Ohio

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 Ohio'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 Ohio'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 Ohio'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 Ohio's AYP rules for academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- We estimate that **10 of 18 elementary schools** and **16 of 18 middle schools** in our sample **failed to make AYP** in 2008 under Ohio's accountability system. (This rate is partly explained by our sample, which intentionally includes some schools with relatively large populations of low-performing students.)
- **Looking across the 28 state accountability systems examined in the study, we find that the number of elementary schools that made AYP in Ohio was exceeded in just 6 other sample states** (Ohio and Illinois tie with 8 elementary schools making AYP) (see Figure 1).[2]
- Nearly all of the schools in our sample that failed to make AYP in Ohio are meeting expected targets for their overall populations[3] but failing because of the performance of individual subgroups, particularly students with disabilities (SWDs)[4] and English language learners.
- A few sample schools that made AYP in Ohio failed to make AYP in most other states. **This is most likely because Ohio's proficiency standards are relatively easy compared to other states, and Ohio's minimum *n* (number of students in sample) size for SWDs is higher than other states, meaning that**

Ohio falls in the upper end of the state distribution in terms of the number of schools that make AYP. In fact, a few sample schools make AYP in Ohio that fail to make AYP in most other states. This is likely because Ohio's proficiency standards are relatively easy compared to other states (most of Ohio's cut scores are below the 35th percentile). Additionally, while Ohio's minimum *n* size for most of its subgroups is a little lower than in other states (30), the state raises its subgroup size to 45 for students with disabilities, meaning fewer of these students are held separately accountable than in other jurisdictions. On the other hand, Ohio does not apply confidence intervals (or margins of error) to their measurements of student proficiency rates. This means that schools in Ohio will have a more difficult time meeting their targets than schools in states that do use confidence intervals.

[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 Ohio Achievement Test.

[2] In 2006, Ohio received approval from the U.S. Department of Education to use a student growth model in its state accountability plan. The data in this study are drawn from 2005–2006 and do not reflect student growth calculations in any way.

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

[4] SWDs are defined as those students following individualized education plans.

