



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 Montana’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 Montana’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 Montana’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 Montana’s AYP rules for academic year 2007–2008 (shortened to “2008” in this report).

Here are some key findings:

- We estimate that **15 of 18 elementary schools and all 18 middle schools** in our sample **failed to make adequate yearly progress** in 2008 under Montana’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**

elementary schools that made AYP in Montana was exceeded in 15 other sample states; Montana ties with 4 other states that each has 3 schools that made AYP (see Figure 1). Montana also joins Idaho, Massachusetts, South Carolina, and North Dakota with no middle schools that made AYP in the sample.

- Some elementary schools in our sample that failed to make AYP in Montana are meeting expected targets for their overall pupil populations² but failed because of the performance of individual subgroups, particularly students with disabilities (SWDs), and English language learners.
- One of the sample middle schools did not make AYP in Montana even though it did so in 23 other states. This may be because some of Montana’s annual measurable objectives (AMOs, the proficiency targets needed to make AYP) are relatively high compared to many of the other states examined. In fact, **the way Montana’s cut scores and annual targets work together may make it difficult for schools to**

Several factors combine to make **Montana’s** AYP rules relatively difficult compared to the other states examined in the study. Montana’s proficiency cut scores in math are relatively high, meaning that a student who meets the math proficiency standards in other states might have a harder time doing so in Montana. In addition, the annual targets in Montana are high compared to other states, meaning that schools in Montana must get larger percentages of their students to the “proficient” level than in many other states in order to make AYP. In fact, from our sample of 36 schools, only three elementary and no middle schools met AYP, and none of these three elementary schools had traditionally academically disadvantaged subgroups (such as low income or African American).

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 Montana Criterion Referenced Test.

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 simply are not treated as their own subgroup.

