

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 Wyoming'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 Wyoming'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 Wyoming's system as well as under the systems of 27 other states. We used school data and proficiency cut score¹ estimates from academic year 2005–2006, but applied them against Wyoming'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 **17 of 18 middle schools** in our sample **failed to make AYP** in 2008 under Wyoming'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 partially explained by Wyoming's proficiency standards which are relatively

¹ A cut score is the minimum score that a student must receive on the Proficiency Assessment for Wyoming Students (PAWS) in order to be considered proficient under Wyoming's accountability system.

² It's important to keep in mind, however, that school size impacts minimum subgroup size (e.g., it makes sense for smaller schools to have smaller *n* sizes).

³ It's important to note that students in subgroups not meeting the minimum *n* sizes are still included for accountability purposes in the overall student calculations; they simply are not treated as their own subgroup.

difficult, compared to other states. In addition, Wyoming's minimum subgroup size is 30, which is smaller than most other states we examined.² This means that more subgroups in Wyoming are held accountable for performance than in other states.

- **Looking across the 28 state accountability systems examined in the study, we find that the number of elementary schools making AYP in Wyoming was exceeded in 20 other sample states (Wyoming ties 5 other states with only 2 elementary schools making AYP). In addition, Wyoming was one of 6 states with a single passing middle school in the sample (see Figure 1).**
- Many of the schools in our sample that failed to make AYP in Wyoming are meeting expected targets for their overall populations but failing because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.³

Only two elementary schools and one middle school in our sample make AYP in 2008 under **Wyoming's** accountability system. This places Wyoming at the lower end of the state distribution in terms of the number of schools making AYP. This is likely due to the fact that Wyoming's proficiency standards are relatively difficult compared to other states. Almost all of Wyoming's cut scores are above the 40th percentile. Moreover, Wyoming's minimum subgroup size is 30, which is smaller than most other states we examined. This means that schools in Wyoming will have more accountable subgroups than would similar schools in other states. Finally, more subgroups in Wyoming's elementary schools failed to meet their reading targets than their math targets. This is probably because the proficiency standards in grades 3-6 reading are higher than those in math.