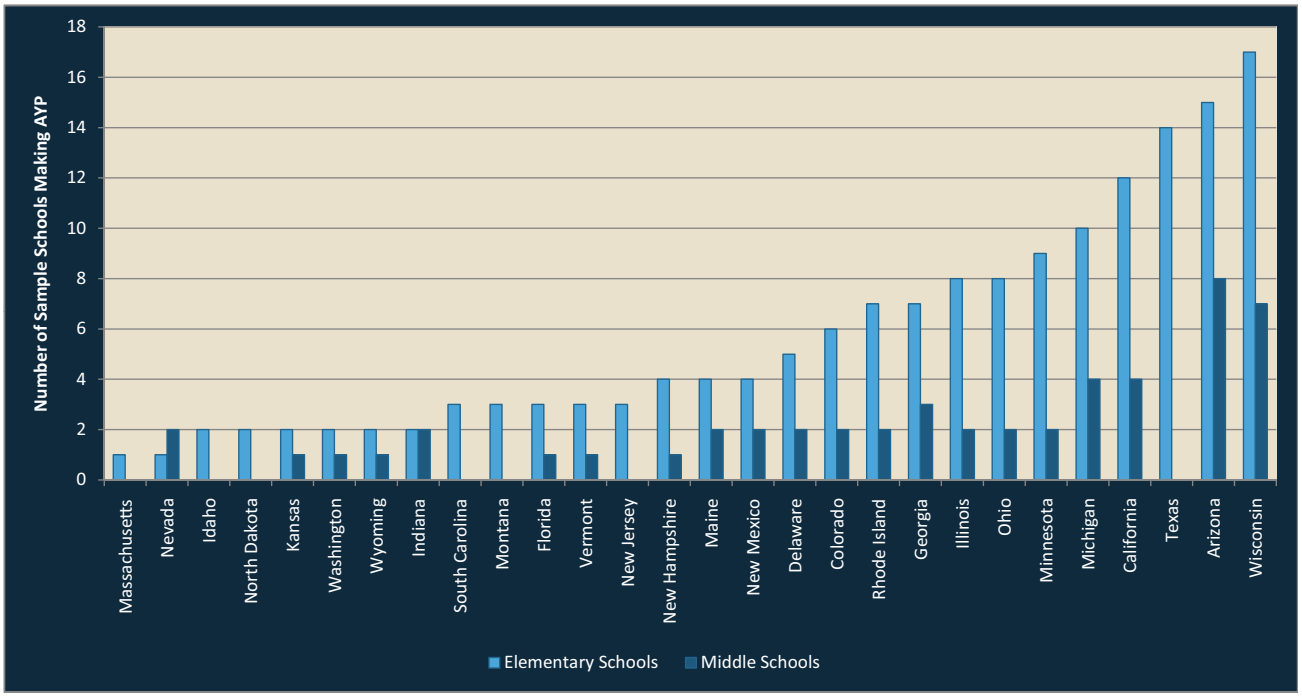

Figure 1. Number of sample schools making AYP by state

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.

fewer SWD subgroups in Ohio (especially at the elementary level) are likely to be held separately accountable for performance.

- As in other states, schools with fewer subgroups attained AYP more easily in Ohio 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 in Ohio had greater difficulty reaching AYP than did elementary schools, primarily because their student populations are larger and therefore have more qualifying subgroups—not because their student achievement is lower than in the elementary schools.

- A strong predictor of whether or not a school will make AYP under Ohio's system is whether it has enough limited English proficient (LEP) students[5] to qualify as a separate subgroup. Almost every single school with even one such subgroup failed to make AYP, in part because these students did not meet the state's targets in reading and math.[6]

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance on Ohio'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

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

[6] 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 Ohio Achievement Test, the standardized state test. Specifically, the U.S. Department of Education has issued new NCLB guidelines in recent years that exclude small percentages of LEP students and SWDs from taking the state test or that allow them to take alternative assessments. In this study, however, no valid MAP scores were omitted from consideration.

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[7] 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 10 pupils, for example—could jeopardize students' confidentiality and risk presenting inaccurate results. (With such small groups, random events, like one student being out sick on test day, could skew the outcome.) Because of this flexibility, states have set widely varying *n* sizes for their subgroups, from as few as 10 youngsters to as many as 100.

Many states have also adopted confidence intervals—basically margins of statistical error—to try to account for potential measurement error within the state test. In some states, these margins are quite wide, which has the effect of making it easier to achieve an annual target.

All of these AYP rules vary by state, which means that a school that makes AYP in Wisconsin or Ohio, for example, might not make it under South Carolina's or Idaho's rules (U.S. Department of Education 2008).

What We Studied

We collected students' MAP test scores from the 2005–2006 academic year from 18 elementary and 18 middle schools around the country. We also collected the NCLB subgroup designations for all students in those schools—in other words, whether they had been classified as members of a minority group or as English language learners, among other subgroups.

The schools were not selected as a representative sample of the nation's population. Instead, we selected the schools because they exhibited a range of characteristics on measures such as academic performance, academic growth, and socioeconomic status (the latter calculated by the percentage of students receiving free or reduced-price lunches). Appendix 1 contains a complete discussion of the methodology for this project along with the characteristics of the school sample.[8]

Proficiency cut score estimates for the Ohio Achievement Test (OAT) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Ohio's definitions of proficiency generally ranked below average compared with the standards set by the other 25 states in that study.

