Arizona

Executive Summary

The intent of the No Child Left Behind (NCLB) Act of 2001 is to hold schools accountable for ensuring that all their students achieve mastery in reading and math, with a particular focus on groups that have traditionally been left behind. Under NCLB, states submit accountability plans to the U.S. Department of Education detailing the rules and policies to be used in tracking the adequate yearly progress (AYP) of schools toward these goals.

This report examines Arizona'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 Arizona'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 Arizona's system as well as under the systems of 27 other states. We used school data and proficiency cut score[1] estimates from academic year 2005–2006, but applied them against Arizona's AYP rules for the academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- We estimate that **3 of 18 elementary schools** and **10 of 18 middle schools** in our sample **failed to make AYP** in 2008 under Arizona's accountability system. Among the 28 accountability systems examined in the study, there's only one state where more schools make AYP than in Arizona (Wisconsin). This makes The Grand Canyon State one of the least restrictive in terms of AYP passage rates (see Figure 1.)[2]

- Several sample schools made AYP in Arizona that failed to make AYP in most other states. This is probably because **Arizona's proficiency standards are relatively easy compared to other states (especially in reading)**. Another reason is that Arizona's definitions for subgroups are grade-based rather than school based, resulting in fewer accountable subgroups (i.e., a school must have at least 40 individuals within a grade for that group to be evaluated). Arizona also uses a very generous confidence interval (or margin of error).

Arizona has several unique characteristics which contribute to the large number of schools making AYP in the state. In fact, only one other state in the study (Wisconsin) deems that more schools make AYP than Arizona does. One of the factors contributing to this is the rule set governing subgroup size. Unlike most states, Arizona considers each grade separately when determining whether a subgroup meets the criteria for accountability, which (for Arizona) is at least 40 students. For instance, a middle school in Arizona with three grades could have *almost* 120 African-American students, all performing poorly, and still make AYP as long as there are fewer than 40 African-American children in each grade. Another factor contributing to the high number of schools making AYP is Arizona's 99 percent confidence interval (i.e., statistical margin of error). This provides schools with greater leniency than the 95 percent confidence interval used by most other states in the study. Finally, Arizona's proficiency standards (or cut scores) are relatively easy in the early grades, compared to other states. In fact, in grades 3-5, the reading cut score is in the 25th percentile range.

---

[1] A cut score is the minimum score a student must receive on the Arizona's Instrument to Measure Standards (AIMS) in order to be considered proficient under Arizona's accountability system.

[2] Note that Arizona received full approval from the U.S. Department of Education to implement a student growth model for the 2006-2007 school year. The current analysis, which draws on data from 2005–2006, does not in any way use or incorporate student growth model calculations.
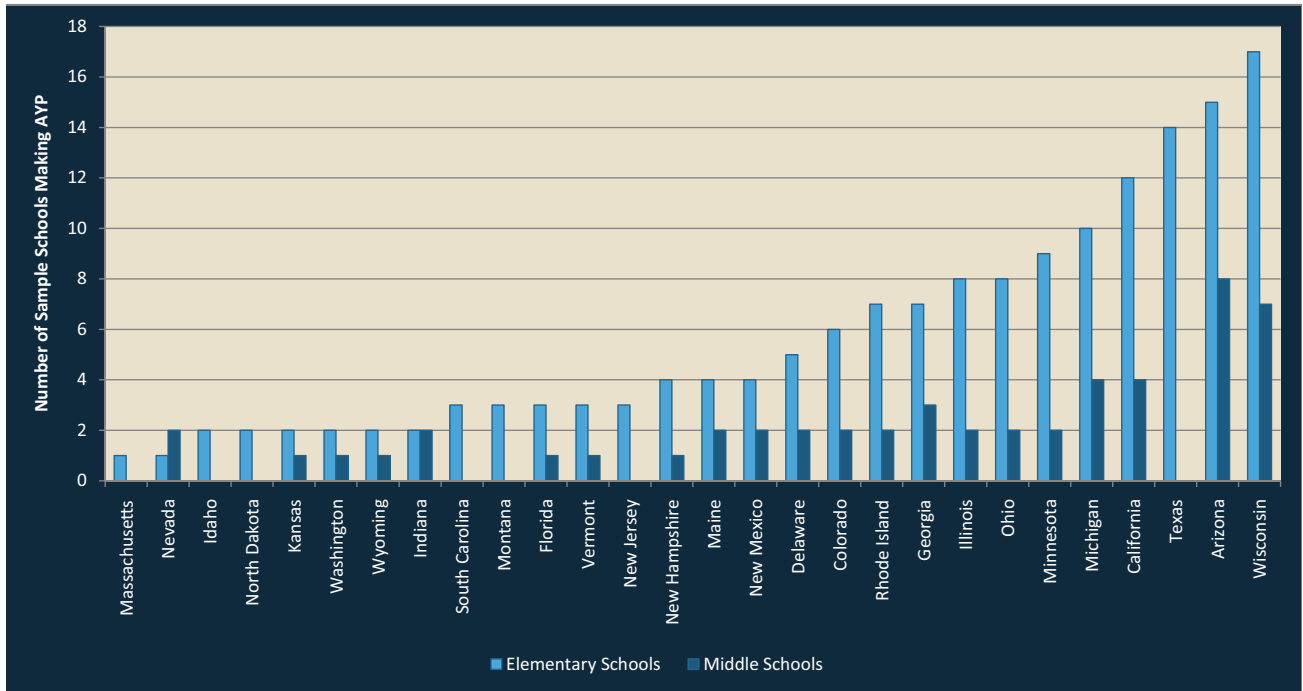
THOMAS B. FORDHAM INSTITUTE

