



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 Idaho’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 Idaho’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 Idaho’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 Idaho’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 middle schools** in our sample failed to make AYP in 2008 under Idaho’s accountability system. The high failure rate is partly explained by our sample, which intentionally includes some schools with a relatively large population of low-performing students. It’s also partly explained by Idaho’s minimum subgroup size (34), which is relatively small in comparison to most other states examined in the study. This means that schools in Idaho will be ac-

countable for more subgroups than would similar schools in other states with larger subgroup sizes.

- Looking across the 28 state accountability systems examined in the study, we find the number of elementary schools that made AYP in **Idaho was exceeded in 20 other states (Idaho ties 5 other states in having just 2 elementary schools that made AYP). Idaho joins Massachusetts, Montana, South Carolina, and North Dakota in having no middle schools that make AYP in our sample (see Figure 1).**
- Many of the schools in our sample that failed to make AYP in Idaho are meeting expected targets for their overall populations but failed because of the performance of individual subgroups, particularly students with disabilities (SWD) and English language learners.²
- Schools with fewer subgroups attained AYP more easily in Idaho than schools with more subgroups, even when their average student performance is much lower. In other words, **schools with greater**

Only two elementary schools and none of the middle schools in our sample made AYP in 2008 under **Idaho’s** accountability system. A number of factors likely contribute to this low number. First, Idaho’s minimum subgroup size is 34, which is relatively small in comparison to most other states examined in the study. This means that schools in Idaho will be accountable to more subgroups than would similar schools in other states with higher subgroup sizes. Not only did many disadvantaged subgroups fail their annual targets in Idaho, quite a few white subgroups failed as well, especially in reading. Another factor which makes it difficult for schools to make AYP in Idaho is that no confidence interval (margin of error) is applied to proficiency rate calculations.

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 Idaho Standards Achievement Tests.

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.

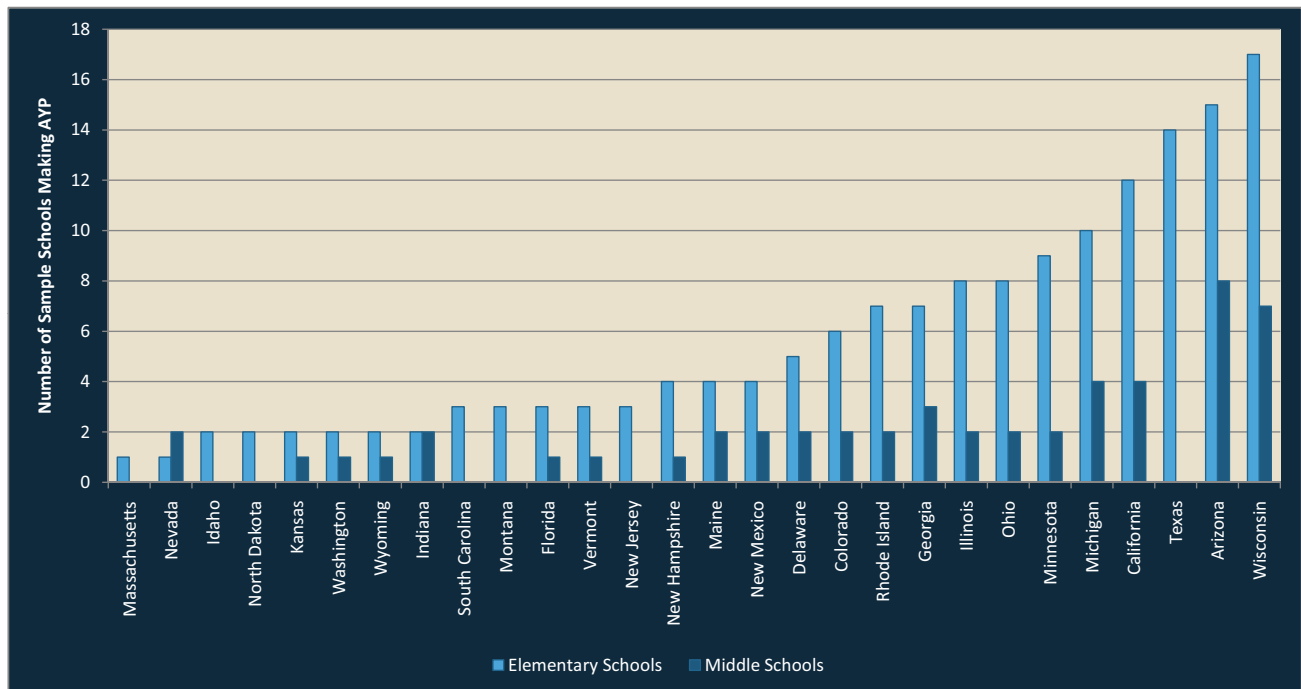


Figure 1. Number of sample schools making AYP by state

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.

diversity and size face greater challenges in making AYP. This is true in other states as well.