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

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

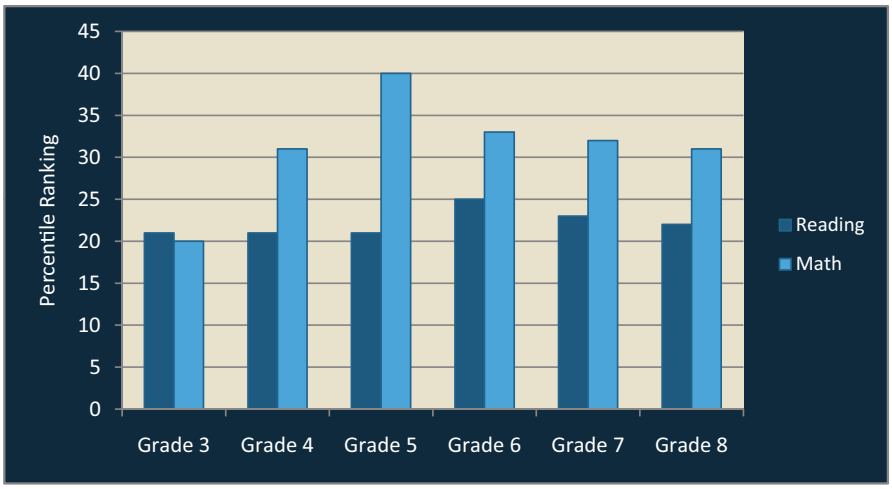

Figure 2. Ohio reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Ohio'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 Ohio's cut scores are at or below the 40th percentile.

Table 1. Ohio AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 30	
	SWDs: 45	
	Low-income students: 30	
	LEP students: 30	
CI	**Applied to proficiency rate calculations?**	
	Not used	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	n/a	77.0
Grade 4	36.0	74.6
Grade 5	n/a	74.6
Grade 6	n/a	80.6
Grade 7	n/a	74.9
Grade 8	n/a	79.0
MATH		
Grade 3	n/a	68.5
Grade 4	36.0	73.7
Grade 5	n/a	59.7
Grade 6	n/a	64.1
Grade 7	n/a	57.8
Grade 8	n/a	58.0