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.

■ Nearly all of the schools in our sample that failed to make AYP in Arizona are meeting expected targets for their overall populations, but failing because of the performance of individual subgroups—particularly students with disabilities (SWDs) at the middle school level.[3]

■ In Arizona, as in most states, schools with fewer subgroups attain AYP more easily than schools with more subgroups, even when their average student performance is lower. In other words, schools with greater diversity and size face greater challenges in making AYP.

■ As in other states, middle schools have greater difficulty reaching AYP in Arizona 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.[4]

■ A strong predictor of a school making AYP under Arizona's system is whether it has enough SWDs to qualify as a separate subgroup. In cases where there were enough students to constitute a separate SWD subgroup, every school with one failed to make AYP.

Introduction

The Proficiency Illusion (Cronin, et al. 2007a) linked student performance on Arizona's test and those of 25 other states to the Northwest Evaluation Association's (NWEA) Measures of Academic Progress (MAP), a computerized adaptive test used in schools nationwide. This

---

[3] SWDs are defined as those students following individualized education plans. We should also note that our subgroup findings for Limited English proficient (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 Arizona's Instrument to Measure Standards (AIMS), 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.

[4] 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.

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 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,[6] among other subgroups.

The schools were not selected as a representative sample of the nation's population. Instead, we selected the schools because they exhibited a range of characteristics on measures such as academic performance, academic growth, and socioeconomic status (the latter calculated by the percentage of students receiving free or reduced-price lunches). Appendix 1 contains a complete discussion of the methodology for this project along with the characteristics of the school sample.[7]

---

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

[6] Note that we use "LEP students" and "English language learners" interchangeably to refer to students in the same subgroup.

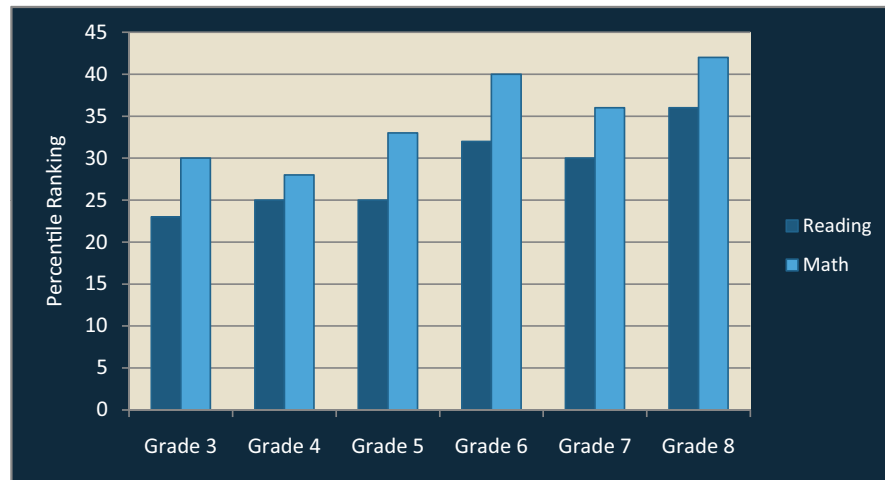[7] We gave all schools in our sample pseudonyms in this report.

