Maine

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 the NCLB accountability system in Maine—particularly how its various rules, criteria, and practices result in schools either making AYP or not making AYP. It also gauges how tough the Maine 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 the Maine 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 the Maine AYP rules for academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- We estimate that **14 of 18 elementary schools** and **16 of 18 middle schools** in our sample failed to make adequate yearly progress in 2008 under Maine'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 Maine's proficiency cut scores which are above average, or relatively difficult, compared with the standards set by the other states in the study. In addition, Maine's minimum subgroup size is 20, which is quite small compared to most other states. This means that schools in Maine will have more accountable subgroups than similar schools in other states.
- Looking across the 28 state accountability systems examined in the study, **12 states passed more of the sample elementary schools than did Maine, while 13 states passed fewer elementary schools. In other words, Maine was about in the middle** (see Figure 1).
- Nearly all of the schools in our sample that failed to make AYP in Maine are meeting expected targets for their overall populations but failing because of the performance of individual subgroups.[2]
- As is the case in other states, schools with fewer subgroups attain AYP more easily in Maine than schools with more subgroups, even when their average student performance is much lower. In other words,

Maine's AYP rules place the state toward the mid to lower end of the state distribution in terms of the number of schools making AYP. Maine's proficiency cut scores generally ranked above average, or relatively difficult, compared with the standards set by the other states in the study. In addition, Maine's minimum subgroup size is 20, which is quite small compared to most other states. This means that more subgroups are held accountable in Maine than would be in other states. In fact, all but two schools with limited English proficient (LEP) or students-with-disabilities (SWD) subgroups failed to make AYP, in part because these students did not meet the state's proficiency targets in math and reading. Students with disabilities had a particularly hard time meeting their AYP targets at the middle school level.

[1] A cut score is the minimum score a student must receive on NEWA's Measures of Academic Progress (MAP) that is equivalent to performing proficient on the Maine Education Assessment (MEA).