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; n/a = not applicable

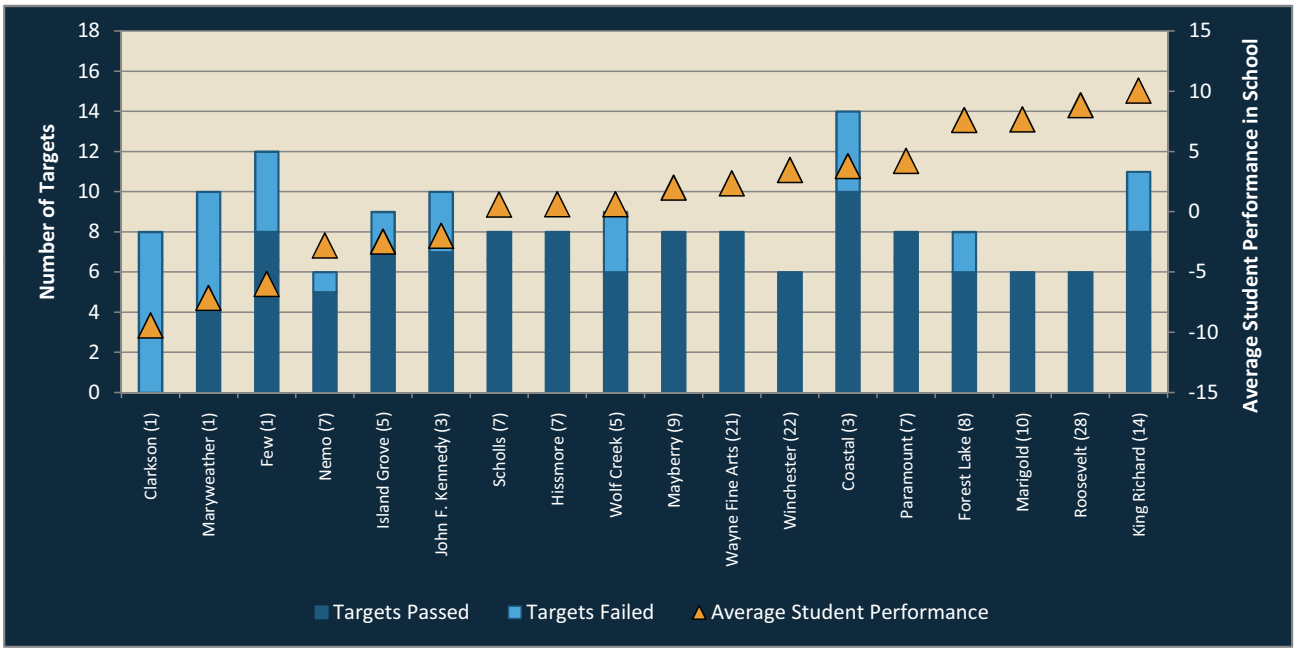

Figure 3. AYP performance of the elementary school sample under Ohio 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under Ohio's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet in order to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not meet them (light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMOs for even a single subgroup didn't make AYP, so any light blue means that the school failed. Forest Lake, for example, met 6 of its 8 targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). This is measured by the average MAP performance of students within the school, and its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

These cut scores were used to estimate whether students would have scored as proficient or better on the Ohio 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 Ohio 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 Ohio's 2008 AYP rules.

Table 1 shows the pertinent Ohio AYP rules that we applied to elementary and middle schools in the current study. Ohio's minimum subgroup size is 30 for three of the four reporting groups (race/ethnicity, low income, and English proficiency), but 45 for the fourth group (students with disabilities), which is higher than most other states we examined.[9]

Specifically, most states have a subgroup size of around 35-40 for reporting purposes but typically don't alter *n* sizes based on particular subgroups. Also unlike most other states, Ohio does not apply confidence intervals (or margins of statistical error) to its measurements of student proficiency rates. This means that schools in Ohio will have a more difficult time meeting their targets than schools in states that do use confidence intervals. Annual targets in Ohio also differ by grade and subject matter (e.g., 57.8% of seventh graders are expected to be proficient in math in 2008; that number changes to 80.6% for sixth graders in reading).