- Middle schools have somewhat greater difficulty reaching AYP in Idaho than do elementary schools, primarily because their student populations are larger and therefore have more qualifying subgroups—not because their student achievement is any lower than in the elementary schools.
- A strong predictor of whether or not a school will make AYP under Idaho’s system is whether it has enough SWDs and enough English language learners to qualify as a separate subgroup. Every school with an SWD or limited English proficient (LEP)³ subgroup failed to make AYP, in part because these students did not meet the state’s proficiency targets in reading or math.⁴

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance on Idaho’s tests and those of 25 other states to the Northwest Evaluation Association’s (NWEA’s) Measures of Academic Progress (MAP), a computerized adaptive test used in schools nationwide. This single common scale permitted cross-state comparisons of each state’s reading and math proficiency standards to measure school performance under the No Child Left Behind (NCLB) Act of 2001. That study revealed profound differences in states’ proficiency standards (i.e., how difficult it is to achieve proficiency on the state test), and even across grades within a single state.

Our study expands on *The Proficiency Illusion* by examining other key factors of state NCLB accountability

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

⁴ 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 Idaho Standards Achievement Tests, the standardized state tests. 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 tests or that allow them to take alternative assessments. In this study, however, no valid MAP scores were omitted from consideration.

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-income⁵ 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 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.⁶

Proficiency cut score estimates for the Idaho Standards Achievement Tests (ISAT) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Idaho'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 Idaho test, given their performance on MAP. Student test data and subgroup designations were then used to determine how these 18 elementary and 18 mid-

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

⁶ We gave all schools in our sample pseudonyms in this report.

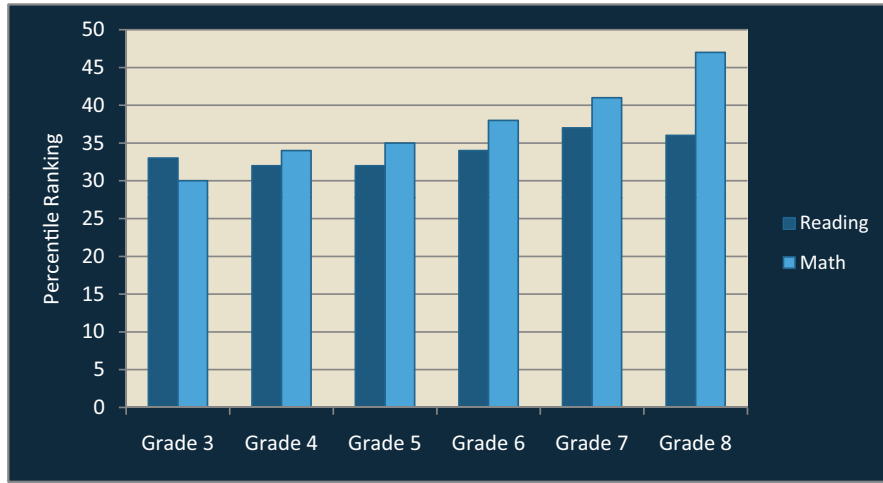


Figure 2. Idaho reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Idaho’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 Idaho’s cut scores are below the 50th percentile.

