Washington

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 Washington'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 Washington'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 Washington's system as well as under the systems 27 other states. We used school data and proficiency cut score[1] estimates from academic year 2005–2006, but applied them against Washington'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 Washington'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.

- It's also partly explained by Washington's somewhat smaller minimum *n* size for its race/ethnicity and low income subgroups, which means more of these

---

[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 Washington Assessment of Student Learning (WASL).

students are held separately accountable in Washington than they might be in other states. In addition, Washington has above average proficiency standards, especially at the middle school level, and relatively high annual targets, especially in grades 3-5 reading. Both these factors potentially hinder a school's chance of making AYP in Washington.

- Looking across the 28 state accountability systems examined in the study, **we find that the number of elementary schools that made AYP in Washington is exceeded in 20 other sample states (Washington ties with 5 other states that each has 2 schools that made AYP). In addition, Washington is one of 6 states with a single middle school making AYP (See Figure 1).**

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

Only two elementary schools and one middle school in our sample make AYP in 2008 under **Washington's** accountability system. One of the main reasons is that the state has a relatively small subgroup size for its minority and low-income students. This means that schools in Washington may have more accountable minority and low-income subgroups than would similar schools in other states. In addition, Washington has above average proficiency standards, especially at the middle school level, and relatively high annual targets, especially in grades 3-5 reading. Even though Washington's 99 percent confidence interval (i.e., statistical margin of error) provides schools with greater leniency than the more commonly used 95 percent confidence interval, these other factors make Washington among the most restrictive states in terms of the number of schools making AYP.

THOMAS B. FORDHAM INSTITUTE
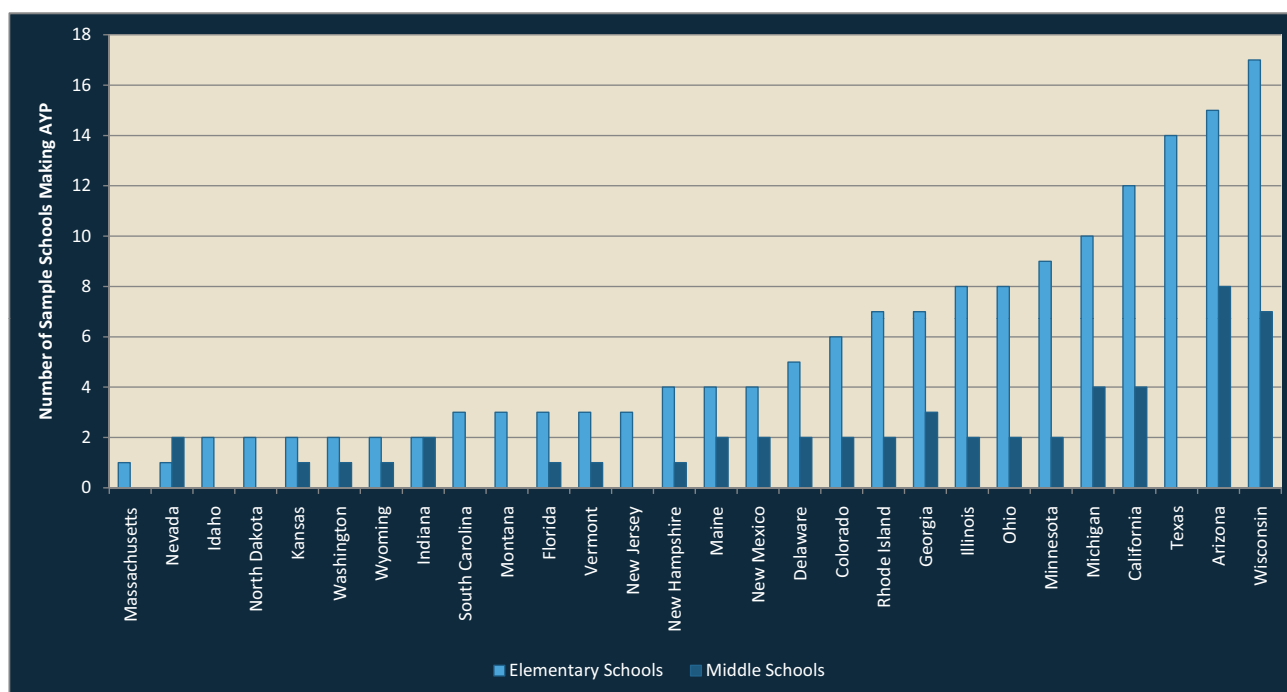
Figure 1. Number of sample schools making AYP by state

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

their overall populations[2] but failing because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.