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

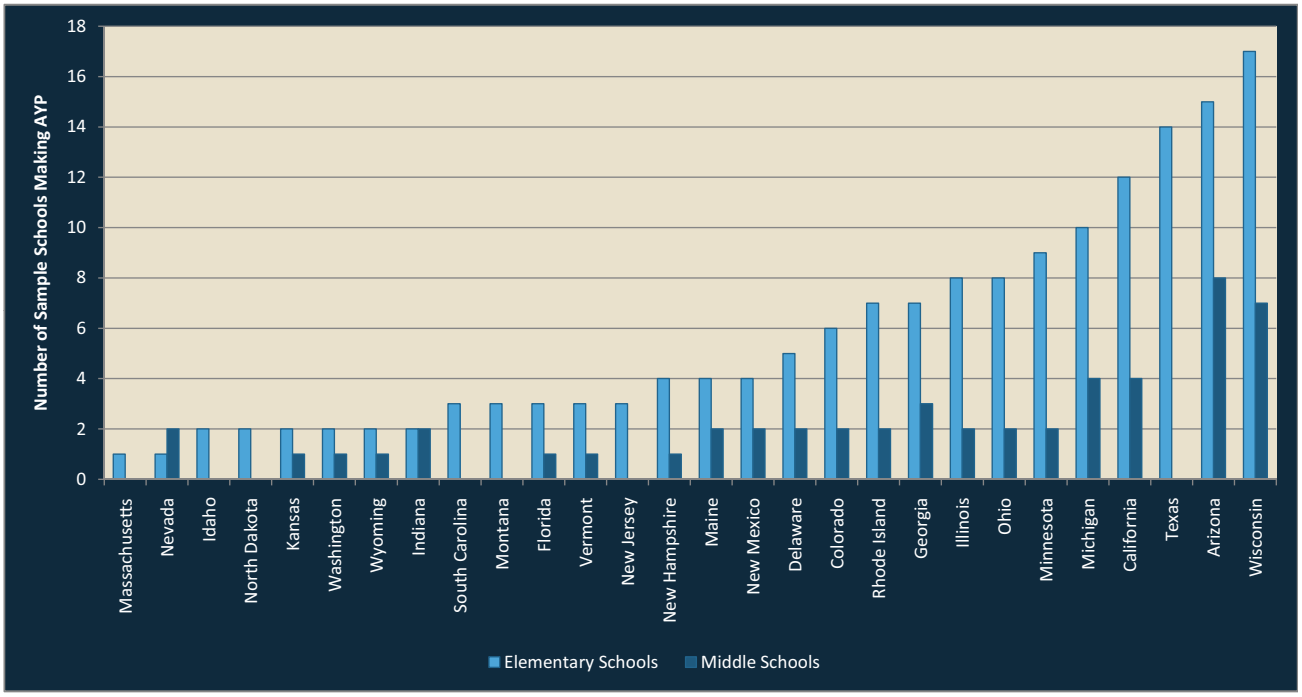

Figure 1. Number of sample schools making AYP by state

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

schools with greater diversity and size face greater challenges in making AYP.

- Middle schools have greater difficulty reaching AYP in Maine than do elementary schools, primarily because their student populations are larger and therefore, have more qualifying subgroups—not because their student achievement is any lower than in the elementary schools.

- A strong predictor of whether or not a school will make AYP under the Maine system is whether it has enough students with disabilities (SWD) or English language learners to qualify as a separate subgroup. Nearly all schools with limited English proficient (LEP)[3] or SWD subgroups failed to make AYP, in part because these students did not meet the state's proficiency targets in reading.[4]

Introduction

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

Our study expands on *The Proficiency Illusion* by examining other key factors of state NCLB accountability plans and how they interact with state proficiency stan-

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

[4] SWDs are defined as those students following individualized education plans. We should also note that our subgroup findings for LEP students and SWDs may be more negative than actual findings, mostly because of the likely differences between how LEP students and SWDs are treated in MAP, the assessment we used in this study, and in the Maine Education Assessment, 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.

dards 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. These AMOs vary by state (as do, of course, the difficulty of the proficiency standards).

States also determine the minimum number of students that must constitute a subgroup in order for its scores to be analyzed separately (also called the minimum *n* [number of students in sample] size). The rationale is that reporting the results of very small subgroups—fewer than ten pupils, for example—could jeopardize students' confidentiality and risk presenting inaccurate results. (With such small groups, random events, like one student being out sick on test day, could skew the outcome.) Because of this flexibility, states have set widely varying *n* sizes for their subgroups, from as few as 10 youngsters to as many as 100.

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

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

What We Studied

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

The schools were not selected as a representative sample of the nation's population. Instead, we selected the schools because they exhibited a range of characteristics on measures such as academic performance, academic growth, and socioeconomic status (the latter calculated by the percentage of students receiving free or reduced-price lunches). Appendix 1 contains a complete discussion of the methodology for this project along with the characteristics of the school sample.[6]

Proficiency cut score estimates for the Maine Education Assessment (MEA) are taken from *The Proficiency Illusion* (as shown in Figure 1), which found that **Maine's proficiency cut scores were generally ranked above average, or relatively difficult, 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 Maine 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 Maine AYP rules for 2008. In other words,

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

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

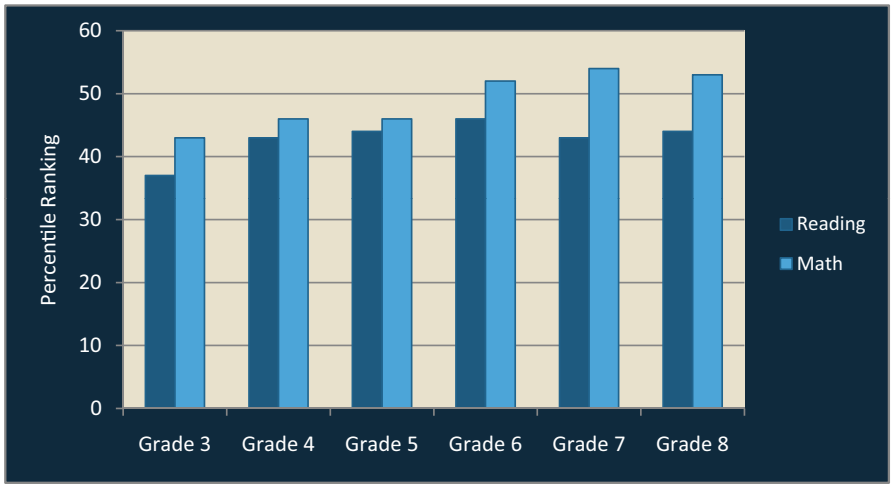

Figure 2. Maine reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Maine's cut scores (as 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 Maine's cut scores are below the 55th percentile.