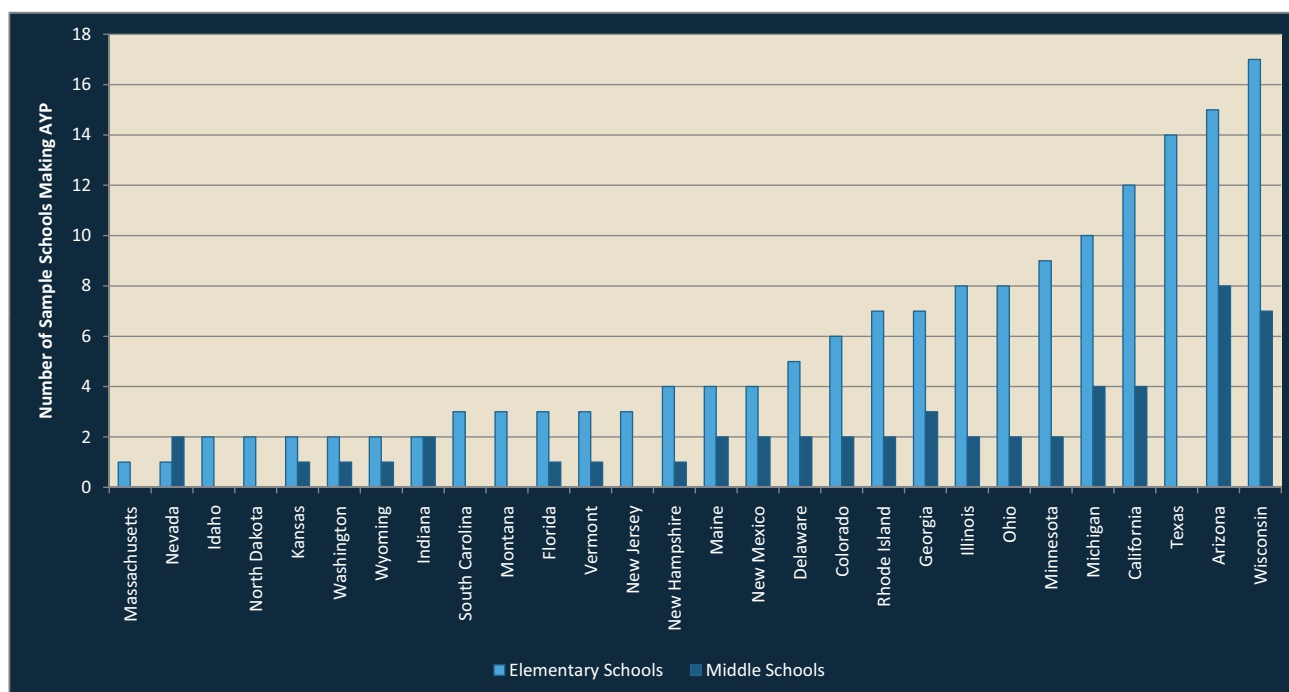


Figure 1. Number of sample schools making AYP by state

Note: Middle schools were not included for Texas and New Jersey; absence of a middle school bar in those states means “not applicable” as opposed to zero. States like Idaho and North Dakota, however, have zero passing middle schools.

make AYP. Specifically, the state’s reading cut scores are fairly low but its annual reading targets are demanding; on the other hand, the state’s math targets are fairly low, and its math cut scores are somewhat high.

- In Montana, as in most states, schools with fewer subgroups attained AYP more easily than schools with more subgroups, even when their average student performance is lower than that in some failing schools. In other words, schools with greater diversity and size face greater challenges in making AYP.
- Montana applies a 95% confidence interval (a statistical margin of error) to its proficiency rate calculations. The confidence interval had little or no impact, however, on final AYP outcomes for sample elementary and middle schools in Montana, partly

because sample schools already missed AYP for their subgroup performance.

- As in other states, middle schools in Montana had greater difficulty reaching AYP than did elementary schools, primarily because their student populations are larger and therefore have more qualifying subgroups—not because their student achievement is lower than in the elementary schools.
- Almost all schools with enough SWDs and limited English proficiency (LEP) students to qualify as separate subgroups failed to meet their targets for those groups.³

Introduction

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

³ Note that we use “LEP students” and “English language learners” interchangeably to refer to students in the same subgroup. 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 Montana Criterion Referenced Test, 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 income⁴ 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 discussion of the methodology for this project along with the characteristics of the school sample.⁵

⁴ Low-income students are those who receive a free or reduced-price lunch.

⁵ We gave all schools in our sample pseudonyms in this report.

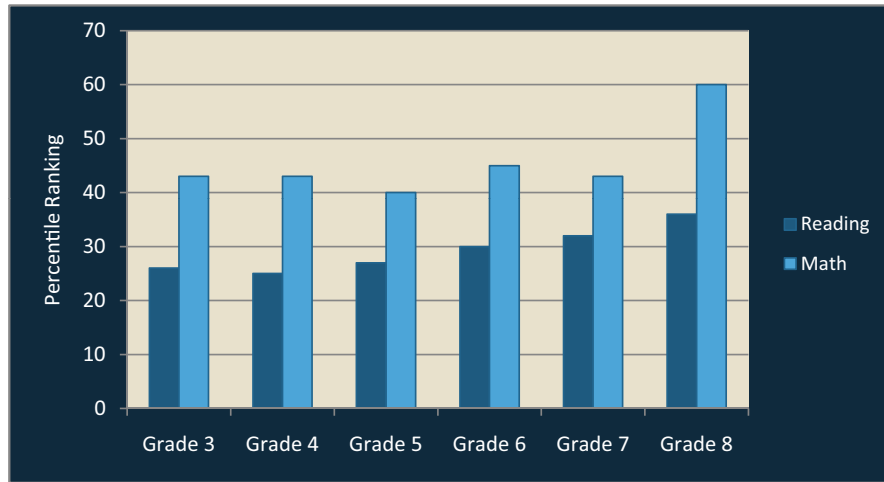


Figure 2. . Montana reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Montana’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 Montana’s cut scores for reading are below the 40th percentile and all cut scores for math are at or above the 60th percentile.