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 stu-

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

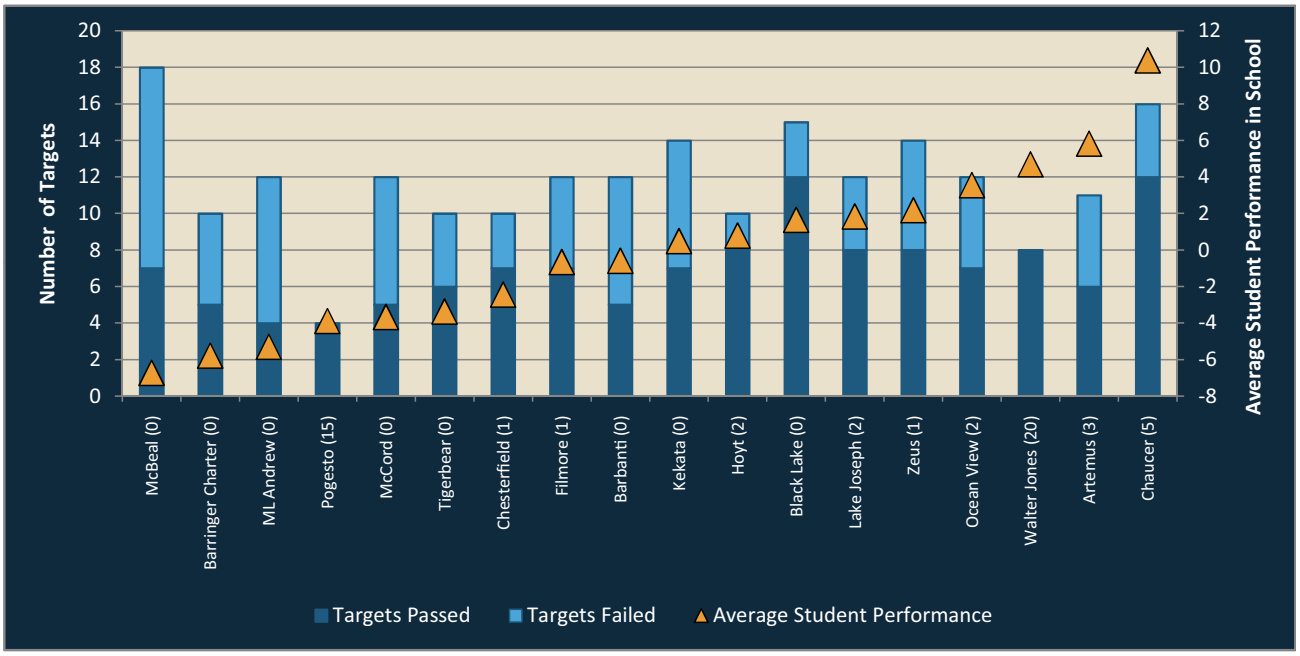

Figure 4. AYP performance of the middle school sample under Ohio 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under Ohio's AYP rules (as described in Table 1). The bars show the number of targets that each school had to meet in order to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not meet them (light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMOs for even a single subgroup did not make AYP, so any light blue means that the school failed. Hoyt, for example, met 8 of its 10 targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). This is measured by the average MAP performance of students within the school, and its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

dents within any failing subgroup by at least 10% relative to the previous year's performance. Because we had access to only a single academic year's data (2005–2006), we were not able to include this in our analysis. As a result, it's possible that some of the schools in our sample that failed to make AYP according to our estimates would have made AYP under real conditions.

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

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

How Did the Sample Schools Fare under Ohio's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Ohio's 2008 AYP rules. **Eight elementary schools made AYP (Scholls, Hissmore, Mayberry, Wayne Fine Arts, Winchester, Paramount, Marigold, and Roosevelt) and 10 failed to make AYP.** The triangles in Figure 3 show the average academic performance of students within the school, with negative values indicating below-grade-level performance for the average student, and positive values indicating above-grade-level performance. The majority of the schools that made AYP are in the right half of the figure, meaning that higher performing students were found at these schools.

Table 2. Elementary subgroup performance of sample schools under the 2008 Ohio 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	54.5%	55.5%	N	N			N	N	N	N					N	N					8	0	0%	N	1
Maryweather	59.8%	63.9%	N	Y			N	N	N	N					N	Y			Y	Y	10	4	40%	N	1
Few	66.0%	67.0%	Y	Y	N	N	N	N	Y	Y					Y	Y			Y	Y	12	8	67%	N	1
Nemo	69.8%	78.1%	Y	Y					N	Y									Y	Y	6	5	83%	N	7
Island Grove	71.4%	77.0%	Y	Y				N	Y	Y					N	Y			Y	Y	9	7	78%	N	4
JFK	74.0%	74.6%	Y	Y	N	N			Y	Y	N	Y							Y	Y	10	7	70%	N	3
Scholls	82.5%	80.3%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	7
Hissmore	81.4%	82.4%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	7
Wolf Creek	73.9%	78.1%	Y	Y				N	N	Y					N	Y			Y	Y	9	6	67%	N	5
Alice Mayberry	80.3%	84.1%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	9
Wayne Fine Arts	82.2%	90.2%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	21
Winchester	78.3%	86.7%	Y	Y											Y	Y			Y	Y	6	6	100%	Y	22
Coastal	81.8%	82.6%	Y	Y	N	N	N	N	Y	Y	Y	Y			Y	Y			Y	Y	14	10	71%	N	3
Paramount	82.2%	82.1%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	7
Forest Lake	89.8%	90.6%	Y	Y	N	N			Y	Y									Y	Y	8	6	75%	N	8
Marigold	91.7%	91.7%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	10
Roosevelt	93.6%	96.9%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	28
King Richard	89.5%	94.2%	Y	Y	N		N	Y	Y	Y					N	Y			Y	Y	11	8	73%	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.

Yet almost without regard to average student performance, the schools that made AYP were primarily those with relatively few qualifying subgroups—and thus the fewest targets to meet. For example, Winchester made it, but had only six targets (two targets in reading and math for its overall student population, two more for its Hispanic subgroup, and two more for its white subgroup).