Table 1. Maine AYP Rules for 2008

Subgroup minimum *n*	Race/ethnicity: 20	
	SWDs: 20	
	Low-income students: 20	
	LEP students: 20	
CI	**Applied to proficiency rate calculations?**	
	Yes; 95% CI	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	n/a	49
Grade 4	34	49
Grade 5	n/a	49
Grade 6	n/a	50
Grade 7	n/a	50
Grade 8	35	50
MATH		
Grade 3	n/a	32
Grade 4	12	32
Grade 5	n/a	32
Grade 6	n/a	33
Grade 7	n/a	33
Grade 8	13	33

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

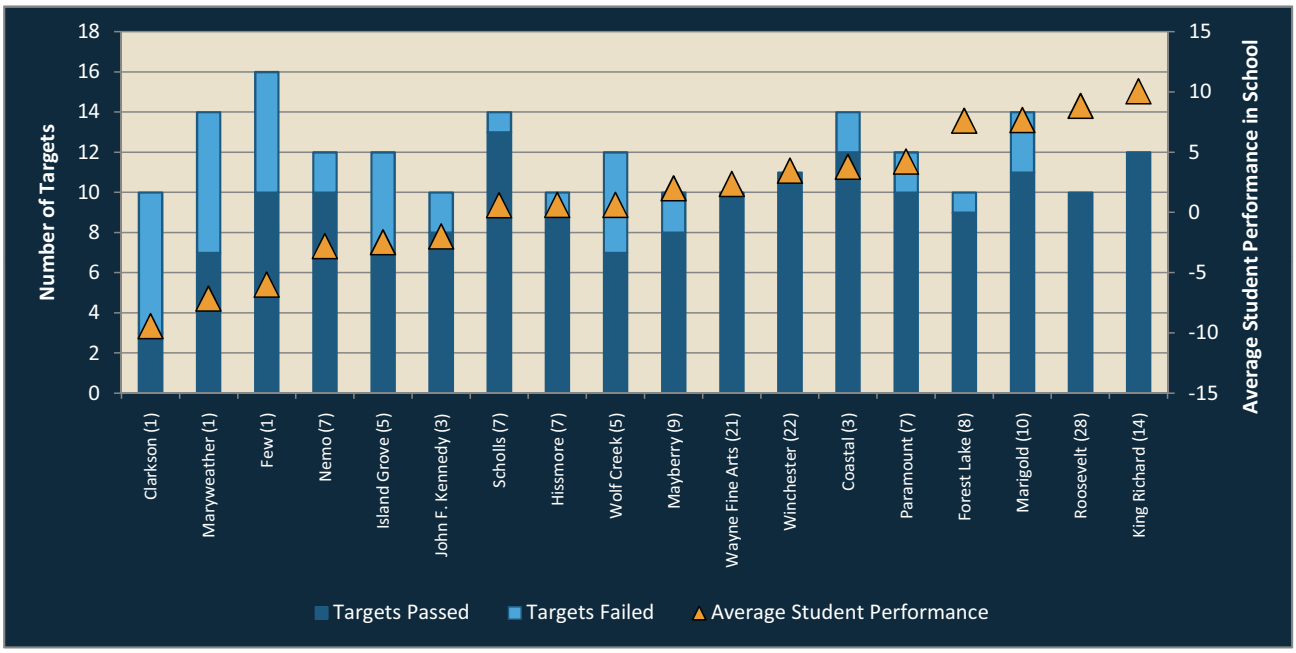

Figure 3. AYP performance of the elementary school sample under Maine's 2008 AYP Rules

Note: This figure indicates how each of the elementary schools within the sample fared under the Maine 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 didn't make AYP, so any light blue means the school failed. Mayberry, for example, met eight of its ten 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.

the school data and our proficiency cut score estimates are from 2005–2006, but we are applying them against the Maine 2008 AYP rules.