Table 1. Montana AYP rules for 2008

Subgroup minimum <i>n</i>	Race/ethnicity: 40	
	SWDs: 40	
	Low-income students: 40	
	LEP students: 40	
CI	Applied to proficiency rate calculations?	
	Yes; 95% CI used	
AMOs	Baseline proficiency levels as of 2002 (%)	2008 targets (%)
READING/LANGUAGE ARTS		
Grade 3	74	83
Grade 4	74	83
Grade 5	74	83
Grade 6	74	83
Grade 7	74	83
Grade 8	74	83
MATH		
Grade 3	51	68
Grade 4	51	68
Grade 5	51	68
Grade 6	51	68
Grade 7	51	68
Grade 8	51	68

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

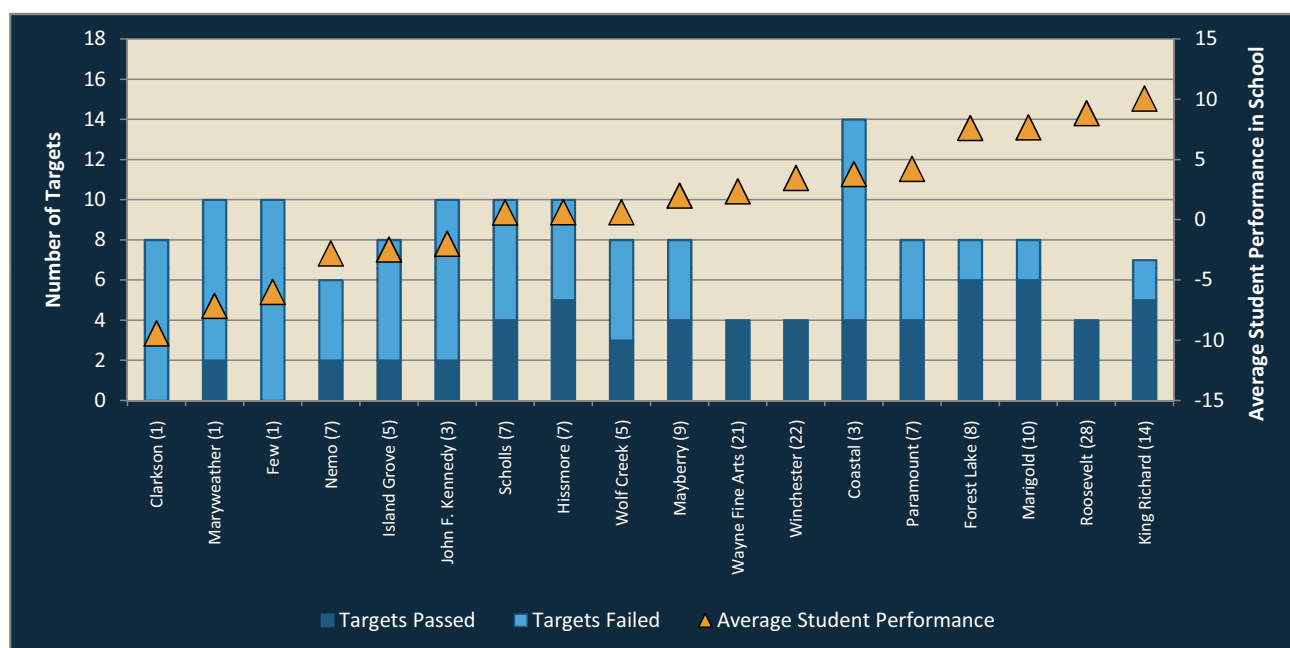


Figure 3. AYP performance of the elementary school sample under Montana’s 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under Montana’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. Marigold Elementary, for example, met six of its eight targets, but because it didn’t meet them all, it didn’t make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). This is measured by the average MAP performance of students within the school, 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

Proficiency cut score estimates for the Montana Criterion-Referenced Test (Montana CRT) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Montana’s proficiency standards in reading ranked about average compared with the standards set by the other 25 states in that study, and its proficiency standards in math ranked above average. These cut scores were used to estimate whether students would have scored as proficient or better on the Montana 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 Montana 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 Montana’s 2008 AYP rules.