■ In Washington, as in most states, schools with fewer subgroups attained 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.

■ As is the case in other states, middle schools have greater difficulty reaching AYP in Washington 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 whether or not a school would make AYP under Washington's system is whether it has enough English language learners to qualify as a separate subgroup. Every single school with a limited English proficient (LEP)[3] subgroup failed to make AYP, in part because these students did not meet the state's targets in reading and math. Likewise, every school with enough qualifying SWDs failed to make AYP.[4]

Introduction

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

---

[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.

[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 and SWDs may be slightly more negative than would be seen under real world conditions. This is mostly due to the differences in testing practices between how LEP students and students with disabilities are treated in the Washington Assessment of Student Learning (WASL) state assessment and in the NWEA's Measures of Academic Progress (MAP), the assessment used in this study. Specifically, the U.S. Department of Education has issued NCLB guidelines permitting schools to exclude small percentages of LEP or disabled students from taking state tests, or providing them alternate assessments. In the current study, however, no valid MAP scores were omitted from consideration.

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

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

NCLB requires every state, as a condition of receiving Title I funding, to implement an accountability system that aims to get 100% of its students to the proficient level on the state test by academic year 2013–2014. In the intervening years, states set annual measurable objectives (AMOs). This is the percentage of students in each school, and in each subgroup within the school (such as low income[5] 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, 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.[6]

---

[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.

Thomas B. Fordham Institute
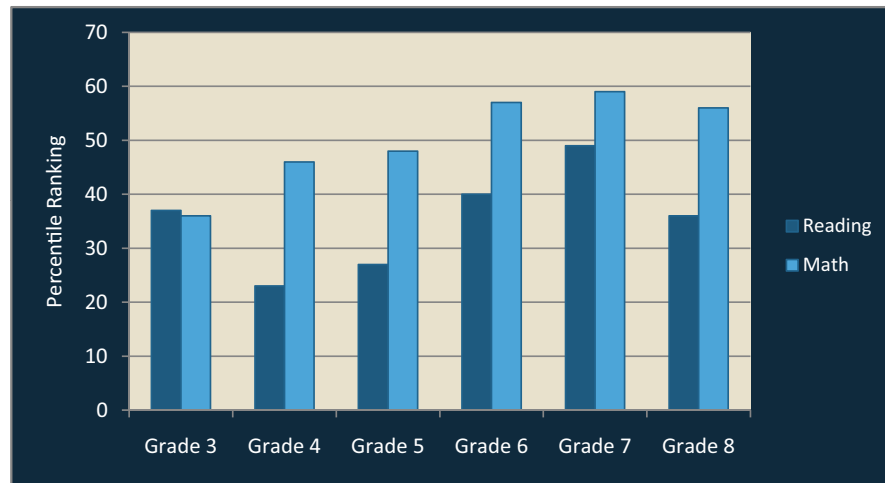
Figure 2. Washington reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Washington's cut scores (proficiency passing scores) for its reading and math tests, as percentiles of the NWEA norm, in grades 3 through 8. Higher percentile ranks are more difficult to achieve. All of Washington's cut scores are below the 60th percentile.

Proficiency cut score estimates for the Washington Assessment of Student Learning (WASL) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Washington's proficiency cut scores generally ranked above the average in difficulty, 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 Washington 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 Washington 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 Washington's 2008 AYP rules.)

Table 1 shows the pertinent Washington AYP rules that were applied to elementary and middle schools in the current study. Washington's minimum *n* sizes, unlike other states, vary according to subgroup. The subgroup size is 30 for the race/ethnicity and low-income subgroups and 40 for SWDs and students with limited English proficiency. For schools with 4000 or more students, the subgroup minimums for these last two groups is 1% of the school population. A subgroup size of 40 is typical, compared to other states in the study, but 30 is a bit lower.[7] This means that schools in Washington may have more accountable subgroups than would similar schools in other states.

Most states examined also apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates. However, Washington's 99% confidence interval provides schools with greater leniency than the more commonly used 95% confidence interval. This means even though the AMO might require a school to attain, for instance, 76.1% reading proficiency among its grade 3 students, and 76.1% reading proficiency among its grade 3 students in each subgroup, the real target can be lower, particularly with smaller groups.