Table 1 shows the pertinent Maine AYP rules that were applied to elementary and middle schools in the current study. Maine's minimum subgroup size is 20, which is quite small compared to most other states examined in the study. This means that schools in Maine will have more accountable subgroups than similar schools in other states. Maine, like the majority of states examined in the study, applies the 95% confidence intervals to their measurements of student proficiency rates, which makes it easier to achieve their annual measurable objectives. So, for instance, even though schools are supposed to get 50% of their grade 6 students to the proficient level on the state reading test, as well as 50% of the grade 6 students in each subgroup, applying the confidence interval means that the real target can be lower, particularly with smaller groups.

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

Furthermore, attendance and test-participation rates are beyond the scope of the study. 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.

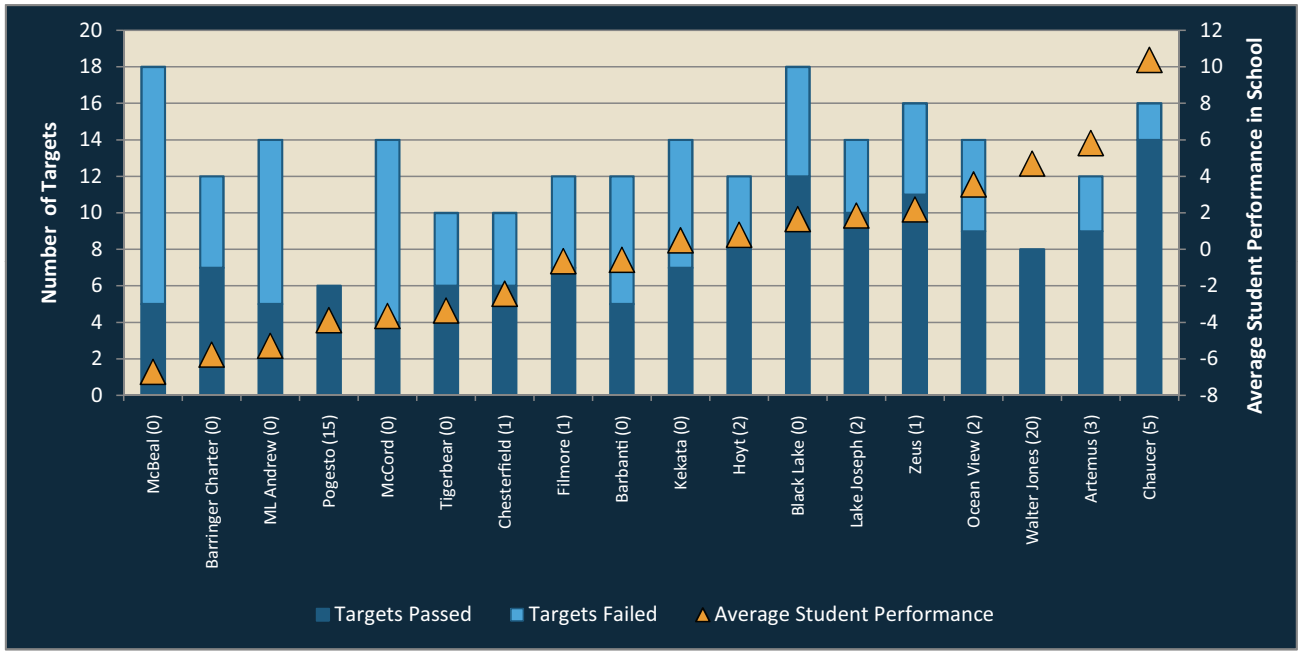

Figure 4. AYP performance of the middle school sample under Maine's 2008 AYP Rules