Table 1 shows the pertinent Montana AYP rules that were applied to elementary and middle schools in the current study. Montana’s minimum subgroup size is 40, which is about average, compared to most other states we examined.⁶

Furthermore, Montana, like most states, applies a 95% confidence interval (or margin of statistical error) to its measurements of student proficiency rates.⁷ So, for instance, even though schools are supposed to get 68% of their grade 3 students to the proficient level on the state math test, as well as 68% of the grade 3 students in each subgroup, applying the confidence interval means that the real target can be lower, particularly with smaller groups.

⁶ It’s worth noting, however, that schools in Montana are likely to be small and an *n* size of 40, though average, may in fact exclude more subgroups than would be the case in states with larger schools overall.

⁷ We also conducted an analysis to show the effect of confidence intervals on the reading and math proficiency rates for elementary and middle schools. We describe those results later in the report.

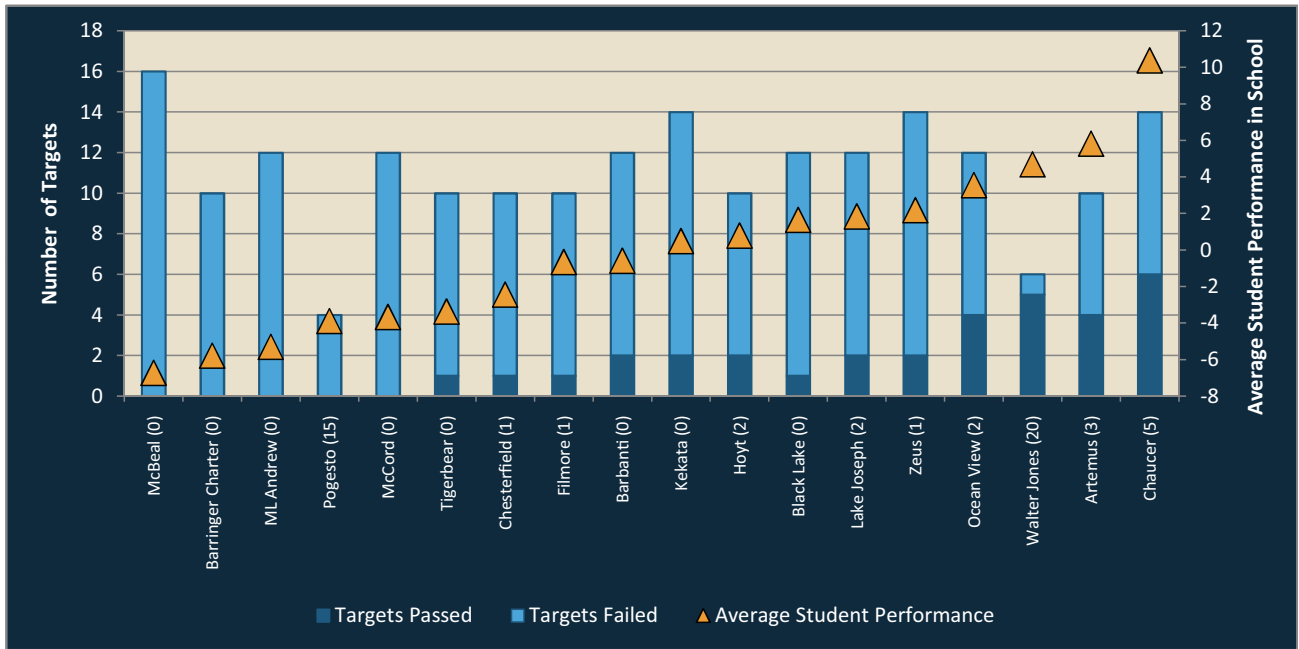


Figure 4. AYP performance of the middle school sample under Montana's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under Montana'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. Walter Jones Middle School, for example, met five of its six targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). This is measured by the average MAP performance of students within the school, and its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