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 by at least 10% the number of non-proficient students within any failing subgroup, relative to the previous year's performance. Because we had access to only a single 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.

---

[7]   School size and *n* size, however, are related (e.g., it makes sense for small schools to have small *n* sizes).

Table 1. Washington AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 30	
	SWDs: 40, or 1% if school population > 4000	
	Low-income students: 30	
	LEP students: 40, or 1% if school population > 4000	
CI	**Applied to proficiency rate calculations?**	
	Yes; 99% CI used	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	n/a	76.1
Grade 4	52.2	76.1
Grade 5	n/a	76.1
Grade 6	n/a	65.1
Grade 7	30.1	65.1
Grade 8	n/a	65.1
MATH		
Grade 3	n/a	64.9
Grade 4	29.7	64.9
Grade 5	n/a	64.9
Grade 6	n/a	58.7
Grade 7	17.3	58.7
Grade 8	n/a	58.7

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

Abbreviations: SWDs = students with disabilities; LEP = limited English proficiency; CI = confidence interval; AMOs = annual measurable objectives; n/a = not available

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

So, to reiterate, AYP decisions in the current study are modeled solely on test performance data for a single year. For each school, overall reading and math proficiency rates are calculated (along with any confidence intervals) to determine whether the overall school population and any qualifying subgroups achieved the annual measurable objectives. A school is deemed to have made AYP if the overall student body and all its qualifying subgroups met or exceeded its annual measurable objectives. Again, Appendix 1 supplies further methodological detail.

How Did the Sample Schools Fare Under Washington's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Washington's 2008 AYP rules. **Only 2 elementary schools out of 16 made AYP.** The triangles in Figure 3 show the average academic performance of students within the school, with negative values indicating below-grade-level performance, and positive values indicating above-grade-level performance. The schools that made AYP are in the right half of the figure,
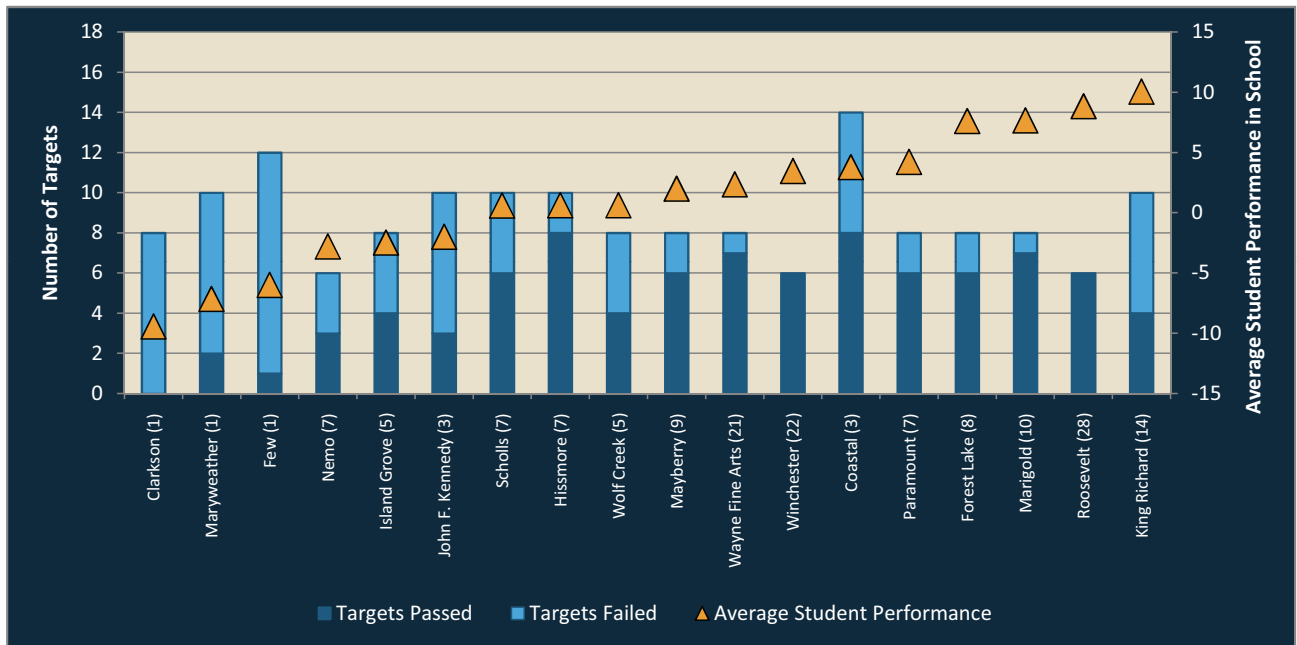
Figure 3. AYP performance of the elementary school sample under Washington's 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under Washington's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet in order to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not (light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMO for even a single subgroup didn't make AYP, so any light blue means the school failed. Marigold Elementary, for example, met seven of its eight targets, but because it did not 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.