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

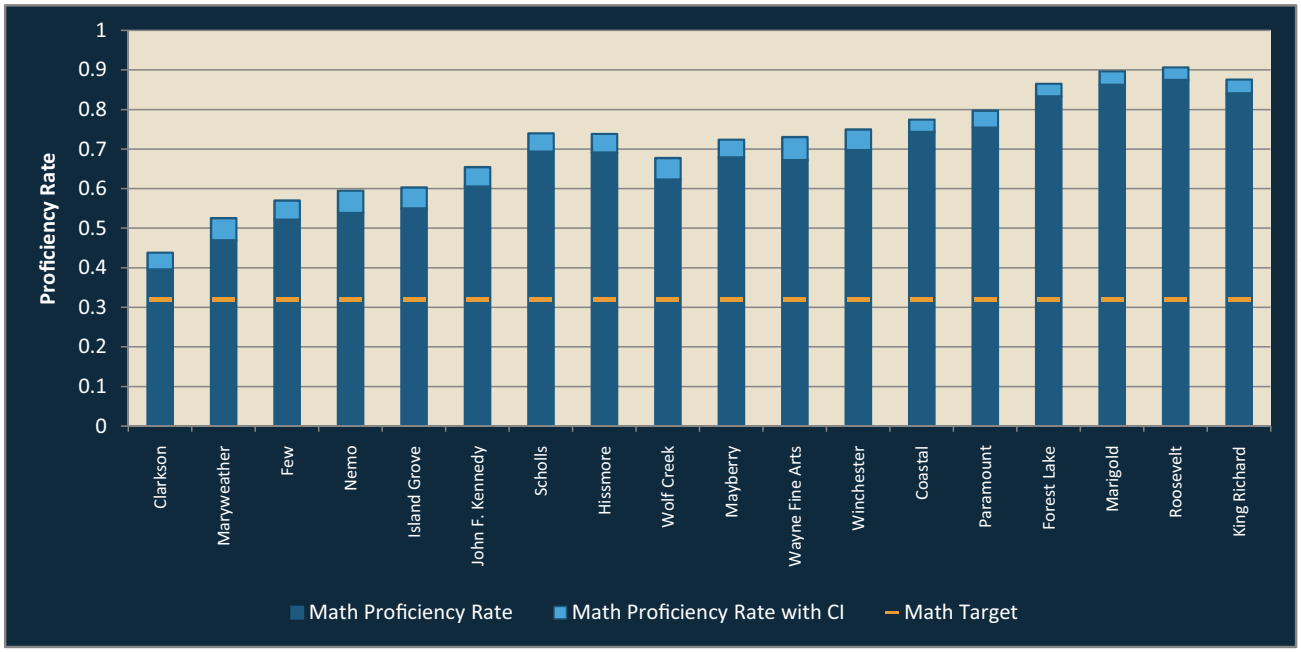

Figure 5. Impact of the confidence interval on elementary school math proficiency rates under Maine's 2008 AYP rules

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 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.