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

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

Proficiency cut score estimates for Arizona's Instrument to Measure Standards (AIMS) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Arizona's definitions of proficiency in reading and math were below-average to average in terms of difficulty, compared to the other states in the study. These cut scores were used to estimate whether students would have scored as proficient or better on the Arizona 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 Arizona 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 Arizona's 2008 AYP rules.

Table 1 shows the pertinent Arizona AYP rules that were applied to elementary and middle schools in this study. Arizona's minimum subgroup size is 40, which is comparable to most other states we examined.[8] However, the size is grade-based, meaning a school must have at least 40 individuals within a grade for that subgroup to be evaluated. Annual targets also change according to grade and subject area. The annual target for grade 3 reading, for example, is 62% of students reaching proficiency; that number changes to 38% for grade 8 math.

Furthermore, although most states apply confidence intervals (or margins of statistical error) to their measurement of student proficiency rates, Arizona's 99% confidence interval gives schools greater leniency than the 95% confidence interval used by most other states. So, for instance, although schools are supposed to get 38% of their eighth grade students to the proficient level on the state math test—and 38% of their students in each subgroup—applying the confidence interval means that the real target can actually 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 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

---

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

Table 1. Arizona AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 40	
	SWDs: 40	
	Low-income students: 40	
	LEP students: 40	
CI	**Applied to proficiency rate calculations?**	
	Yes; 99% CI used	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	44.0	62.6
Grade 4	45.0	56.0
Grade 5	32.0	54.6
Grade 6	45.0	56.0
Grade 7	49.0	59.2
Grade 8	31.0	54.0
MATH		
Grade 3	32.0	54.6
Grade 4	54.0	63.2
Grade 5	20.0	46.6
Grade 6	43.0	54.4
Grade 7	48.0	58.4
Grade 8	7.0	38.0

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

schools. In addition, federal law requires 95% of each school's students—and 95% of the students in each school's subgroup—to participate in testing.