meaning the higher performing students were found at these schools.

Yet almost without regard to average student performance, the only schools actually to make AYP were those with relatively few qualifying subgroups—and thus the fewest targets to meet (because each subgroup has separate targets). For example, Winchester and Roosevelt made AYP, but have only six targets each. Each had to meet two targets in reading and math for their overall student population, two more targets for their white subgroup, and two more targets for an additional subgroup—Hispanic for Winchester and low income for Roosevelt.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Washington AYP rules. **Out of eighteen in our sample, only one made AYP—** a high-performance school (Walter Jones), that 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 the application of the confidence interval. The orange lines show the annual measurable objective needed to meet AYP. These figures show that three of the sample elementary schools (Island Grove, JFK, and Wolf Creek) and one of the middle schools (Kekata) is assisted by the confidence intervals (note how the orange line falls within the light blue bar). However, we know from Figures 3 and 4 that all of these schools still fail to make AYP because of low subgroup performance.

The effect of confidence intervals on reading proficiency rates for elementary and middle schools is much the same (not shown). In reading, four elementary schools
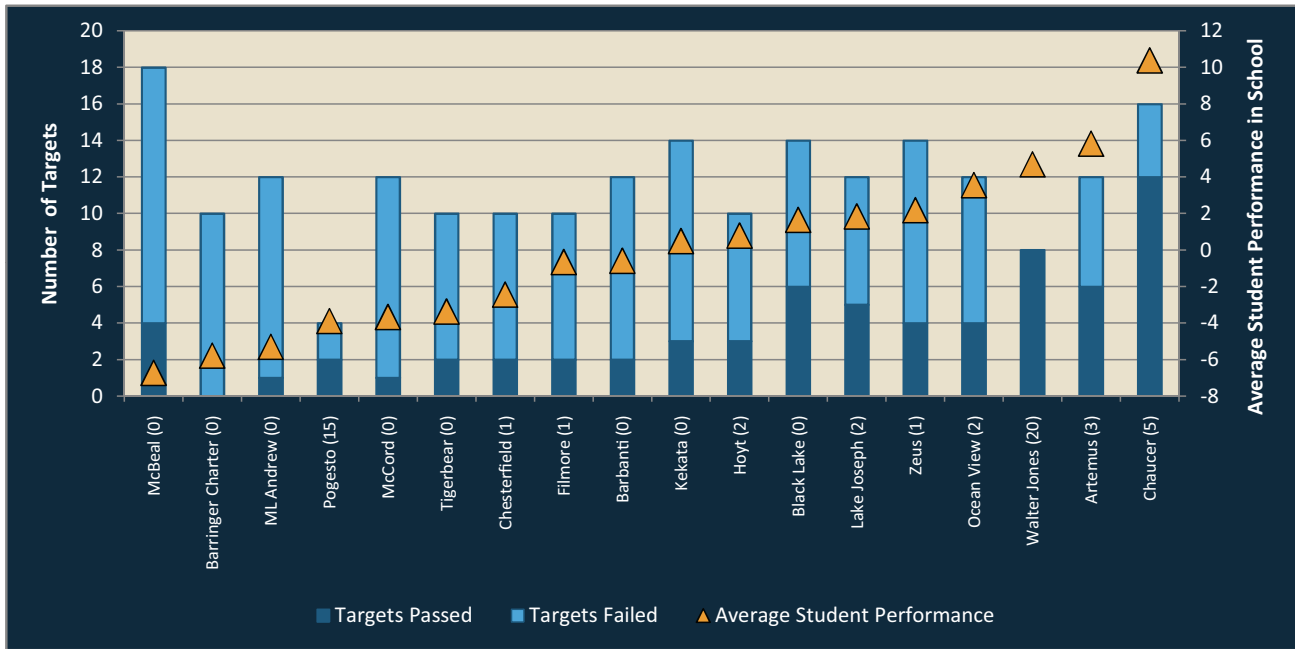
Figure 4. AYP performance of the middle school sample under Washington's 2008 AYP rules