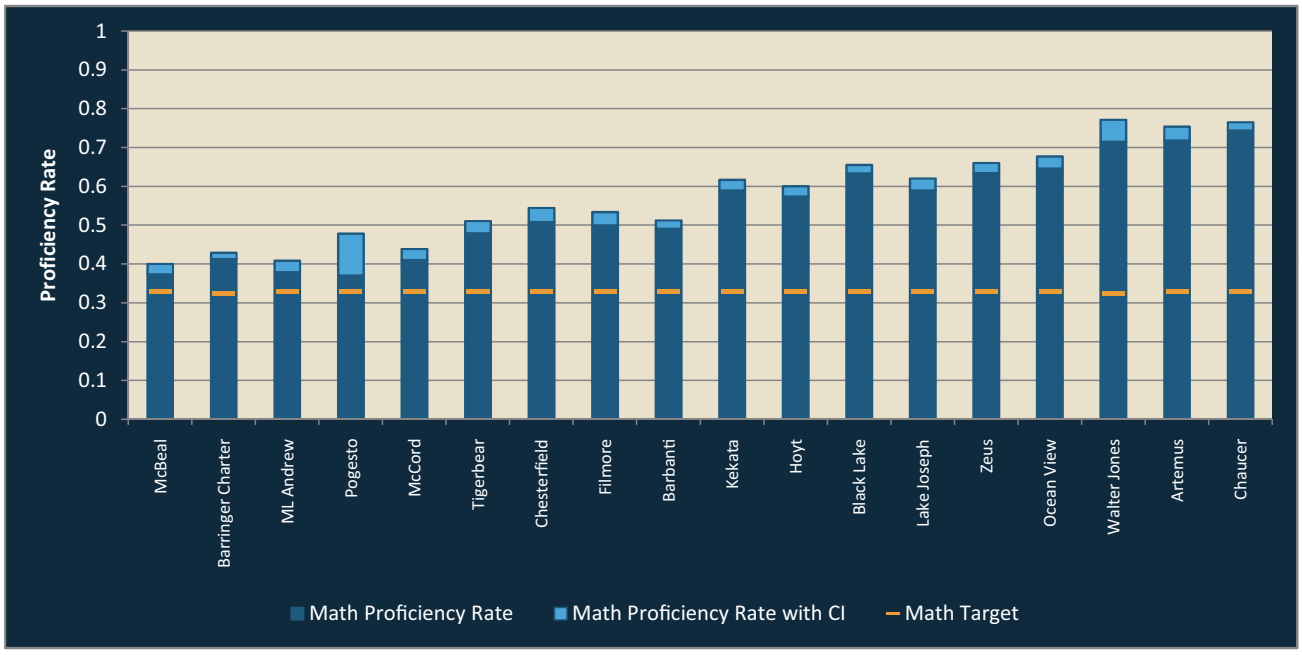

Figure 6. Impact of the confidence interval on middle school math proficiency rates under Maine's 2008 AYP rules

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 none of the sample middle 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.

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 Maine's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Maine's 2008 AYP rules. **Only 4 schools (Wayne Fine Arts, Winchester, Roosevelt, and King Richard) made AYP while 14 failed to make it**. The triangles in the Figure 3 show the average academic performance of students within the school, with negative values indicating below-grade-level performance for the average student, and positive values indicating above-grade-level performance. All schools making AYP are in the right half of the figure, meaning that these schools contain the highest performing students.

Yet almost without regard to average student performance, the only schools actually to make AYP are those with relatively few qualifying subgroups—and thus the fewest targets to meet. For example, Wayne Fine Arts made AYP, but only has ten targets.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Maine AYP rules. **Out of 18 in our sample, only 2 middle schools make AYP** – 1 low-performance school (Pogesto), and 1 high-performance school (Walter Jones), both of which have few qualifying subgroups.

Figures 5 and 6 indicate the degree to which schools' overall math proficiency rates are aided by Maine's confidence interval for elementary and middle schools, respectively. 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 applying the confidence interval. The orange lines show the AMO needed to meet AYP. **These figures show that none of the sample elementary or middle schools were assisted**