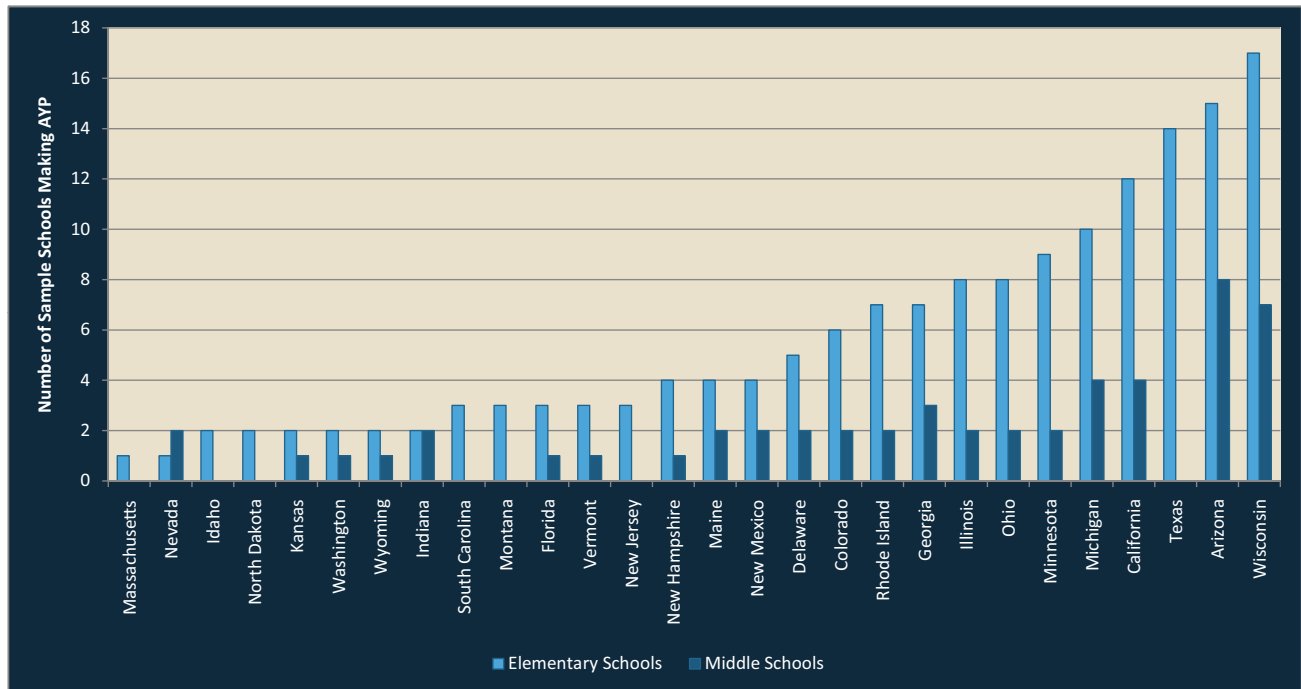


Figure 1. Number of sample schools making AYP by state

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.

- In Wyoming, 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 with greater diversity and size face greater challenges in making AYP.
- Like most other states, Wyoming applies a confidence interval (or statistical margin of error) to its measures of proficiency. **However, partly because of Wyoming’s relatively low annual targets in math and reading, the confidence interval has little or no effect on AYP decisions for the sample elementary and middle schools in the state.**
- As in other states, middle schools have greater difficulty reaching AYP in Wyoming 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 a school making AYP under Wyoming’s system is whether it has enough English language learners to qualify as a separate subgroup. Every school with a subgroup of students with limited English proficiency (LEP)⁴ failed to make AYP, in part because these students did not meet the state’s targets in reading and/or math. Likewise, all schools with enough qualifying SWDs failed to meet their AYP targets in reading.⁵

Introduction

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

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

⁵ Incidentally, reading cut scores in Wyoming are higher than math cut scores in grades 3-6. In addition, 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 Proficiency Assessment for Wyoming Students (PAWS), the standardized state test. Specifically, the U.S. Department of Education has issued new NCLB guidelines in recent years that exclude small percentages of LEP students and SWDs from taking the state test or that allow them to take alternative assessments. In this study, however, no valid MAP scores were omitted from consideration.

other states to the Northwest Evaluation Association's (NWEA's) Measures of Academic Progress (MAP), a computerized adaptive test used in schools nationwide. This single common scale permitted cross-state comparisons of each state's reading and math proficiency standards to measure school performance under the No Child Left Behind (NCLB) Act of 2001. That study revealed profound differences in states' proficiency standards (i.e., how difficult it is to achieve proficiency on the state test), and even across grades within a single state.

Our study expands on *The Proficiency Illusion* by examining other key factors of state NCLB accountability plans and how they interact with state proficiency standards to determine whether the schools in our sample made adequate yearly progress (AYP) in 2008. Specifically, we estimated how a single set of schools, drawn from around the country, would fare under the differing rules for determining AYP in 28 states (the original 25 in *The Proficiency Illusion* plus 3 others for which we now have cut score estimates). In other words, if we could somehow move these entire schools—with their same mix of characteristics—from state to state, how would they fare in terms of making AYP? Will schools with high-performing students consistently make AYP? Will schools with low-performing students consistently fail to make AYP? If AYP determinations for schools are not consistent across states, what leads to the inconsistencies?

NCLB requires every state, as a condition of receiving Title I funding, to implement an accountability system that aims to get 100% of its students to the proficient level on the state test by academic year 2013-2014. In the intervening years, states set annual measurable objectives (AMOs). This is the percentage of students in each school, and in each subgroup within the school (such as low income⁶ 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, 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, 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

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.