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Ohio AYP rules. **Of 18 middle schools in our sample, only 2 made AYP**—one low-performance school (Pogesto) and one high-performance school (Walter Jones), both of which have relatively few qualifying subgroups.

Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low or middling performance can still make AYP when the school has fewer targets to meet because it has fewer subgroups. These figures do not, however, indicate which subgroups failed or passed in which school. Information on individual subgroup performance appears in Tables 2 and 3 for elementary and middle schools, respectively.

Table 3. Middle school subgroup performance of sample schools under the 2008 Ohio 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	59.7%	65.5%	N	Y	N	N	N	N	N	N	N	Y	Y	Y	N	N	N	Y	Y	Y	18	7	39%	N	0
Barringer Charter	60.0%	71.8%	N	Y	N	N			N	Y	N	Y			Y	Y					10	5	50%	N	0
ML Andrew	58.6%	71.6%	N	Y	N	N			N	N	N	N			N	Y			Y	Y	12	4	33%	N	0
Pogesto	64.8%	75.9%	Y	Y															Y	Y	4	4	100%	Y	15
McCord Charter	60.4%	73.2%	Y	Y	N	N			N	N	N	N			N	Y			Y	Y	12	5	42%	N	0
Tigerbear	68.5%	68.9%	Y	Y	N	N			Y	Y	N	N							Y	Y	10	6	60%	N	0
Chesterfield	73.8%	74.0%	Y	Y	N	N			Y	Y	Y	N							Y	Y	10	7	70%	N	1
Filmore	70.5%	80.0%	Y	Y	N	N	N	N	Y	Y					N	Y			Y	Y	12	7	58%	N	1
Barbanti	67.7%	75.6%	Y	Y	N	N	N	N	N	N					N	Y			Y	Y	12	5	42%	N	0
Kekata	75.6%	76.7%	Y	Y	N	N	N	N	Y	Y	N	Y			N	N			Y	Y	14	7	50%	N	0
Hoyt	78.2%	80.9%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	2
Black Lake	80.9%	80.9%	Y	Y	N	N	N		Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	15	12	80%	N	0
Lake Joseph	77.3%	84.6%	Y	Y	N	N	N	N	Y	Y					Y	Y			Y	Y	12	8	67%	N	2
Zeus	80.6%	81.6%	Y	Y	N	N	N	N	Y	Y	Y	Y			N	N			Y	Y	14	8	57%	N	1
Ocean View	82.3%	89.4%	Y	Y	N	Y	N	N	N	Y					N	Y			Y	Y	12	7	58%	N	2
Walter Jones	84.3%	89.1%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	20
Artemus	83.8%	86.1%	Y	Y		N			N	N			Y	Y	N	N			Y	Y	11	6	55%	N	3
Chaucer	89.5%	92.6%	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.

Tables 2 and 3 show which subgroups qualified for evaluation at each school (i.e., whether the number of students within that subgroup exceeded the state's minimum *n*), and whether that subgroup passed or failed. Although all schools are evaluated on the proficiency rate of their overall population, potential subgroups that are separately evaluated for AYP include SWDs, students with LEP, low-income students, and the following race/ethnic categories: African American, Asian/Pacific Islander, Hispanic/Latino, American Indian/Alaska Native, and white. Tables 2 and 3 also show whether a school met AYP under the 2008 Ohio 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:

- Overwhelmingly, schools met their targets for their overall student populations. Only one elementary school (Clarkson) failed to meet its math and reading targets for its overall school population. One additional elementary school (Maryweather) failed to meet its overall math target.