Table 1. Idaho AYP rules for 2008

Subgroup minimum <i>n</i>	Race/ethnicity: 34	
	SWDs: 34	
	Low-income students: 34	
	LEP students: 34	
CI	Applied to proficiency rate calculations?	
	CI not used	
AMOs	Baseline proficiency levels as of 2002 (%)	2008 targets (%)
READING/LANGUAGE ARTS		
Grade 3	66	78
Grade 4	66	78
Grade 5	66	78
Grade 6	66	78
Grade 7	66	78
Grade 8	66	78
MATH		
Grade 3	51	70
Grade 4	51	70
Grade 5	51	70
Grade 6	51	70
Grade 7	51	70
Grade 8	51	70

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

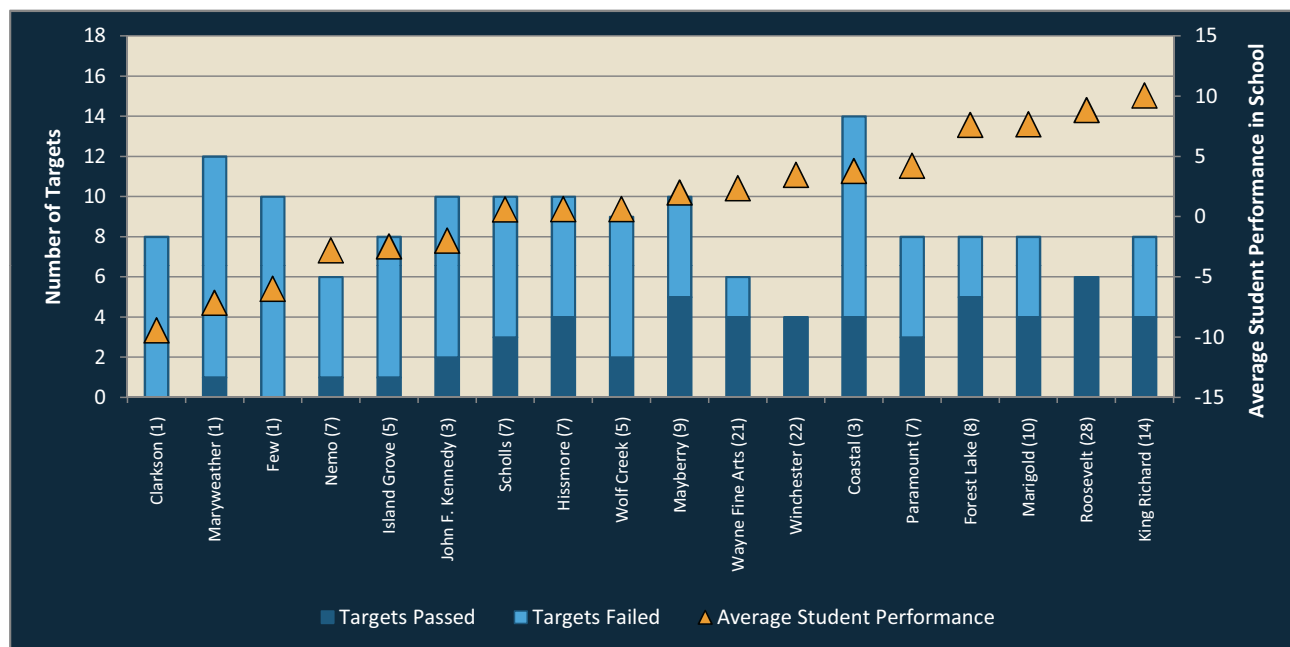


Figure 3. AYP performance of the elementary school sample under Idaho's 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under Idaho'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. Wayne Fine Arts, for example, met four 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 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 in the study.

dle schools would have fared under Idaho AYP rules for 2008. (In other words, the school data are from 2005–2006, as are our proficiency cut score estimates, but we are applying them against Idaho's 2008 AYP rules.)

Table 1 shows the pertinent Idaho AYP rules that we applied to elementary and middle schools in the current study. **Idaho's minimum subgroup size is 34, which is relatively small in comparison to most other states examined in the study. This means that schools in Idaho will be accountable for more subgroups than would similar schools in other states with larger subgroup sizes.**⁷

Furthermore, although the majority of states examined in the study apply confidence intervals to their student proficiency rates, Idaho does not. This means that **Idaho schools will have greater difficulty achieving their AMOs than equivalent schools in other states that re-**

port a confidence interval around their school proficiency rates.