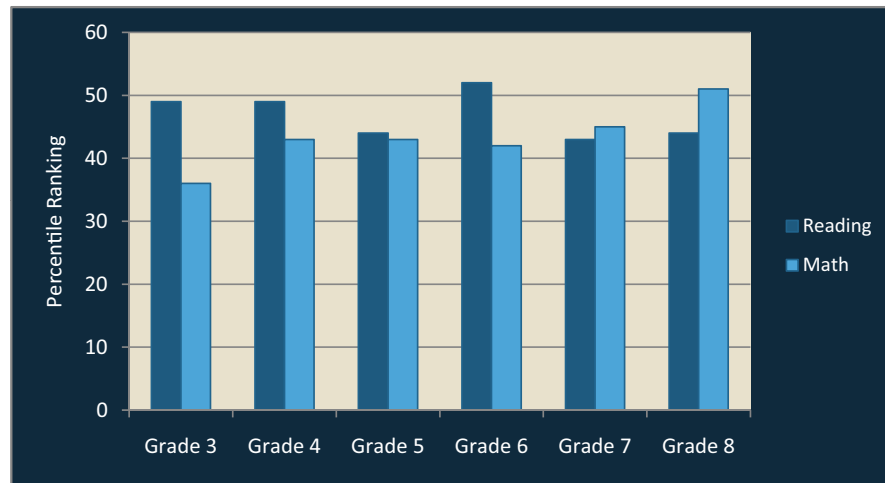


Figure 2. Wyoming reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Wyoming’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 Wyoming’s cut scores are at or below the 52nd percentile.

Proficiency cut score estimates for the Proficiency Assessment for Wyoming Students (PAWS) were estimated using the same methods as in *The Proficiency Illusion* (as shown in Figure 2), and were above average in difficulty, compared to 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 Wyoming test, given their performance on MAP. Student test data and subgroup designations are then used to determine how these 18 elementary and 18 middle schools would have fared under Wyoming 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 Wyoming’s 2008 AYP rules.

Table 1 shows the pertinent Wyoming AYP rules that were applied to elementary and middle schools in this study. Wyoming’s minimum subgroup size is 30, which is smaller than most other states we examined.⁸ As do most of the states examined in the current study, Wisconsin applies a 95% confidence interval – essentially a statistical margin of error—to their proficiency rate measurements. So, for instance, although schools are supposed to get 53.6% of their grade 3–5 students (and 53.6% of the grade 3–5 students in each subgroup) to

the proficient level on the state reading test, applying the confidence interval means that the real target can actually 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

⁸ Keep in mind, however, that school size and *n* size are related (e.g., small *n* sizes make sense for small schools).

Table 1. Wyoming AYP rules for 2008

Subgroup minimum <i>n</i>	Race/ethnicity: 30	
	SWDs: 30	
	Low-income students: 30	
	LEP students: 40	
CI	Applied to proficiency rate calculations?	
	Yes; 95% CI used	
AMOs	Baseline proficiency levels as of 2002 (%)	2008 targets (%)
READING/LANGUAGE ARTS		
Grade 3	n/a	53.6
Grade 4	30.4	53.6
Grade 5	n/a	53.6
Grade 6	n/a	56.3
Grade 7	n/a	56.3
Grade 8	34.5	56.3
MATH		
Grade 3	n/a	49.2
Grade 4	23.8	49.2
Grade 5	n/a	49.2
Grade 6	n/a	50.2
Grade 7	n/a	50.2
Grade 8	25.3	50.2

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 available

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 Wyoming's AYP Rules?

Figure 3 illustrates the AYP performance of the sample

elementary schools under Wyoming's 2008 AYP rules. **Only 2 elementary schools made AYP while 16 failed to make 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. The two passing schools (Wayne Fine Arts and Roosevelt) are in the right half of the figure, meaning that the higher performing students were found at these schools.

Yet these two schools also have relatively few qualifying subgroups—and thus the fewest targets to meet (because each subgroup has separate targets). For exam-