To reiterate, then, AYP decisions in the current study are modeled solely on test performance data for a single academic year. For each school, we calculated reading and math proficiency rates (along with any confidence intervals) to determine whether the overall school population 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 Arizona's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Arizona's 2008 AYP rules. **Only 3 of the 18 elementary schools failed to make AYP under the Arizona rules.** The triangles in Figure 3 show the average academic performance of students within the school, with negative values indicating below-grade-level performance for the average student, and positive values indicating above-grade-level performance. The two schools with lowest average student performance (Clarkson and Maryweather) both fail to make AYP, as does one of the schools with higher average student
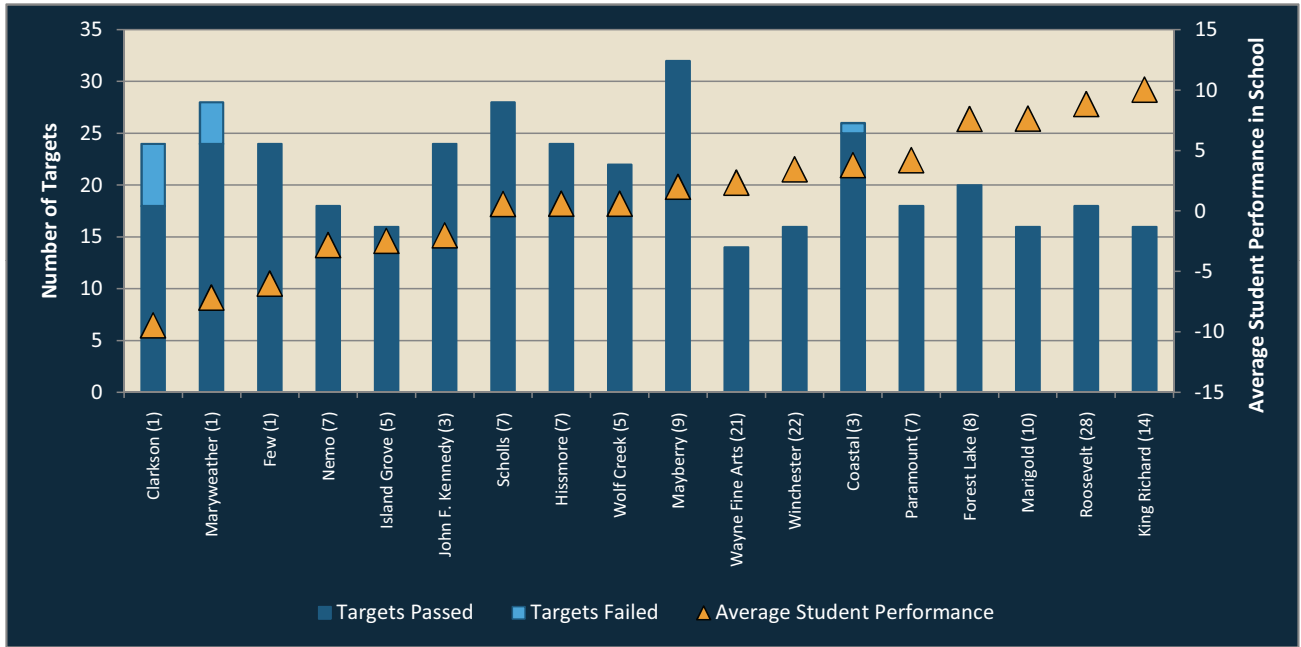
Figure 3. AYP Performance of the elementary school sample under Arizona's 2008 AYP rules

Note: This figure indicates how each elementary school within the sample fared under Arizona'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 the school failed. Coastal Elementary, for example, met 25 of its 26 targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles) which is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states, out of 28, in which that school would have made AYP.

performance (Coastal). All three schools that failed to make it, however, have between 24 and 28 targets to meet, as opposed to the schools that made AYP, which have, on average, only 20 targets to meet.[9]

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Arizona AYP rules. **Out of 18 middle schools in our sample, 8 made AYP –** three low-performance schools (Pogesto, Chesterfield, and Filmore), and five high-performance schools (Lake Joseph, Ocean View, Walter Jones, Artemus, and Chaucer). As with the sample elementary schools, schools that made AYP tended to have fewer targets to meet than schools that didn't make AYP.

Figure 5 indicates the degree to which elementary schools'

math proficiency rates are aided by the confidence interval. On this figure, the darker portions of the bars show the actual proficiency rates at each school, and the lighter portions of the 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. The figure shows that none of the sample elementary schools was assisted by the confidence intervals, because the math targets in Arizona are low relative to the schools' overall performance. Although not shown, this same trend held true for middle school math and reading proficiency rates at the middle and elementary school levels as well. **Because of the relatively easy targets established by Arizona's annual measurable objectives, confidence intervals have little impact on whether schools make AYP.**[10]

---

[9] Recall that Arizona has more targets because each grade level is considered a group unto itself. For instance, a middle school in Arizona with three grades and four subgroups has 3 × 4 × 2 (subjects) or 24 targets.