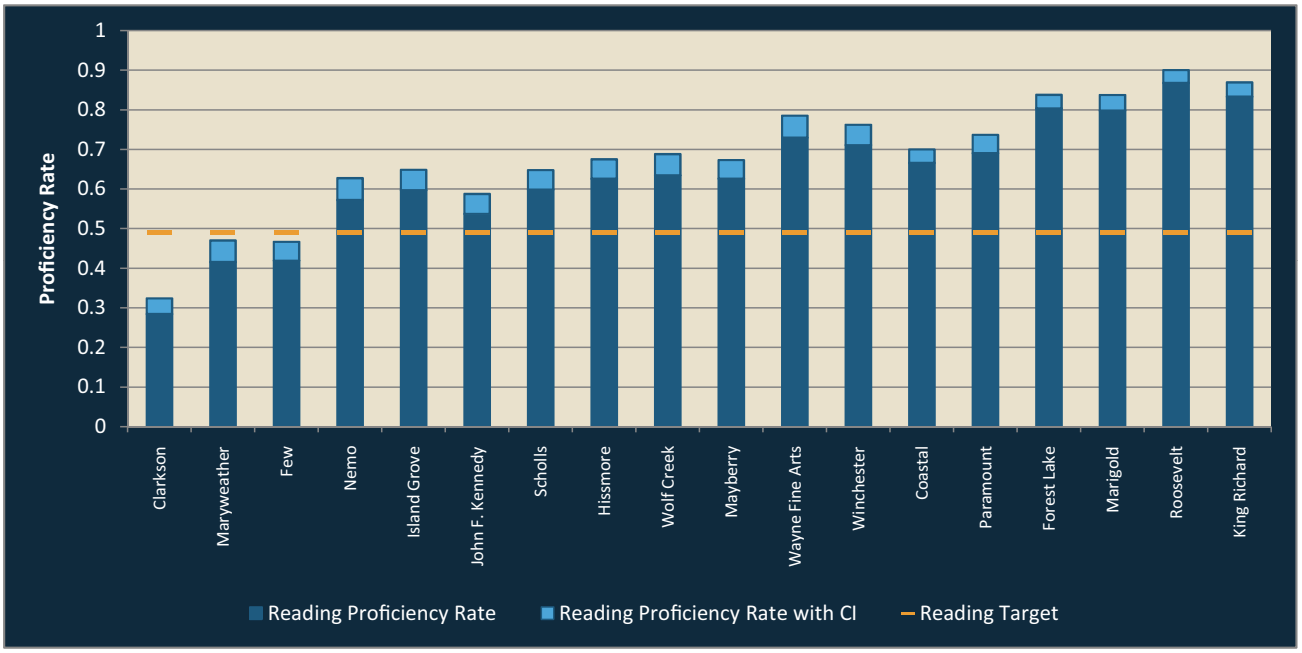

Figure 7. Impact of the confidence interval on elementary school reading proficiency rates under Maine's 2008 AYP rules

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 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.

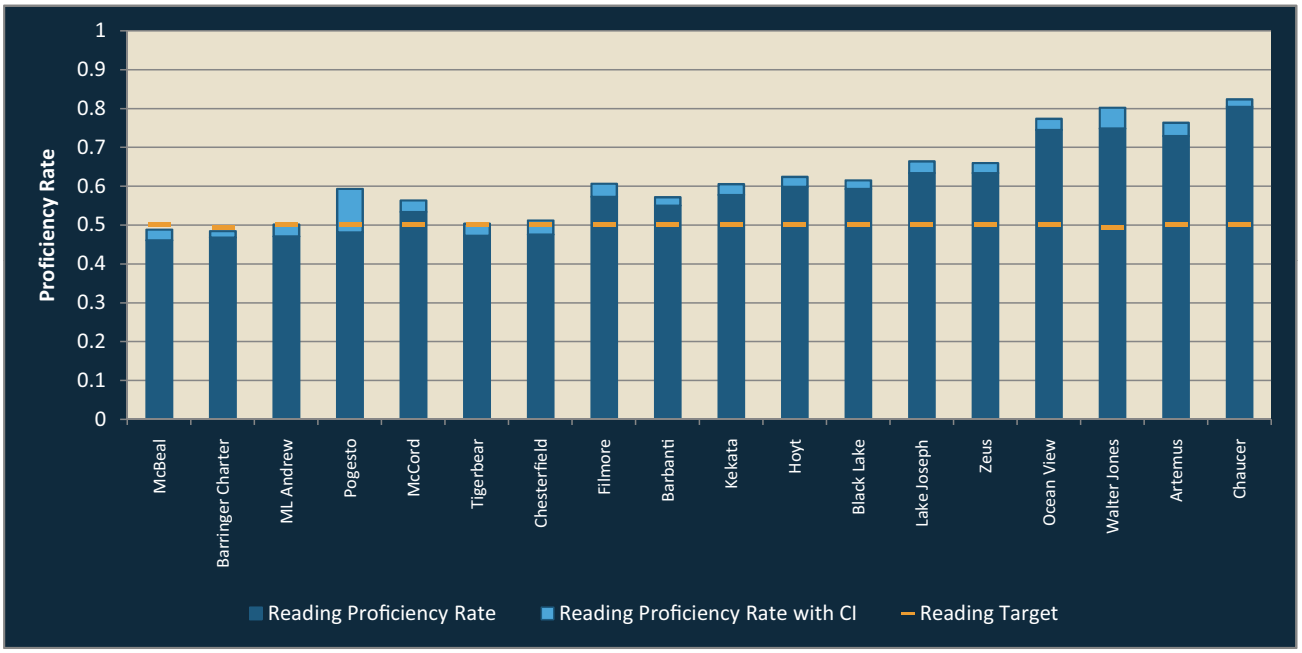

Figure 8. Impact of the confidence interval on middle school reading proficiency rates under Maine's 2008 AYP rules