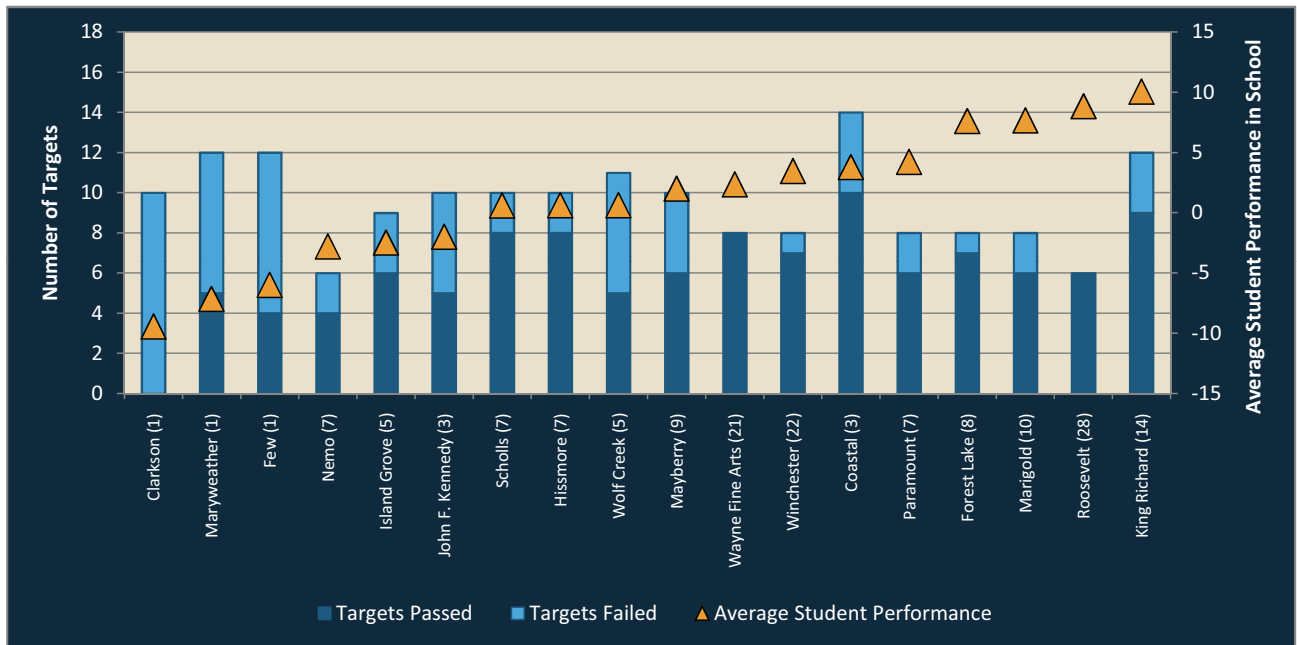


Figure 3. AYP Performance of the elementary school sample under Wyoming's 2008 AYP rules

Note: This figure indicates how each elementary school within the sample fared under Wyoming's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet 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. Winchester Elementary, 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), which is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

ple, Wayne Fine Arts passed, but had only eight targets—two targets in reading and math for their overall student population, two more for their low-income subgroup, two more for their African American subgroup, and two for their Caucasian subgroup.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Wyoming AYP rules. **Out of 18 middle schools in our sample, only 1 passed**—a high-performance school (Walter Jones) which has 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 dark blue bars show the actual proficiency rates at each school, and the light blue bars show the degree to which these proficiency rates were increased by applying the confidence interval. The or-

ange lines show the annual measurable objective needed to meet AYP. These figures show that none of the sample elementary schools and only one of the sample middle schools (Pogesto) was assisted by the confidence interval, because the math targets in Wyoming are so low, relative to the sample schools' overall performance. Moreover, we know from Figure 4 that Pogesto still failed to meet its targets for one of its subgroups, so even though it met its overall target through use of the confidence interval, the final AYP outcome was not affected.

The effect of confidence intervals on elementary and middle school reading proficiency rates is much the same (not shown). In reading, one elementary school (Nemo) and one middle school (Filmore) 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 subgroups. **In short, applying the**