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

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

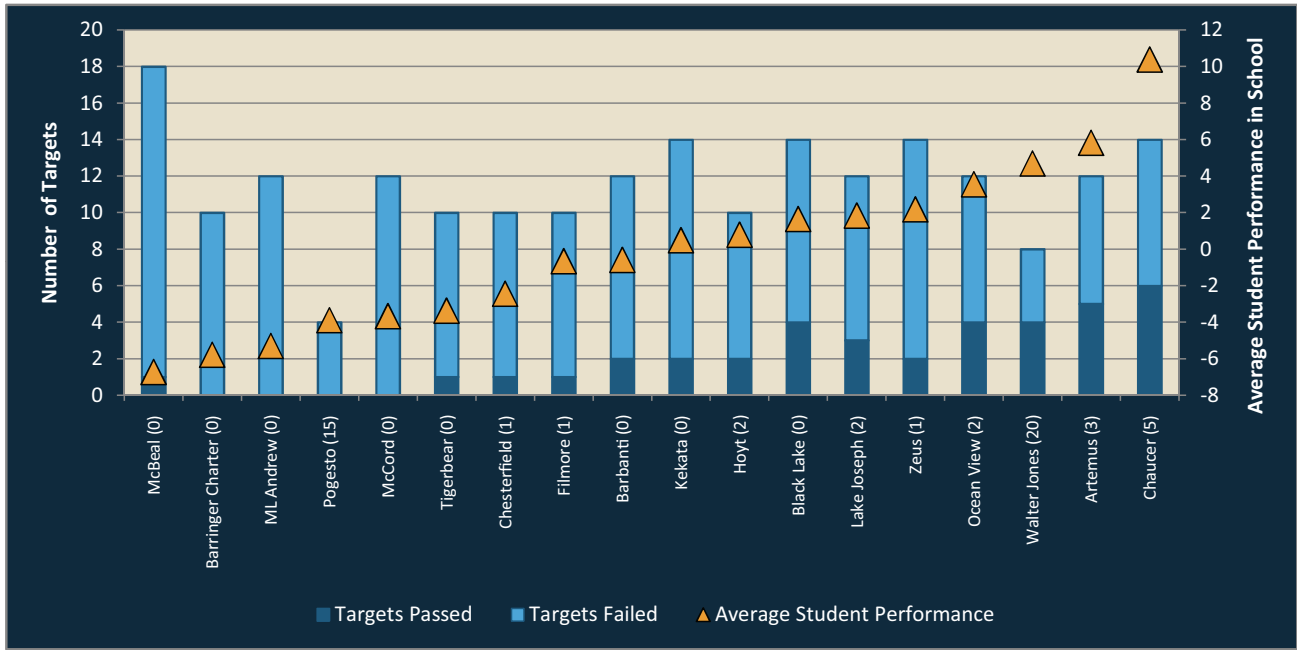


Figure 4. AYP performance of the middle school sample under Idaho's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under Idaho'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 fails to meet the AMOs for even a single subgroup didn't make AYP, so any light blue means that the school failed. Walter Jones, for example, met four 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 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

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 Idaho's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Idaho's 2008 AYP rules. **Only**

2 elementary schools made AYP while 16 failed to make it. The triangles in the figure 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 passing schools are in the right half of the figure, meaning that the highest performing students were found at these schools.

Yet, almost without regard to average student performance, the only schools that actually made AYP were those with relatively few qualifying subgroups—and thus the fewest targets to meet (because each subgroup has its own separate targets). Only Winchester and Roosevelt passed, and they had just four and six targets, respectively. Each had to make AYP for its overall student population in reading and math (two targets) and for its white population (two more targets); Roosevelt also had to make AYP for its low-income population (two targets).