[10] 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 Figure 5. However, we chose not to show how the confidence interval impacted subgroup performance because it would have added greatly to this report's length and complexity.
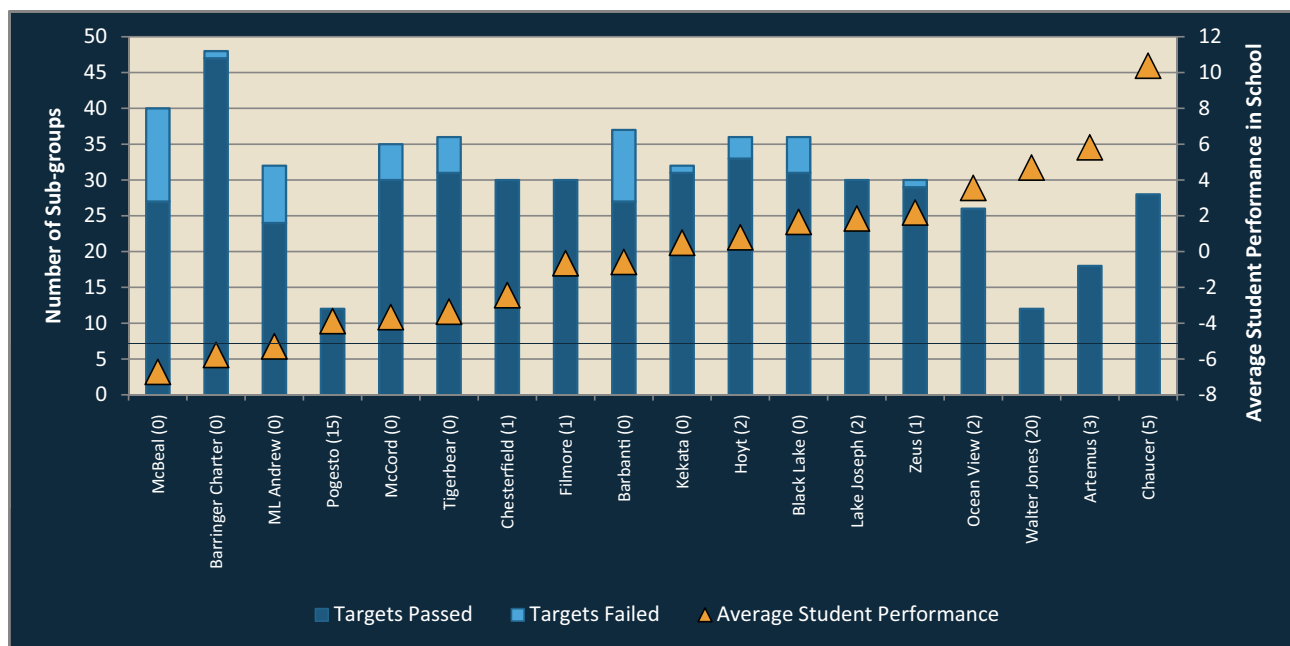
Figure 4. AYP performance of the middle school sample under Arizona's 2008 AYP rules

Note: This figure shows how each middle school would have faired under Arizona'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 AMO for even a single subgroup did not make AYP, so any light blue means the school failed. Zeus Middle School, for example, met 29 of its 30 targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles) which is measured by average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states, out of 28, in which that school would make AYP.
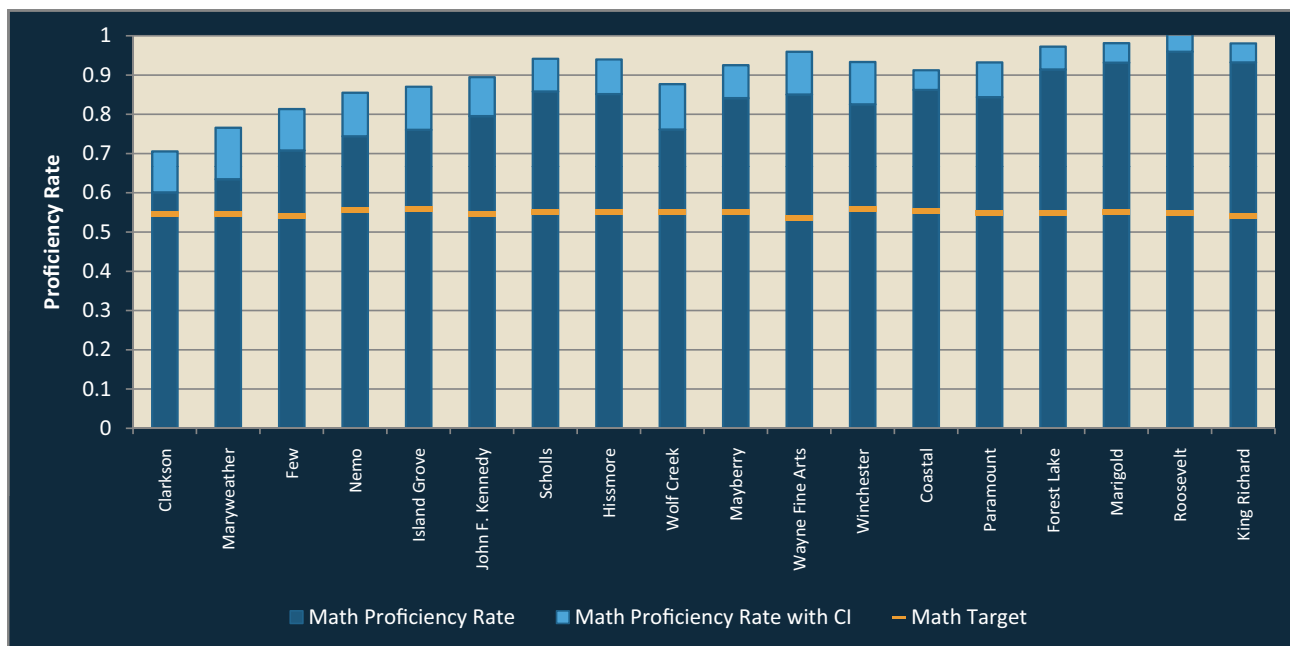
Figure 5. Impact of the confidence interval on elementary school math proficiency rates