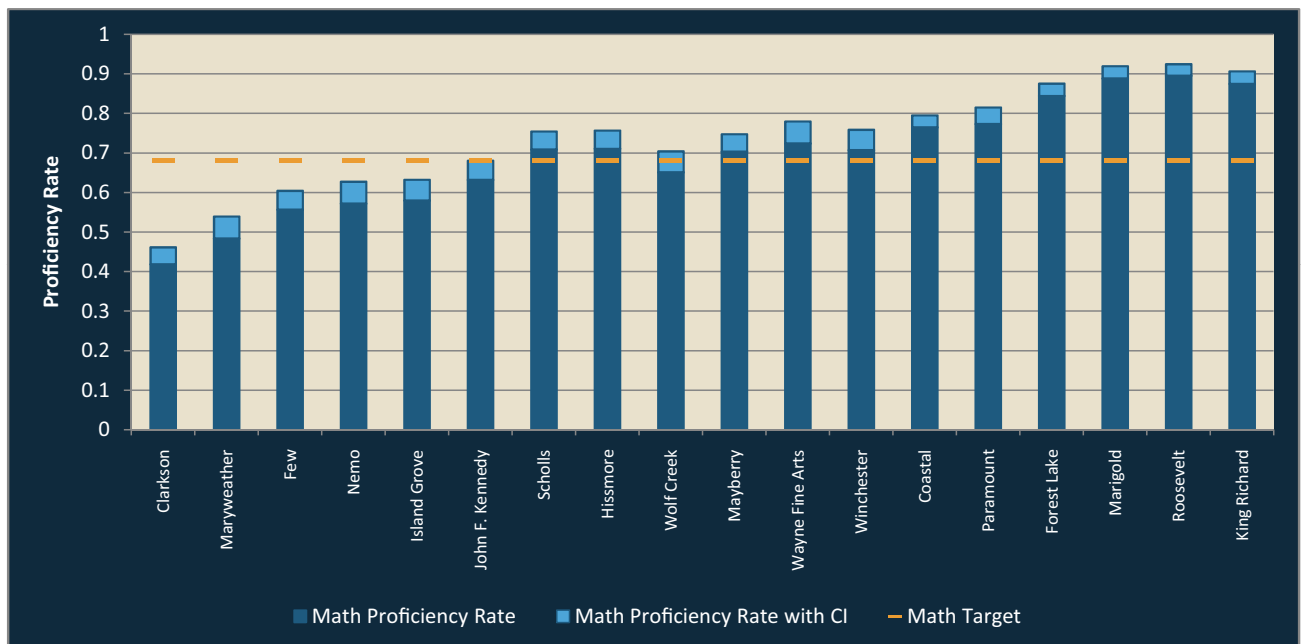


Figure 5. Impact of the confidence interval on elementary school math proficiency rates under Montana's 2008 AYP rules