Table 2. Elementary school subgroup performance of sample schools under the 2008 Idaho 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	51.1%	40.8%	N	N			N	N	N	N					N	N					8	0	0%	N	1	
Maryweather	57.1%	50.7%	N	N	N	N	N	N	N	N					N	N				Y	N	12	1	8%	N	1
Few	64.6%	51.9%	N	N	N	N	N	N	N	N					N	N					10	0	0%	N	1	
Nemo	66.0%	67.0%	N	N					N	N									Y	N	6	1	17%	N	7	
Island Grove	69.3%	66.7%	N	N					N	N					N	N				Y	N	8	1	13%	N	4
JFK	72.9%	60.4%	Y	N	N	N			N	N	N	N							Y	N	10	2	20%	N	3	
Scholls	81.7%	68.8%	Y	N	N	N			Y	N	N	N							Y	N	10	3	30%	N	7	
Hissmore	80.6%	72.1%	Y	N	N	N			Y	N	Y	N							Y	N	10	4	40%	N	7	
Wolf Creek	72.5%	67.6%	Y	N		N			N	N					N	N				Y	N	9	2	22%	N	5
Alice Mayberry	77.2%	75.1%	Y	N	N	N			Y	N	Y	N							Y	Y	10	5	50%	N	9	
Wayne Fine Arts	79.3%	81.0%	Y	Y					N	N									Y	Y	6	4	67%	N	21	
Winchester	78.8%	79.1%	Y	Y															Y	Y	4	4	100%	Y	22	
Coastal	82.2%	74.8%	Y	N	N	N	N	N	Y	N	N	N			N	N				Y	Y	14	4	29%	N	3
Paramount	81.0%	76.1%	Y	N					N	N					N	N				Y	Y	8	3	38%	N	7
Forest Lake	88.5%	84.4%	Y	Y	N	N			Y	N									Y	Y	8	5	63%	N	8	
Marigold	91.0%	85.6%	Y	Y	N	N			N	N									Y	Y	8	4	50%	N	10	
Roosevelt	93.6%	91.5%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	28	
King Richard	89.9%	88.8%	Y	Y	N	N			N	N									Y	Y	8	4	50%	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.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Idaho AYP rules. **None of the 18 schools in our sample passed**—even Walter Jones, the middle school that makes AYP in the greatest number of states (20) or the school with the highest performing students (Chaucer) didn't make AYP in Idaho.

Where Do Schools Fail?

Figure 3 illustrates how some elementary schools with middling performance can still make AYP when the

school has fewer targets to meet because it has fewer subgroups. Figures 3 and 4 do not, however, indicate which subgroups failed in which school. Information on individual subgroup performance appears in Tables 2 and 3 for elementary and middle schools, respectively.

Tables 2 and 3 show which subgroups qualified for evaluation at each school (i.e., whether the number of students within that subgroup exceeded the state's minimum *n*), and whether that subgroup passed or failed. Although all schools are evaluated on the proficiency rate of their overall population, potential sub-

Table 3. Middle school subgroup performance of sample schools under the 2008 Idaho 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	48.3%	52.8%	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	18	1	6%	N	0
Barringer Charter	53.3%	57.0%	N	N	N	N			N	N	N	N			N	N						10	0	0%	N	0
ML Andrew	48.5%	56.8%	N	N	N	N			N	N	N	N			N	N						12	0	0%	N	0
Pogesto	44.4%	61.1%	N	N																N	N	4	0	0%	N	15
McCord Charter	50.8%	60.3%	N	N	N	N			N	N	N	N			N	N						12	0	0%	N	0
Tigerbear	60.9%	55.7%	N	N	N	N			N	N	N	N								Y	N	10	1	10%	N	0
Chesterfield	62.5%	57.8%	N	N	N	N			N	N	N	N								Y	N	10	1	10%	N	1
Filmore	62.0%	66.4%	N	N	N	N			N	N					N	N				Y	N	10	1	10%	N	1
Barbanti	59.0%	61.7%	N	N	N	N	N	N	N	N					N	N				Y	Y	12	2	17%	N	0
Kekata	68.4%	66.1%	N	N	N	N	N	N	N	N	N	N			N	N				Y	Y	14	2	14%	N	0
Hoyt	69.2%	68.6%	N	N	N	N			N	N	N	N								Y	Y	10	2	20%	N	2
Black Lake	73.2%	68.8%	Y	N	N	N			N	N	N	N	Y	Y	N	N				Y	N	14	4	29%	N	0
Lake Joseph	70.1%	72.5%	Y	N	N	N	N	N	N	N					N	N				Y	Y	12	3	25%	N	2
Zeus	72.2%	71.6%	Y	N	N	N	N	N	N	N	N	N			N	N				Y	N	14	2	14%	N	1
Ocean View	74.3%	81.3%	Y	Y	N	N	N	N	N	N					N	N				Y	Y	12	4	33%	N	2
Walter Jones	82.0%	82.9%	Y	Y					N	N					N	N				Y	Y	8	4	50%	N	20
Artemus	81.0%	78.1%	Y	Y	N	N			N	N			Y	N	N	N				Y	Y	12	5	42%	N	3
Chaucer	83.2%	86.6%	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.

groups 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.

