N	N	N	N	N	N	N			Y	Y	N	N			Y	Y	14	5	36%	N	1
Barbanti	62.4%	64.8%	N	N	N	N	N	N	N	N	Y	Y	Y	Y	N	N			Y	Y	16	6	38%	N	0
Kekata	70.4%	68.5%	Y	N	N	N	N	N	N	N	N	N	Y	Y	N	N			Y	Y	16	5	31%	N	0
Hoyt	71.6%	71.7%	Y	N	N	N	N	N	N	N	N	N			N	N			Y	Y	14	3	21%	N	2
Black Lake	75.6%	71.2%	Y	N	N	N	N	N	Y	N	N	N	Y	Y	Y	N	Y	Y	Y	N	18	8	44%	N	0
Lake Joseph	71.7%	77.2%	Y	Y	N	N	N	N	Y	N	Y	Y			N	N			Y	Y	14	7	50%	N	2
Zeus	74.2%	72.7%	Y	N	N	N	N	N	N	N	Y	N	Y	Y	N	N			Y	N	16	5	31%	N	1
Ocean View	76.7%	83.2%	Y	Y	N	N	N	N	N	N	Y	Y	Y	Y	N	N			Y	Y	16	8	50%	N	2
Walter Jones	86.0%	83.4%	Y	Y	Y	Y			Y	Y	Y	N			Y	Y			Y	Y	12	11	92%	N	20
Artemus	81.7%	80.5%	Y	Y	N	N			N	N	Y	Y	Y	Y	N	N			Y	Y	14	8	57%	N	3
Chaucer	85.1%	86.9%	Y	Y	N	N	N	N	Y	N	Y	Y	Y	Y	Y	Y			Y	Y	16	11	69%	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.

Characteristics of Schools that Did and Didn't Make AYP

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

But North Dakota is also home to a few anomalies. First, consider Winchester Elementary (see Figure 3). It made AYP in 22 of the 28 states in our sample, yet not in North Dakota. Examining Table 2, one can see that Winchester didn't meet reading targets for its LEP or SWD subgroups, although the school's overall reading proficiency rate was 82%. Second, look at Walter Jones Middle School (Figure 4). Even with its relatively high average performance it didn't make AYP in North Dakota, but made AYP in 20 of 28 states. Like Winchester, it missed the AYP mark in North Dakota probably because of North Dakota's relatively small minimum *n*.

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 North Dakota 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	18	10	16
Students with limited English proficiency	13	5	11
Low-income students	18	2	14
African-American students	13	1	8
Asian/Pacific Islander students	6	0	0
Hispanic students	15	2	11
American Indian/Alaska Native students	3	0	2
White students	18	0	2

Table 5. Summary of subgroup performance of sample middle schools under the 2008 North Dakota 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	17	16	16
Students with limited English proficiency	13	12	12
Low-income students	18	13	16
African-American students	16	9	11
Asian/Pacific Islander students	10	0	0
Hispanic students	18	11	14
American Indian/Alaska Native students	2	1	1
White students	18	0	7

Note: The relatively high number of qualifying subgroups for African American, Asian/Pacific Islander, and Hispanic students is largely due to Nebraska's minimum n size of 10.

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 North Dakota's AYP rules (and AMOs) for 2008. Among this sample, only two elementary schools and no middle schools—two in all from a sample of 36—would have made AYP in North Dakota. Looking across the 28 state accountability systems examined in the study, this puts North Dakota at the low end of the distribution in terms of the number of schools making AYP (see Figure 1). North Dakota's small minimum *n* size and fairly high AMOs likely lead to the large number of schools that failed to make AYP.

The overriding goal of the federal 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, North Dakota'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 students. Many of the sample schools met the North Dakota reading and math targets for their student populations as a whole (more so in math than reading). In the pre-NCLB era, such schools might have been considered to be effective or at least not needing 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 having fewer subgroups enhances the likelihood of making AYP? And in the case of North Dakota, that small subgroup sizes and high annual targets make it nearly impossible for schools to be viewed as successful? Even if actual participation guidelines for English language learners and SWDs are more generous under the current state assessment system,[11] doesn't the failure of these students to meet North Dakota'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 LEP students and SWDs, as for other students, but when so few schools are able to meet the goal, perhaps that indicates that the goal is unrealistic. These will be critical considerations for Congress as it takes up NCLB reauthorization in the future.

Limitations

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

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

[11] See footnote 5.

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.