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

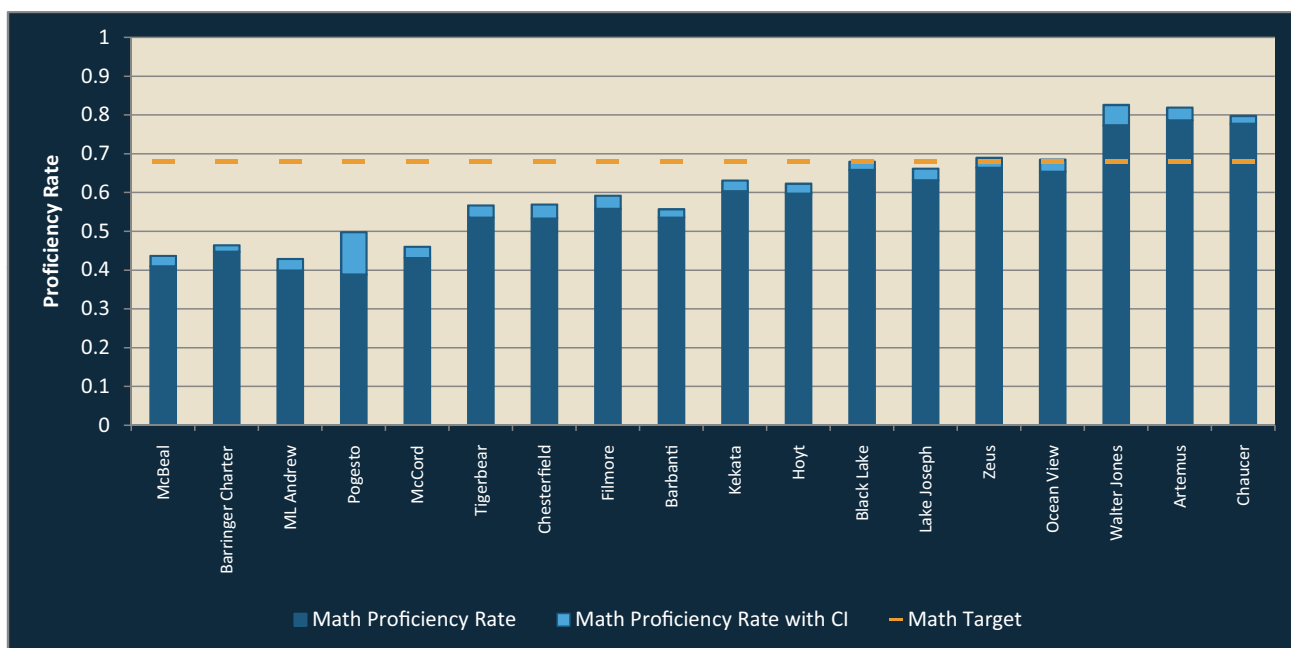


Figure 6. Impact of the confidence interval on middle school math proficiency rates under Montana’s 2008 AYP rules

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 two of the sample elementary schools, Black Lake and Zeus, were assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.

Note that we were unable to examine the impact of NCLB’s “safe harbor” provision. This provision permits a school to make AYP even if some of its subgroups fail, as long as it reduces the number of nonproficient students within any failing subgroup by at least 10% relative to the previous year’s performance. Because we had access to only a single academic year’s data (2005–2006), we were not able to include this in our analysis. As a result, it’s possible that some of the schools in our sample that failed to make AYP according to our estimates would have made AYP under real conditions.

Furthermore, attendance and test participation rates are beyond the scope of the study. Note that most states include attendance rates as an additional indicator in their NCLB accountability system for elementary and middle schools. In addition, federal law requires 95% of each school’s students—and 95% of the students in each subgroup—to participate in testing.

To reiterate, then, AYP decisions in the current study are modeled solely on test performance data for a single academic year. For each school, we calculated reading and

math proficiency rates (along with any confidence 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 Montana’s AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Montana’s 2008 AYP rules. Only 3 elementary schools made AYP while 15 failed to make it. The triangles in Figure 3 show the average academic performance of students within the school, with negative values indicating below-grade-level performance for the average student, and positive values indicating above-grade-level performance. All passing schools are in the right half of the figure, meaning that the higher performing students were found at these schools.

Yet almost without regard to average student performance, the only schools made AYP were those with relatively few

Table 2. Elementary school subgroup performance of sample schools under the 2008 Montana 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	41.9%	47.3%	N	N			N	N	N	N					N	N					8	0	0%	N	1	
Maryweather	48.4%	57.1%	N	N			N	N	N	N					N	N				Y	Y	10	2	20%	N	1
Few	55.7%	59.5%	N	N	N	N	N	N	N	N					N	N					10	0	0%	N	1	
Nemo	57.2%	75.3%	N	N					N	N									Y	Y	6	2	33%	N	7	
Island Grove	58.0%	72.4%	N	N					N	N					N	N				Y	Y	8	2	25%	N	4
JFK	63.2%	67.5%	Y	N	N	N			N	N	N	N							Y	N	10	2	20%	N	3	
Scholls	70.9%	74.7%	Y	N	N	N			Y	N	N	N							Y	Y	10	4	40%	N	7	
Hissmore	71.1%	77.5%	Y	N	N	N			Y	N	Y	N							Y	Y	10	5	50%	N	7	
Wolf Creek	65.1%	73.5%	Y	N					N	N					N	N				Y	Y	8	3	38%	N	5
Alice Mayberry	70.3%	80.3%	Y	Y					N	N	N	N								Y	Y	8	4	50%	N	9
Wayne Fine Arts	72.4%	86.8%	Y	Y															Y	Y	4	4	100%	Y	21	
Winchester	70.8%	83.9%	Y	Y															Y	Y	4	4	100%	Y	22	
Coastal	76.5%	79.6%	Y	N	N	N	N	N	Y	N	N	N			N	N				Y	Y	14	4	29%	N	3
Paramount	77.3%	79.9%	Y	Y					N	N					N	N				Y	Y	8	4	50%	N	7
Forest Lake	84.5%	87.6%	Y	Y	N	N			Y	Y									Y	Y	8	6	75%	N	8	
Marigold	88.8%	89.5%	Y	Y	Y	N			Y	N									Y	Y	8	6	75%	N	10	
Roosevelt	89.6%	94.2%	Y	Y															Y	Y	4	4	100%	Y	28	
King Richard	87.5%	89.8%	Y	Y	N	N			Y										Y	Y	7	5	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.