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Ohio 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	5	5	4
Students with limited English proficiency	7	5	6
Low-income students	17	4	2
African-American students	6	1	0
Asian/Pacific Islander students	0	0	0
Hispanic students	9	5	1
American Indian/Alaska Native students	0	0	0
White students	17	0	0

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Ohio 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	15	15	15
Students with limited English proficiency	9	9	8
Low-income students	17	7	5
African-American students	11	6	4
Asian/Pacific Islander students	4	0	0
Hispanic students	14	9	4
American Indian/Alaska Native students	1	1	0
White students	17	0	0

- Three sample middle schools (McBeal, Barringer, and ML Andrew) failed to meet their math targets for their overall populations.
- One elementary school (Nemo) met its math and reading targets for every subgroup except low-income students.
- One elementary school (Forest Lake) met all its targets except for students with disabilities.
- Low-income students tended to meet their annual targets, especially in reading at the elementary level. But all schools with qualifying LEP and SWD subgroups failed to make AYP.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively. First, the performance of students with disabilities is proving quite challenging for schools under

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	8	10	2	16
Average student body size	256	344	124	951
Average % low income	37	54	42	45
Average % nonwhite	36	45	27	46
Average performance†	3.72	-0.77	0.40	-0.11
Average % growth‡	113	116	109	97
Average number of targets to meet	7	10	6	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.

Ohio's system, particularly in middle schools, where this subgroup tends to have enough students to meet the state's minimum *n* size of 45. In fact, all but one SWD subgroup in the study (at Ocean View) failed to meets its AYP targets. Students with limited English proficiency are also struggling to meet the state's targets; almost every school with a large enough LEP population to qualify as a separate subgroup failed to meet its reading targets for these students.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Ohio's NCLB accountability system is, in many respects, behaving like those in other states. For example, among the elementary schools in our sample, Roosevelt, Winchester, and Wayne Fine Arts all made AYP in the greatest number of states—28, 22, and 21, respectively. And these schools all made AYP in Ohio, too. Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make AYP in Ohio.

But Ohio is also home to a few anomalies. First, consider Mayberry Elementary (see Figure 3). It failed to make AYP in 19 of the 28 states in our sample, yet made AYP in Ohio. In examining Table 2, we can see that Mayberry didn't meet the minimum numbers for the LEP or SWD subgroups, which created difficulty for so many other schools within the sample. With fewer accountable subgroups and with relatively easy proficiency standards (Figure 2), Mayberry attained AYP in Ohio, even when other schools with higher average performance failed. This seems to be the case for a few other elementary schools (Hissmore, Paramount, and Marigold) and for at least one middle school (Pogesto).

This is consistent with the patterns shown in Table 6, which compares schools making and not making AYP on a number of academic and demographic dimensions. Within the sample, passing schools do indeed show higher average student performance, but they also differ in the following ways: they have smaller student populations (dramatically so at the middle school level) and fewer subgroups (and thus fewer targets to meet).

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 Ohio's AYP rules (and AMOs) for 2008. We found that 8 elementary schools and 2 middle schools—10 in all, from a sample of 36—would have made AYP in Ohio. Looking across the 28 state accountability systems examined in the study, this puts Ohio towards the high end of the sample distribution in terms of the number of schools making AYP (see Figure 1). Part of the reason that some schools made AYP in Ohio and not in other states is that Ohio's proficiency standards are relatively easy. In addition, Ohio's minimum *n* size for SWDs is higher than in other states, meaning that fewer SWD subgroups in Ohio (particularly at the elementary level) are likely to be held accountable for performance.

Because the overriding goal of NCLB is to eliminate educational disparities within and across states, it's important to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, Ohio's NCLB accountability system is working exactly as Congress intended: identifying as "needing attention" schools with relatively high test score averages that mask low performance for particular groups of students, such as low-income students. Almost all of the sample schools met the Ohio reading and math targets for their overall populations, 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, 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 (and in Ohio, that those subgroup *n* sizes change based on subgroup classification)? Even if actual participation guidelines for English language learners and students with disabilities are more generous under the current state assessment system,[10] doesn't the massive failure of these students, especially in middle schools, to meet Ohio's targets indicate that a new approach is needed for holding schools accountable for their performance? Yes, schools should redouble their efforts to boost achievement for LEP students and students with disabilities, as for other students, but when almost no school is able to meet the goal, perhaps that indicates that the goal is unrealistic. These will be critical considerations for Congress as it takes up NCLB 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,

[10] See footnote 6.

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.