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that none of the sample elementary schools was assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.

Table 2. Elementary school subgroup performance of sample schools under the 2008 Arizona AYP rules

SCHOOL PSEUDONYM	Overall Proficiency Rate Math	Reading	Overall M	R	SWDs M	R	LEP Students M	R	Low-income Students M	R	AA M	R	Asian M	R	Hispanic M	R	AI/AN M	R	White M	R	AYP Targets Required	Targets MET	% of Targets Met	School Met AYP?	Number of states in which school met AYP?
Clarkson	70.6%	58.1%	Y	N					Y	N					Y	N					24	18	75%	N	1
Maryweather	76.6%	68.2%	Y	Y			Y	N	Y	Y					Y	N					28	24	86%	N	1
Few	81.3%	70.6%	Y	Y					Y	Y					Y	Y					24	24	100%	Y	1
Nemo	85.5%	85.4%	Y	Y															Y	Y	18	18	100%	Y	7
Island Grove	87.0%	83.1%	Y	Y															Y	Y	16	16	100%	Y	5
JFK	89.5%	78.7%	Y	Y					Y	Y									Y	Y	24	24	100%	Y	3
Scholls	94.2%	84.7%	Y	Y			Y	Y	Y	Y									Y	Y	28	28	100%	Y	7
Hissmore	94.0%	86.7%	Y	Y					Y	Y									Y	Y	24	24	100%	Y	7
Wolf Creek	87.7%	85.3%	Y	Y			Y	Y											Y	Y	22	22	100%	Y	5
Alice Mayberry	92.5%	88.7%	Y	Y			Y	Y	Y	Y	Y	Y							Y	Y	32	32	100%	Y	9
Wayne Fine Arts	95.9%	96.4%	Y	Y															Y	Y	14	14	100%	Y	21
Winchester	93.3%	94.2%	Y	Y															Y	Y	16	16	100%	Y	22
Coastal	91.2%	85.3%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	26	25	96%	N	3
Paramount	93.2%	88.6%	Y	Y															Y	Y	18	18	100%	Y	7
Forest Lake	97.3%	94.7%	Y	Y					Y	Y									Y	Y	20	20	100%	Y	8
Marigold	98.1%	94.7%	Y	Y															Y	Y	16	16	100%	Y	10
Roosevelt	100.4%	99.7%	Y	Y															Y	Y	18	18	100%	Y	28
King Richard	98.1%	96.3%	Y	Y															Y	Y	16	16	100%	Y	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. Unlike most states, Arizona schools consider each grade separately when determining whether the minimum *n* size is exceeded for a particular subgroup. This means that Arizona schools may be required to meet up to 18 targets for each grade (2 targets each–math and reading–for the overall population, SWDs, LEP, low income, African American, Asian, Hispanic, American Indian, and white). This is, of course, provided that there are sufficient numbers of students within the grade to exceed the state's minimum *n* size of 40 in every subgroup. (In actuality, it's much harder to exceed the minimum *n* size when individual grade levels are considered versus the school as a whole.) In this table, for example, we see that Clarkson Elementary met the minimum *n* size for its overall, Hispanic, and low income subgroups. However, to preserve space, each grade is not displayed separately. Consequently, the number of AYP targets required at Clarkson (24) and the number of targets met (18), let us know that the school failed to meet all of its required subgroup targets, but we don't know in which grades.

Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low average student performance can still make AYP when the school has relatively few targets to meet because it has fewer subgroups. These figures do not, however, indicate which subgroups failed or passed in which school. Tables 2 and 3 list information on individual subgroup for ele-

mentary 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 Arizona 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	62.9%	66.0%	Y	Y	N	N	N	N	N	N					N	N			Y	Y	40	27	68%	N	0
Barringer Charter	66.9%	69.4%	Y	Y					Y	Y	Y	N			Y	Y					48	47	98%	N	0
ML Andrew	63.9%	71.6%	Y	Y	N	N			N	Y	N	N			Y	Y			Y	Y	32	24	75%	N	0
Pogesto	77.7%	92.1%	Y	Y																	12	12	100%	Y	15
McCord Charter	65.8%	72.9%	Y	Y			Y	Y	N	N	N	Y			Y	Y			Y	Y	35	30	86%	N	0
Tigerbear	73.2%	71.5%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	36	31	86%	N	0
Chesterfield	78.4%	75.1%	Y	Y					Y	Y	Y	Y							Y	Y	30	30	100%	Y	1
Filmore	76.4%	82.2%	Y	Y					Y	Y					Y	Y			Y	Y	30	30	100%	Y	1
Barbanti	69.6%	75.0%	Y	Y	N	N			N	N	N				Y	Y			Y	Y	37	27	73%	N	0
Kekata	80.4%	77.9%	Y	Y	N	Y			Y	Y	Y	Y							Y	Y	32	31	97%	N	0
Hoyt	81.7%	80.9%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	36	33	92%	N	2
Black Lake	83.5%	80.3%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	36	31	86%	N	0
Lake Joseph	82.1%	86.5%	Y	Y					Y	Y					Y	Y			Y	Y	30	30	100%	Y	2
Zeus	83.7%	82.2%	Y	Y			N		Y	Y									Y	Y	30	29	97%	N	1
Ocean View	86.4%	91.4%	Y	Y					Y	Y					Y	Y			Y	Y	26	26	100%	Y	2
Walter Jones	100.0%	99.9%	Y	Y																	12	12	100%	Y	20
Artemus	90.3%	92.5%	Y	Y					Y	Y									Y	Y	18	18	100%	Y	3
Chaucer	91.4%	93.1%	Y	Y					Y				Y	Y	Y	Y			Y	Y	28	28	100%	Y	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. Unlike most states, Arizona schools consider each grade separately when determining whether the minimum *n* size is exceeded for a particular subgroup. This means that Arizona schools may be required to meet up to 18 targets for each grade (2 targets each–math and reading–for the overall population, SWDs, LEP, low income, African American, Asian, Hispanic, American Indian, and white). This is, of course, provided that there are sufficient numbers of students within the grade to exceed the state's minimum *n* size of 40 in every subgroup. (In actuality, it's much harder to exceed the minimum *n* size when individual grade levels are considered versus the school as a whole.) In this table, for example, we see that Barringer Charter met the minimum *n* size for its overall, African American, Hispanic, and low income subgroups. However, to preserve space, each grade is not displayed separately. Consequently, the number of AYP targets required at Barringer Charter (48) and the number of targets met (47), let us know that the school failed to meet all of its required subgroup targets, but we don't know in which grades.

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. Tables 2 and 3 also show whether a school met AYP under the 2008 Arizona 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:

- No elementary schools failed to meet their overall targets for math.

- One elementary school (Clarkson) failed to meet the overall target for reading.

- All middle schools met overall targets for reading and math.