qualifying subgroups—and thus the fewest targets to meet. For example, Wayne Fine Arts and Winchester passed, but had only four targets each. Each must make AYP for its overall student population in reading and math (two targets) and for its white population (two more targets).

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Montana AYP rules. **Of 18 middle schools in our sample, none passed.**

Figures 5 and 6 indicate the degree to which schools' overall math proficiency rates are aided by Montana's

confidence interval for elementary and middle schools, respectively. On these figures, the darker portion of the bars show the actual proficiency rates at each school, and the lighter portion of the bars show the degree to which these proficiency rates are 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 (JFK and Wolf Creek) and three middle schools (Black Lake, Zeus, and Ocean View) are assisted by the confidence intervals to meet their overall math targets (note how the

Table 3. Middle school subgroup performance of sample schools under the 2008 Montana 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	41.0%	56.8%	N	N	N	N	N	N	N	N	N	N			N	N	N	N	N	N	16	0	0%	N	0
Barringer Charter	44.8%	62.9%	N	N	N	N			N	N	N	N			N	N					10	0	0%	N	0
ML Andrew	39.9%	59.9%	N	N	N	N			N	N	N	N			N	N			N	N	12	0	0%	N	0
Pogesto	38.9%	68.5%	N	N															N	N	4	0	0%	N	15
McCord Charter	43.2%	63.4%	N	N	N	N			N	N	N	N			N	N			N	N	12	0	0%	N	0
Tigerbear	53.6%	59.9%	N	N	N	N			N	N	N	N							Y	N	10	1	10%	N	0
Chesterfield	53.3%	63.2%	N	N	N	N			N	N	N	N							Y	N	10	1	10%	N	1
Filmore	55.8%	71.4%	N	N	N	N			N	N					N	N			Y	N	10	1	10%	N	1
Barbanti	53.6%	66.0%	N	N	N	N	N	N	N	N					N	N			Y	Y	12	2	17%	N	0
Kekata	60.4%	68.5%	N	N	N	N	N	N	N	N	N	N			N				Y	Y	14	2	14%	N	0
Hoyt	59.7%	72.3%	N	N	N	N			N	N	N	N							Y	Y	10	2	20%	N	2
Black Lake	65.8%	73.3%	N	N	N	N			N	N	N	N			N	N			Y	N	12	1	8%	N	0
Lake Joseph	63.2%	76.3%	N	N	N	N	N	N	N	N					N	N			Y	Y	12	2	17%	N	2
Zeus	66.4%	74.3%	Y	N	N	N	N	N	N	N	N	N			N	N			Y	N	14	2	14%	N	1
Ocean View	65.4%	83.7%	Y	Y	N	N	N	N	N	N					N	N			Y	Y	12	4	33%	N	2
Walter Jones	77.3%	85.1%	Y	Y					Y	N									Y	Y	6	5	83%	N	20
Artemus	78.6%	82.0%	Y	Y	N	N			N	N					N	N			Y	Y	10	4	40%	N	3
Chaucer	77.7%	87.9%	Y	Y	N	N	N	N	N	N			Y	Y	N	N			Y	Y	14	6	43%	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.

orange line falls within the light blue band). Figures 3 and 4 show, however, that all five of these schools still fail to meet some of their subgroup targets. The same is true for reading (not shown). So, although a few schools met their overall targets with the help of the confidence interval, they still missed subgroup targets, and therefore, failed to make AYP. Overall, the confidence interval had little or no impact on final AYP outcomes for sample elementary and middle schools in Montana.⁸

Where Do Schools Fail?