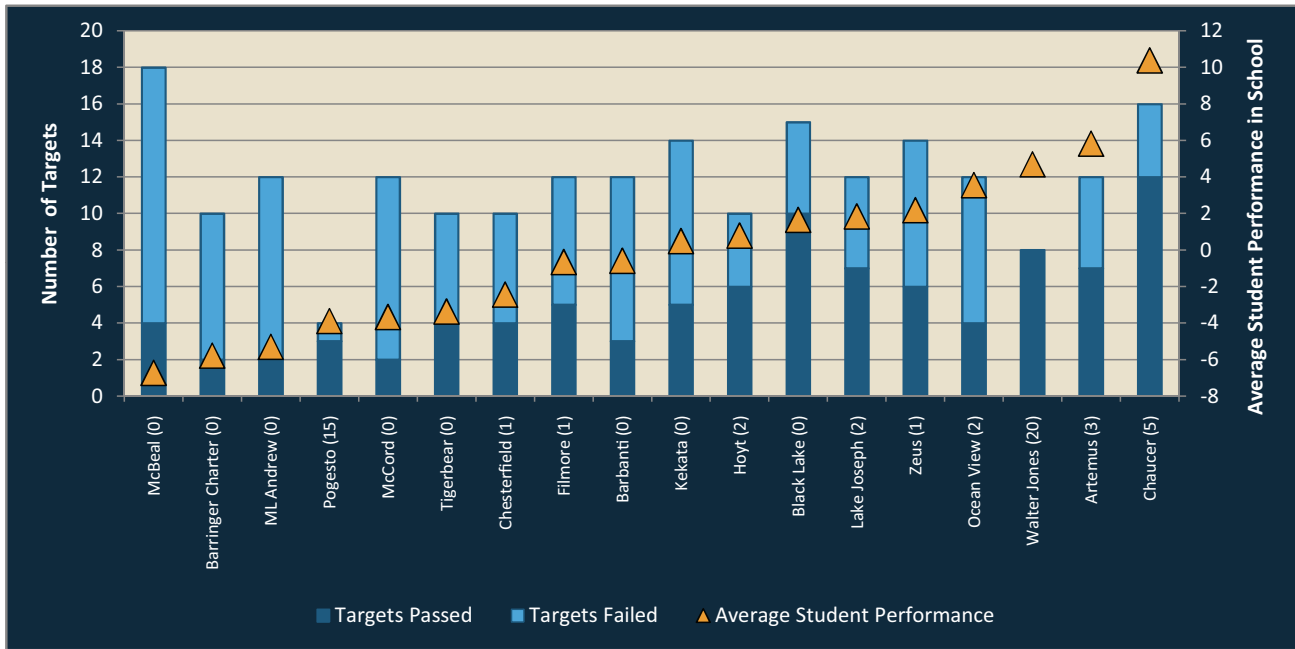


Figure 4. AYP performance of the middle school sample under Wyoming's 2008 AYP rules

Note: This figure indicates how each middle school within the sample fared under Wyoming's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet 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 did not make AYP, so any light blue means that the school failed to make AYP. Hoyt, 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), which is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