Note: This figure indicates how each of the middle schools within the sample would have fared under Washington's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet in order to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not (light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMO for even a single subgroup didn't make AYP, so any light blue means the school failed. Pogesto, for example, meets two of its four targets, but because it did not 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 Washington'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 three of the sample elementary schools (Island Grove, JFK, and Wolf Creek) were assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.
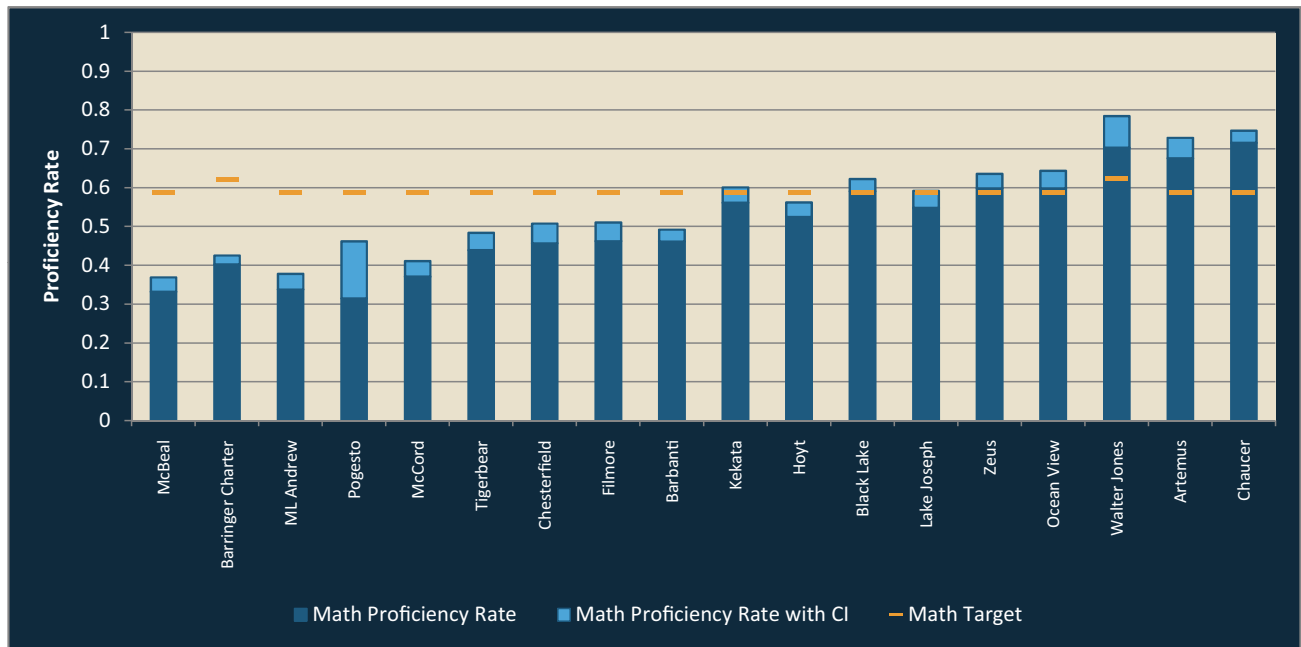
Figure 6. Impact of the confidence interval on middle school mathematics proficiency rates under Washington's 2008 AYP rules

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 (Kekata) 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.

(Nemo, Island Grove, Scholls and Wolf Creek) and two middle schools (Pogesto and Black Lake) are able to meet the overall target with the confidence interval, but still fail to meet their targets for all required subgroups. **In short, the application of the confidence interval has little effect on final AYP decisions for the sample schools in Washington, even though it does help some schools to meet their overall targets.**[8]

Where do schools fail?

Figures 3 and 4 illustrate how schools with low or middling performance can still pass AYP when the school has few targets to meet, thanks to fewer subgroups. However, these figures do not indicate which subgroups failed or passed in which school. Information on individual subgroup performance appears in Tables 2 and 3 for elementary and middle schools, respectively.