Figures 3 and 4 illustrate that a few elementary schools with only middling 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 in which school. Information on individual subgroup performance appears in Tables 2 and 3 for elementary and middle schools, respectively.

⁸ 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 4. Summary of subgroup performance of sample elementary schools under the 2008 Montana 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	7	8
Students with limited English proficiency	4	4	4
Low-income students	15	9	13
African-American students	5	4	5
Asian/Pacific Islander students	0	0	0
Hispanic students	7	7	7
American Indian/Alaska Native students	0	0	0
White students	16	0	1

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Montana AYP rules

SUBGROUP	Number of schools with qualifying subgroups	Number of schools where subgroup failed to meet math target	Number of schools where subgroup failed to meet reading target
Students with disabilities	16	16	16
Students with limited English proficiency	7	7	7
Low-income students	17	16	17
African-American students	10	10	10
Asian/Pacific Islander students	1	0	0
Hispanic students	13	13	13
American Indian/Alaska Native students	1	1	1
White students	17	4	9

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 Montana 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:

- Five elementary schools failed to meet both the math and reading targets for their overall school population. Five more elementary schools failed to meet their overall targets in reading.
- Most middle schools failed to meet their overall reading and math targets.
- Two (Forest Lake and King Richard) of the 15 failing elementary schools missed only for the SWD subgroup.
- One middle school (Walter Jones) failed only for its low-income subgroup.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively.⁹ First, almost every school with a large enough academically disadvantaged population to qualify as a separate subgroup (e.g., low income, African American, Hispanic) failed to meet its targets for these students. Students with disabilities and limited English proficiency did just as poorly, failing in every elementary or middle school in which that subgroup was accountable. Second, elementary schools did slightly better than middle schools because they have fewer subgroups.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Montana'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 Montana, too (though they are the only 3 to do so). Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed in Montana.

But Montana is also home to at least one anomaly. Consider Walter Jones (see Figure 4). It made AYP in 20 of

the 28 states in our sample, but not in Montana. In examining Table 3, we can see that Walter Jones failed to meet the reading target for its low-income subgroup. Although Montana's reading cut scores at the middle school grades are fairly low (except at eighth grade), its annual targets are relatively high (i.e., 83% are expected to reach proficiency) compared with many other states. This may account for the fact that this group missed its target, even though it passed in most other states.

Other state reports contain a section comparing some of the characteristics of the sample schools that made AYP versus those that did not. In Montana, none of the sample middle schools made AYP, and among elementary schools, the only striking difference between schools that made AYP and those that didn't is that the former had fewer subgroups.

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 Montana's AYP rules (and AMOs) for 2008. We found that only 3 elementary schools and no middle schools—3 in all, from of a sample of 36—would have made AYP in Montana. Looking across the 28 state accountability systems examined in the study, this puts Montana in the lower middle part of the sample distribution, as shown in Figure 1. It's worth noting that the way Montana's cut scores and annual targets work together may make it difficult for schools to make AYP.

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, Montana'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. Many of the sample elementary and middle

⁹ Recall that elementary schools did better on Montana's math test than middle school students did, perhaps because Montana's proficiency scores are lower in reading (see Figure 2).

schools met their reading and math targets for their student populations as a whole, that is, without considering subgroup results. 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 students aren't meeting state standards. Disaggregating data by race, income, and so on has made those students visible. That is surely a positive step.

Yet NCLB's design flaws are also readily apparent. Does it make sense that having fewer subgroups enhances the likelihood of making AYP? Even if actual participation

guidelines for English language learners and students with disabilities are more generous under the current state assessment system,¹⁰ doesn't the massive failure of these students to meet Montana'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 pupils, 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 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.

¹⁰ See footnote 3.