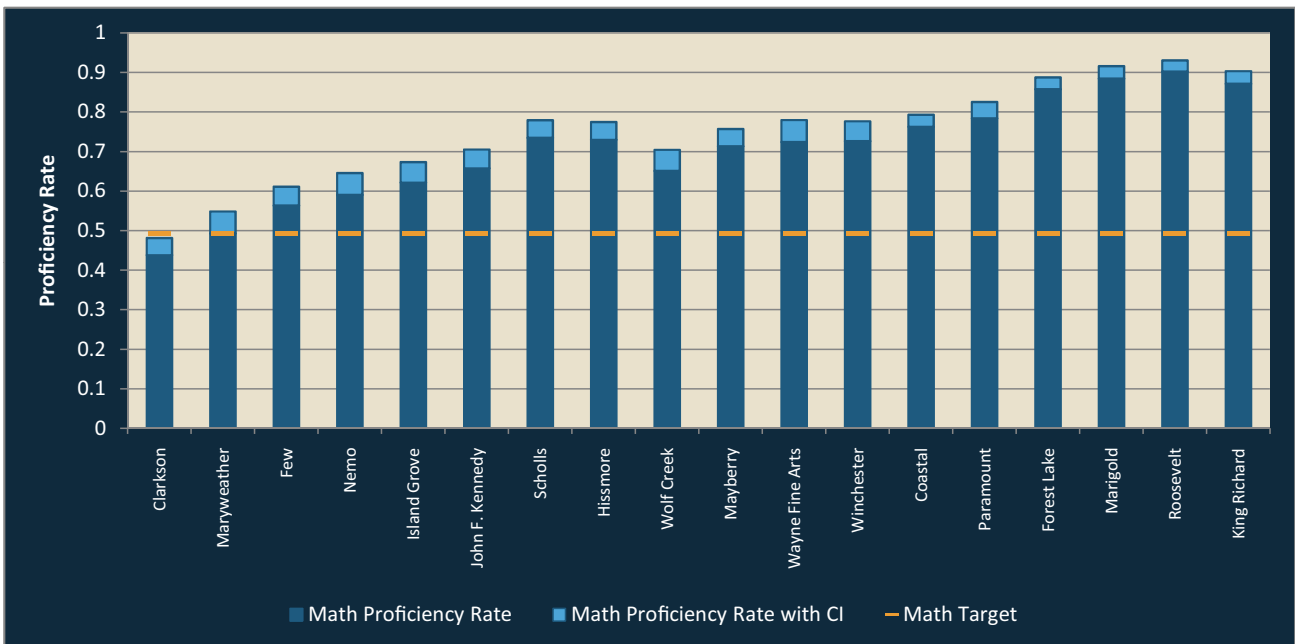


Figure 5. Impact of the confidence interval on elementary school mathematics proficiency rates under Wyoming's 2008 AYP rules

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that none of the sample elementary 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.

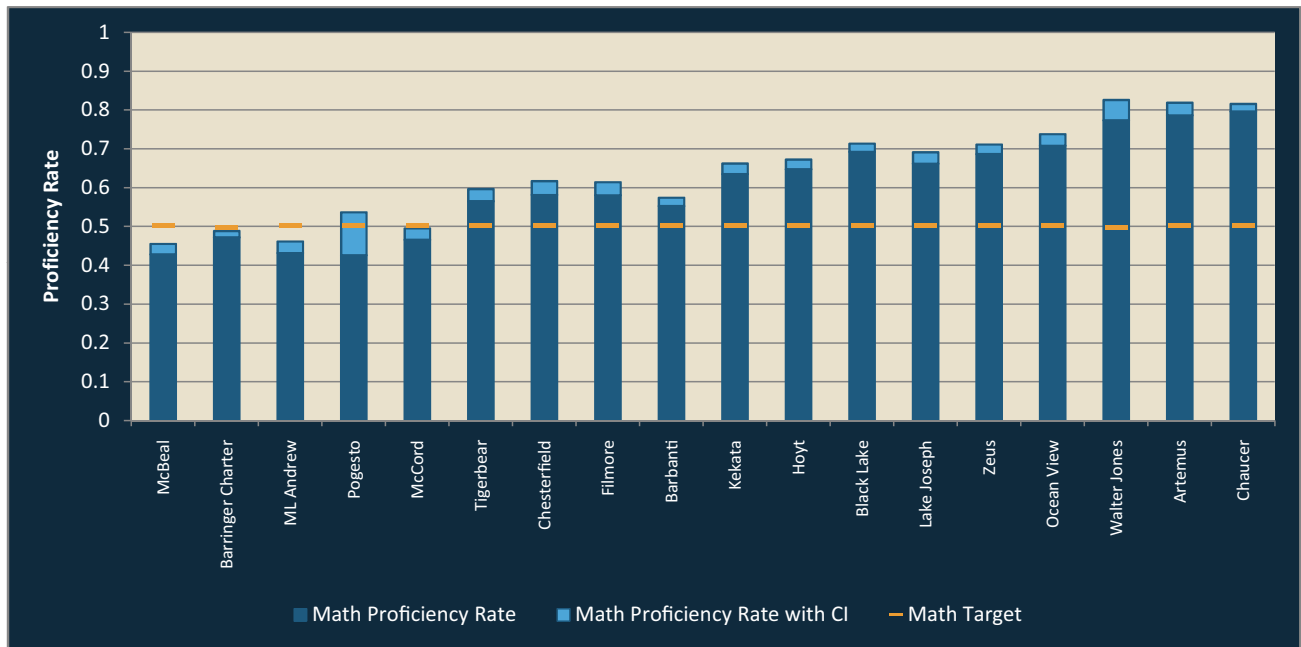


Figure 6. Impact of the confidence interval on middle school mathematics proficiency rates under Wyoming's 2008 AYP rules

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that one of the sample 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.

confidence interval has little or no effect on AYP decisions for the sample elementary and middle schools in Wyoming.⁹

Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low or mid-level performance can still pass AYP when the school has fewer targets to meet because it has to fewer subgroups. These figures do not, however, indicate which subgroups failed or passed in which school. Tables 2 and 3 list information on individual subgroup performance 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 Wyoming 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:

- Most schools, especially at the elementary level, met their targets for their overall student population.
- Four elementary schools, however, failed to meet the reading targets for their overall school population.
- One elementary school (Clarkson) failed to meet the math targets for its overall population.
- Eight middle schools failed to meet overall proficiency targets in reading, math, or both.