Tables 2 and 3 show which subgroups qualified for evaluation at each school (i.e., whether the number of students within that subgroup exceeded the state's minimum n), and whether that subgroup passed or failed. While all schools are evaluated on the proficiency rate of their overall population, potential subgroups that are separately evaluated for AYP purposes include SWDs, students with LEP, low-income students, and the following race/ethnic categories: African American (AA), Asian/Pacific Islander (Asian), Hispanic/Latino (Hispanic), American Indian/Alaska Native (AI/AN), and white. Tables 2 and 3 also show whether a school made AYP under the Washington rules, and the total number of states within the study in which the school made AYP.

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

- Three elementary schools (Clarkson, Maryweather,

---

[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 may be 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 subgroup performance of sample schools under the 2008 Washington 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	40.7%	43.7%	N	N			N	N	N	N					N	N					8	0	0%	N	1
Maryweather	47.5%	55.7%	N	N			N	N	N	N					N	N			Y	Y	10	2	20%	N	1
Few	53.3%	56.0%	N	N	N	N	N	N	N	N					N	N			Y	N	12	1	8%	N	1
Nemo	55.8%	68.8%	N	Y					N	N									Y	Y	6	3	50%	N	7
Island Grove	58.8%	71.2%	Y	Y					N	N					N	N			Y	Y	8	4	50%	N	4
JFK	62.1%	63.8%	Y	N	N	N			N	N	N	N							Y	Y	10	3	30%	N	3
Scholls	72.4%	72.1%	Y	Y	N	N			Y	Y	N	N							Y	Y	10	6	60%	N	7
Hissmore	70.7%	75.2%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Wolf Creek	63.3%	70.3%	Y	Y					N	N					N	N			Y	Y	8	4	50%	N	5
Alice Mayberry	69.3%	75.4%	Y	Y					Y	N	Y	N							Y	Y	8	6	75%	N	9
Wayne Fine Arts	67.2%	85.1%	Y	Y					Y	Y	N	Y							Y	Y	8	7	88%	N	21
Winchester	70.3%	80.6%	Y	Y											Y	Y			Y	Y	6	6	100%	Y	22
Coastal	73.6%	80.3%	Y	Y	N	N	N	N	Y	Y	N	N			Y	Y			Y	Y	14	8	57%	N	3
Paramount	77.3%	77.2%	Y	Y					Y	N					Y	N			Y	Y	8	6	75%	N	7
Forest Lake	85.0%	86.0%	Y	Y	N	N			Y	Y									Y	Y	8	6	75%	N	8
Marigold	85.3%	89.9%	Y	Y	Y	N			Y	Y									Y	Y	8	7	88%	N	10
Roosevelt	88.2%	92.5%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	28
King Richard	83.8%	91.5%	Y	Y	N	N			N	N					N	N			Y	Y	10	4	40%	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.

and Few) failed to meet both the reading targets and math targets for their overall school population.

- Three elementary schools (Hissmore, Marigold, and Forest Lake) met all targets except for their SWDs, and one school (Wayne Fine Arts) met all required targets except for its African American subgroup.

- Eight of the sample middle schools failed to meet both their reading and math targets for their overall population.

- None of the schools with qualifying SWD subgroups and LEP subgroups met AYP.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. First, the performance of SWDs and students with LEP are particularly challenging for Washington schools. Almost every single school with a large enough population of students in these groups to exceed the minimum *n* size failed to meet their subgroup targets. Nearly all of the traditionally academically-disadvantaged subgroups (e.g., low income and African American) also struggle under Washington's accountability system.