The school-by-school findings in Tables 2 and 3 show that:

- Most elementary schools met targets in math, but not in reading, for their overall student populations.
- Almost no subgroups at the elementary or middle

school level met math or reading targets, except for white youngsters in math.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. We see that every school with large enough populations of students with disabilities, LEP, Hispanic, or American Indian/Alaska Natives to qualify as separate subgroups failed to meet its reading and math targets for these students. In fact, the only subgroups where *any* schools in the sample met their targets in both reading and math were Asian/Pacific Islander, and white.

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 APY Idaho 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	11	10	11
Students with limited English proficiency	4	4	4
Low-income students	17	11	16
African-American students	5	3	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	7

Table 5. Summary of subgroup performance of sample middle schools under 2008 APY Idaho 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	17	17
African-American students	10	10	10
Asian/Pacific Islander students	4	1	2
Hispanic students	14	14	14
American Indian/Alaska Native students	1	1	1
White students	17	3	9

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Idaho's NCLB accountability system (at least in terms of the elements studied here) failed more schools than do most states in our sample. In fact, only two states (Massachusetts and Nevada) failed more elementary schools than Idaho. Similarly, Idaho is one of only five states (along

with Massachusetts, South Dakota, Montana, and North Dakota) that have zero passing middle schools in our sample (see Figure 1).

The most likely explanation for the difference in Idaho, relative to the other states examined, has to do with Idaho's more stringent AYP rules. Defining subgroups at 34 means that a school in Idaho will have more subgroups and consequently more chances to fail to make

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	2	16	0	18
Average student body size	225	315	n/a	859
Average % low income	13	50	n/a	45
Average % nonwhite	25	43	n/a	44
Average performance [†]	6.16	0.61	n/a	-0.05
Average % growth [‡]	121	114	n/a	98
Average number of targets to meet	5	9	n/a	12

[†] Student performance is measured by NWEA's MAP assessment and is expressed as an index of grade level normative performance. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance.

[‡] Average growth refers to improvement from fall to spring on the NWEA MAP assessments, averaged across all students within the school. Growth is expressed as an index value relative to NWEA norms and is scaled as a percentage. Thus, 100% means that students at the school are achieving normative levels of growth for their age and grade. Less than 100% growth means that the average student is increasing *by less* than normative amounts, while percentages over 100 mean that the average student is *exceeding* normative growth expectations.

AYP.⁸ Additionally, Idaho does not use a confidence interval as do most of the other states examined. This means that its schools have a more difficult time meeting their targets compared to states that use confidence intervals.

This is consistent with the patterns shown in Table 6, which compares schools making and not making AYP on several academic and demographic dimensions. Within the sample, schools that make AYP do indeed show higher average student performance, but they also differ in the following ways: they have smaller student populations, fewer subgroups (and thus fewer targets to meet), and lower percentages of low-income students.

Concluding Observations

This study examined the test performance data of students from 18 elementary and 18 middle schools across the country to see how these schools would fare under Idaho's AYP rules (and AMOs) for 2008. We found that only 2 elementary schools and no middle schools—2 in all from a sample of 36—would have made AYP in

Idaho. Looking across the 28 state accountability systems, this puts Idaho near the lower end of the sample distribution in terms of the number of schools making AYP (see Figure 1).

Several other factors are important to note for Idaho. First, Idaho's minimum subgroup size is relatively small in comparison to most other states examined in the study, meaning that schools in Idaho will be accountable for more subgroups than would similar schools in other states with higher subgroup sizes. Finally, Idaho, unlike most other states, does not apply confidence intervals so schools will have greater difficulty achieving their annual targets than equivalent schools in other states that report a confidence interval.

The overriding goal of the federal NCLB is to eliminate educational disparities within and across states; it is important to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, Idaho's NCLB accountability system is working exactly as Congress intended:

⁸ It is worth noting, however, that schools in Idaho are likely to be small and an *n* size of 34 probably makes sense.

identifying as needing attention those schools with relatively high test score averages that mask low performance for particular groups of students, such as low-income or Hispanic students. A moderate number of schools made annual targets in Idaho 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 pupils were not meeting state standards. Disaggregating data by race, income, and so on has made those students visible. That is surely a positive step.

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

much influence over making AYP? Does it make sense that having fewer subgroups enhances the likelihood of making AYP? Even if actual participation guidelines for English language learners and SWDs are more generous under the current state assessment system,⁹ doesn't the failure of these students to meet Idaho'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.

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.

⁹ See footnote 4.