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 length and complexity.

Table 2. Elementary school subgroup performance of sample schools under the 2008 Wyoming 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	43.8%	23.4%	N	N	N	N	N	N	N	N					N	N					10	0	0%	N	1
Maryweather	49.3%	35.6%	Y	N	N	N	N	N	Y	N					Y	N			Y	Y	12	5	42%	N	1
Few	56.4%	36.4%	Y	N	N	N	N	N	Y	N					Y	N			Y	N	12	4	33%	N	1
Nemo	59.1%	49.3%	Y	Y					N	N									Y	Y	6	4	67%	N	7
Island Grove	62.2%	53.1%	Y	Y				N	Y	N					Y	N			Y	Y	9	6	67%	N	4
JFK	65.8%	45.9%	Y	N	N	N			Y	N	Y	N							Y	Y	10	5	50%	N	3
Scholls	73.5%	53.9%	Y	Y	Y	N			Y	Y	Y	N							Y	Y	10	8	80%	N	7
Hissmore	73.0%	57.3%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Wolf Creek	65.1%	58.9%	Y	Y	N	N		N	Y	N					N	N			Y	Y	11	5	45%	N	5
Alice Mayberry	71.4%	58.1%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	9
Wayne Fine Arts	72.4%	69.5%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	21
Winchester	72.6%	67.3%	Y	Y	Y	N									Y	Y			Y	Y	8	7	88%	N	22
Coastal	76.3%	65.1%	Y	Y	Y	N	Y	N	Y	Y	Y	N			Y	N			Y	Y	14	10	71%	N	3
Paramount	78.4%	67.2%	Y	Y					Y	N					Y	N			Y	Y	8	6	75%	N	7
Forest Lake	85.8%	75.0%	Y	Y	Y	N			Y	Y									Y	Y	8	7	88%	N	8
Marigold	88.5%	78.3%	Y	Y	Y	N			Y	N									Y	Y	8	6	75%	N	10
Roosevelt	90.2%	84.1%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	28
King Richard	87.2%	82.3%	Y	Y	Y	N	Y	N	Y	N					Y	Y			Y	Y	12	9	75%	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.

- Three of the 16 elementary schools (Hissmore, Winchester, and Forest Lake) that failed to make AYP missed only for the SWD subgroup.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. First, the performance of students with disabilities is proving challenging for schools under Wyoming's system, particularly in middle schools, where this subgroup tends to have enough students to meet the state's minimum *n* of 30. In fact, all middle schools with qualifying SWD subgroups failed to meet targets for this subgroup in both reading and math. Students with LEP are also struggling

to meet the state's targets; every school with a LEP population large enough to qualify as a separate subgroup failed to meet its reading targets for these students. Low-income students also struggled; more than half of the schools with low-income subgroups failed to meet their proficiency targets for this group.

Other state reports contain a section comparing some of the characteristics of the sample schools that made AYP vs. those that did not. In Wyoming, such comparisons are less helpful, given that there were so few schools making AYP (only two elementary and one middle school).