Table 3. Middle school subgroup performance of sample schools under the 2008 Washington 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	33.2%	47.5%	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	40.2%	56.1%	N	N	N	N			N	N	N	N			N	N					10	0	0%	N	0
ML Andrew	33.8%	51.1%	N	N	N	N			N	N	N	N			N	N			N	Y	12	1	8%	N	0
Pogesto	31.5%	53.7%	N	Y															N	Y	4	2	50%	N	15
McCord Charter	37.1%	54.8%	N	N	N	N			N	N	N	N			N	N			N	Y	12	1	8%	N	0
Tigerbear	43.9%	51.0%	N	N	N	N			N	N	N	N			N	N			Y	Y	10	2	20%	N	0
Chesterfield	45.7%	50.1%	N	N	N	N			N	N	N	N			N	N			Y	Y	10	2	20%	N	1
Filmore	46.2%	60.2%	N	N	N	N			N	N					N	N			Y	Y	10	2	20%	N	1
Barbanti	46.2%	55.9%	N	N	N	N	N	N	N	N					N	N			Y	Y	12	2	17%	N	0
Kekata	56.1%	60.1%	Y	N	N	N	N	N	N	N	N	N			N	N			Y	Y	14	3	21%	N	0
Hoyt	52.5%	62.4%	N	Y	N	N			N	N	N	N							Y	Y	10	3	30%	N	2
Black Lake	59.1%	64.1%	Y	Y	N	N					N	N	Y	Y	N	N			Y	Y	14	6	43%	N	0
Lake Joseph	54.8%	67.3%	Y	Y	N	N	N	N	N	Y					N	N			Y	Y	12	5	42%	N	2
Zeus	59.9%	66.2%	Y	Y	N	N	N	N	N	N	N	N			N	N			Y	Y	14	4	29%	N	1
Ocean View	59.8%	75.0%	Y	Y	N	N	N	N	N	N					N	N			Y	Y	12	4	33%	N	2
Walter Jones	70.3%	81.7%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	20
Artemus	67.6%	73.8%	Y	Y	N	N			N	N			Y	Y	N	N			Y	Y	12	6	50%	N	3
Chaucer	71.6%	82.3%	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.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Washington's NCLB accountability system is, in some respects, behaving like those in other states. For example, Roosevelt and Winchester are among the schools that make AYP in the greatest number of states—28 and 22, respectively. And these schools make AYP in Washington, too. Likewise, the elementary and middle schools that fail to make AYP in the greatest number of states also fail to make AYP in Washington.

But Washington is home to at least one anomaly. Consider Wayne Fine Arts Elementary (see Figure 3). It failed to make AYP in Washington, but makes AYP in 21 other states in our sample. Examining Table 2, one can see that Wayne Fine Arts failed to meet the minimum numbers for its LEP or SWD subgroups, which create difficulty for so many other schools within the sample. It did, however, miss the math target for its African American subgroup, possibly because of Washington's harder than average proficiency standards.

The differences between schools that did and didn't

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Washington 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	17	8	10
African-American students	6	4	4
Asian/Pacific Islander students	0	0	0
Hispanic students	9	6	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 Washington 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	15	14
African-American students	11	10	10
Asian/Pacific Islander students	4	0	0
Hispanic students	14	12	12
American Indian/Alaska Native students	1	1	1
White students	17	3	0

make AYP under Washington's accountability system can be seen in Table 6, which compares them on a number of 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 much smaller student populations, fewer subgroups (and thus fewer targets to meet), and much lower percent-

ages of low income students.

The picture for middle schools is similar. The one middle school that made AYP had slightly higher performing students, on average, than middle schools that didn't, as well as a drastically smaller enrollment, a smaller nonwhite population, and fewer subgroups (and thus targets to meet).

Thomas B. Fordham Institute

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	2	16	1	17
Average student body size	225	315	165	900
Average % low income	13	50	38	45
Average % nonwhite	25	43	33	45
Average performance†	6.16	0.61	4.69	-0.33
Average % growth‡	121	114	111	97
Average number of targets to meet	6	9	8	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.

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 Washington's AYP rules and annual measurable objectives for 2008. We found that only two elementary schools and one middle school—three in all from a total of 36—would have made AYP in Washington. Looking across the 28 state accountability systems examined in the study, this puts Washington at the low end of the distribution in terms of the number of schools making AYP (as shown in Figure 1). This high failure rate is partly explained by Washington's somewhat smaller minimum *n* size for its race/ethnicity and low income subgroups, which means more of these students are held separately accountable in Washington than they might be in other states. In addition, Washington has above average proficiency standards, especially at the middle school level, and relatively high annual targets, especially in reading for grades 3 through 5. All of these factors potentially hinder a school's chance of making AYP in The Evergreen State.

The overriding goal of the No Child Left Behind act

(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, Washington's No Child Left Behind 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 schools met the Washington math and reading targets for their student populations as a whole (i.e., without considering subgroup results). 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, etc. has made those students visible. That is surely a good thing.

Yet NCLB's design flaws are also readily apparent. Does it make sense that 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 Washington's targets (especially at the middle school level where more of them qualify) 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.

THOMAS B. FORDHAM INSTITUTE