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Arizona 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	1	0	1
Students with limited English proficiency	4	0	1
Low-income students	9	0	1
African-American students	2	0	0
Asian/Pacific Islander students	0	0	0
Hispanic students	3	0	2
American Indian/Alaska Native students	0	0	0
White students	15	0	0

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Arizona 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	7
Students with limited English proficiency	3	1	2
Low-income students	16	4	3
African-American students	8	2	2
Asian/Pacific Islander students	1	0	0
Hispanic students	9	1	1
American Indian/Alaska Native students	0	0	0
White students	15	0	0

- One elementary school (Coastal) met every target except for the reading target for its SWDs.

- Five middle schools (Tigerbear, Kekata, Hoyt, Black Lake, and Zeus) met all targets except for SWDs.

- One middle school (Barringer Charter) met every target except for one ethnic minority group.

Tables 4 and 5 summarize subgroup performance for ele-mentary and middle schools, respectively. As shown, the performance of SWDs is proving most challenging for schools under Arizona's system, particularly in middle schools, where this subgroup tends to have enough stu-dents to meet the state's minimum n of 40. In fact, every school within the sample with qualifying SWDs failed to make AYP. (However, it's well worth noting that only one school met the minimum *n* size for SWD subgroups at the elementary level.)

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	15	3	8	10
Average student body size	299	333	587	1077
Average % low income	41	75	34	54
Average % nonwhite	34	72	43	45
Average performance†	2.32	-4.26	2.41	-2.03
Average % growth‡	118	100	106	92
Average number of targets to meet	20	26	23	36

† 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.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Arizona's NCLB accountability system is, in some respects, behaving similarly to those in other states. All the sample schools that fail under Arizona rules failed in most of the other states examined in this study. For example, among the elementary schools in our sample, Clarkson and Maryweather both failed in Arizona (Figure 3), and these two schools failed in all but one of the 28 states examined in this study. Likewise, all the failing middle schools in Figure 4 also failed in the majority of the other states examined in the study.

However, on the whole, Arizona's AYP rules are generally more lenient than in other states. Many sample elementary schools (e.g., Few, Island Grove, and JFK) and middle schools (e.g., Chesterfield and Filmore) that failed to make AYP in most other states make it in Arizona. This is most likely attributable to Arizona's minimum subgroup policy, which considers grades separately, meaning that an Arizona school will have fewer accountable subgroups than a similar school in another state. Arizona's subgroup policies,

along with relatively easy annual targets relative to student performance, mean that schools made AYP more easily in Arizona than in many other states.

Despite its greater leniency, the rule set in Arizona showed certain trends that were similar for other states as well. Schools that made AYP in Arizona tended to have higher average student performance than schools that didn't, though schools with more targets to meet tended not to do as well as schools with fewer targets.

This is illustrated in Table 6, which compares schools that did and didn't make AYP on a number of academic and demographic dimensions in Arizona. 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, particularly in middle schools, fewer subgroups (and thus fewer targets to meet), and lower percentages of low income students.

Concluding Observations

This study evaluated the test performance data of students from 18 elementary and 18 middle schools across

the country to see how these schools would fare under Arizona's AYP rules (and AMOs) for 2008. We found that 15 elementary schools and 8 middle schools—23 in all, from a sample of 36—would have made AYP in Arizona. Compared to the other 27 states examined, this places Arizona at the high end of the distribution in terms of the number of schools making AYP (see Figure 1). In addition, some sample schools make AYP in Arizona that fail to make AYP in most other states. This is most likely because Arizona's proficiency standards are relatively easy compared to other states and its particular rules result in fewer accountable subgroups.

Because the overriding goal of the federal 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, Arizona'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. All the sample schools, save one, make AYP in Arizona for their student populations as a whole (i.e., without considering sub-group 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 weren't meeting state standards. Disaggregating data by race, income, and so on. has made those students visible. That is surely a positive step.

Yet NCLB's design flaws are also readily apparent. Does it make sense that having fewer subgroups enhances the likelihood of making AYP? Is it "fair" for a state to have such generous margins of error and low elementary school cut scores? Does it make sense that the size of a school's enrollment has so much influence over making AYP? 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.

THOMAS B. FORDHAM INSTITUTE