Table 3. Middle school subgroup performance of sample schools under the 2008 Wyoming 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	42.8%	44.2%	N	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	47.2%	41.8%	N	N	N	N			N	N	N	N			Y	Y					10	2	20%	N	0	
ML Andrew	43.2%	45.5%	N	N	N	N			N	N	N	N			N	N				Y	Y	12	2	17%	N	0
Pogesto	42.6%	44.4%	Y	N																Y	Y	4	3	75%	N	15
McCord Charter	46.6%	51.3%	N	N	N	N			N	N	N	N			N	N				Y	Y	12	2	17%	N	0
Tigerbear	56.5%	44.9%	Y	N	N	N			Y	N	N	N								Y	Y	10	4	40%	N	0
Chesterfield	58.1%	46.8%	Y	N	N	N			Y	N	N	N								Y	Y	10	4	40%	N	1
Filmore	58.0%	55.0%	Y	Y	N	N	N	N	Y	N					N	N				Y	Y	12	5	42%	N	1
Barbanti	55.2%	53.9%	Y	N	N	N	N	N	N	N					N	N				Y	Y	12	3	25%	N	0
Kekata	63.5%	55.9%	Y	Y	N	N	N	N	Y	N	N	N			N	N				Y	Y	14	5	36%	N	0
Hoyt	64.7%	58.1%	Y	Y	N	N			Y	N	Y	N								Y	Y	10	6	60%	N	2
Black Lake	69.2%	57.5%	Y	Y	N	N	Y		Y	N	Y	N	Y	Y	Y	N				Y	Y	15	10	67%	N	0
Lake Joseph	66.2%	61.1%	Y	Y	N	N	N	N	Y	Y					Y	N				Y	Y	12	7	58%	N	2
Zeus	68.6%	61.1%	Y	Y	N	N	N	N	Y	N	Y	N			N	N				Y	Y	14	6	43%	N	1
Ocean View	70.8%	72.0%	Y	Y	N	N	N	N	N	N					N	N				Y	Y	12	4	33%	N	2
Walter Jones	77.3%	70.9%	Y	Y					Y	Y					Y	Y				Y	Y	8	8	100%	Y	20
Artemus	78.6%	69.9%	Y	Y	N	N			Y	N			Y	Y	N	N				Y	Y	12	7	58%	N	3
Chaucer	79.6%	79.4%	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.

In general, schools not making AYP had higher numbers of accountable subgroups than did schools making AYP, but other striking differences were not apparent.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Wyoming's NCLB accountability system is, in many respects, behaving like those in other states. For example, among the elementary schools in our sample, Roosevelt and Wayne Fine Arts made AYP in the greatest number of

states—28 and 21, respectively. And these schools made AYP in Wyoming, too. Likewise, the elementary and middle schools that fail to make AYP in the greatest number of states also fail AYP in Wyoming.

But Wyoming is home to at least one anomaly. Consider Winchester Elementary (see Figure 3). It made AYP in 22 of the 28 states in our sample, yet failed to make AYP in Wyoming. Examining Table 2, one can see that Winchester meets the minimum number (30) for the SWD subgroup, and does not meet its reading target, probably due to harder than average proficiency cut scores.

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Wyoming 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	13	7	13
Students with limited English proficiency	7	3	7
Low-income students	17	2	11
African-American students	6	0	4
Asian/Pacific Islander students	0	0	0
Hispanic students	9	2	7
American Indian/Alaska Native students	0	0	0
White students	17	0	1

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Wyoming 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	9	8	8
Low-income students	17	6	14
African-American students	11	7	10
Asian/Pacific Islander students	4	0	0
Hispanic students	14	9	11
American Indian/Alaska Native students	1	1	1
White students	17	0	0

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 Wyoming's AYP rules and AMOs for 2008. We found that only 2 elementary schools and 1 middle school—3 in all from a total of 36—would have made AYP in Wyoming. Looking across the 28 state accountability systems

examined in the study, we find that the number of elementary schools making AYP in Wyoming was exceeded in 20 other sample states (Wyoming ties 5 other states with only 2 elementary schools making AYP). This is partly due to Wyoming's proficiency standards which are relatively difficult compared to other states, and Wyoming's comparatively small minimum subgroup size, meaning that more subgroups in Wyoming are likely held accountable for performance than in other states.

Because 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, Wyoming'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 Wyoming math and reading targets for their student populations as a whole. In the pre-NCLB era, such schools might have been considered to be effective or at least not in need of improvement, even though sizable numbers of their pupils 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 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,¹⁰ doesn't the disproportionate failure of these students to meet Wyoming's targets indicate that a new approach is needed for holding schools accountable for the performance of these students? Yes, schools should redouble their efforts to boost achievement for ELL students and students with disabilities, as for other students, but when almost no school is able to meet the goal, perhaps that indicates that the goal is unrealistic. These will be critical considerations for Congress as it takes up NCLB 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

¹⁰ See footnote 5.

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.