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that two of the sample middle schools (Pogesto and Chesterfield) 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 Maine 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	39.6%	28.5%	Y	N	N	N	N	N	Y	N					Y	N					10	3	30%	N	1
Maryweather	47.0%	41.6%	Y	N	N	N	Y	N	Y	N	Y	N			Y	N			Y	Y	14	7	50%	N	1
Few	52.2%	41.9%	Y	N	Y	N	Y	N	Y	N	Y	Y			Y	N	Y	N	Y	Y	16	10	63%	N	1
Nemo	54.0%	57.2%	Y	Y	Y	N			Y	N	Y	Y			Y	Y			Y	Y	12	10	83%	N	7
Island Grove	55.0%	59.7%	Y	Y	N	N	N	N	Y	Y					Y	N			Y	Y	12	7	58%	N	4
JFK	60.6%	53.7%	Y	Y	Y	N			Y	Y	Y	N							Y	Y	10	8	80%	N	3
Scholls	69.4%	59.9%	Y	Y	Y	N	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	13	93%	N	7
Hissmore	69.2%	62.6%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	7
Wolf Creek	62.4%	63.5%	Y	Y	N	N	Y	N	Y	N					Y	N			Y	Y	12	7	58%	N	5
Alice Mayberry	67.9%	62.6%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	9
Wayne Fine Arts	67.2%	73.0%	Y	Y					Y	Y	Y	Y			Y	Y			Y	Y	10	10	100%	Y	21
Winchester	69.8%	71.1%	Y	Y	Y	Y			Y	Y				Y	Y	Y			Y	Y	11	11	100%	Y	22
Coastal	74.4%	66.6%	Y	Y	Y	N	Y	N	Y	Y	Y	Y			Y	Y			Y	Y	14	12	86%	N	3
Paramount	75.5%	69.0%	Y	Y	Y	Y	N	N	Y	Y					Y	Y			Y	Y	12	10	83%	N	7
Forest Lake	83.4%	80.4%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	8
Marigold	86.3%	79.8%	Y	Y	Y	N	Y	N	Y	Y			Y	Y	Y	N			Y	Y	14	11	79%	N	10
Roosevelt	87.5%	86.8%	Y	Y					Y	Y	Y	Y			Y	Y			Y	Y	10	10	100%	Y	28
King Richard	84.1%	83.3%	Y	Y	Y	Y	Y	Y	Y	Y					Y	Y			Y	Y	12	12	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.

by the confidence intervals, because the math targets in Maine are low, relative to the schools' overall performance. In other words, the sample schools met the targets without the assistance of the confidence interval.

The effect of confidence intervals on reading proficiency rates for elementary and middle schools is similar (Figures 7 and 8). In reading, none of the elementary schools make use of the confidence interval to meet the overall target. Two of the sample middle schools (Pogesto and Chesterfield) met the overall target with the help of the confidence interval (see Figure 8), but we know that Chesterfield still failed to meet all its subgroup targets (Figure 4). **In short, the application of the confidence interval has only modest impact on AYP decisions for the sample schools in Maine.**[7]

[7] 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 through 8. 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 3. Middle school subgroup performance of sample schools under the 2008 Maine 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	37.4%	46.1%	Y	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N	N	Y	Y	18	5	28%	N	0
Barringer Charter	41.3%	46.8%	Y	N	N	N			Y	N	Y	N			Y	Y			Y	Y	12	7	58%	N	0
ML Andrew	37.9%	47.1%	Y	Y	N	N	N	N	N	N	N	N			Y	N			Y	Y	14	5	36%	N	0
Pogesto	37.0%	48.1%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	15
McCord Charter	41.0%	53.4%	Y	Y	N	N	N	N	N	N	N	N			N	N			Y	Y	14	4	29%	N	0
Tigerbear	47.9%	47.2%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	0
Chesterfield	50.9%	47.6%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	1
Filmore	50.0%	57.3%	Y	Y	N	N	N	N	Y	Y					Y	N			Y	Y	12	7	58%	N	1
Barbanti	49.1%	55.0%	Y	Y	N	N	N	N	N	N					Y	N			Y	Y	12	5	42%	N	0
Kekata	58.9%	57.8%	Y	Y	N	N	N	N	Y	N	Y	N			Y	N			Y	Y	14	7	50%	N	0
Hoyt	57.4%	59.9%	Y	Y	N	N			Y	Y	Y	N			Y	N			Y	Y	12	8	67%	N	2
Black Lake	63.4%	59.3%	Y	Y	N	N	Y	N	Y	N	Y	N	Y	Y	Y	N	Y	Y	Y	Y	18	12	67%	N	0
Lake Joseph	59.0%	63.4%	Y	Y	N	N	N	N	Y	Y	Y	Y			Y	Y			Y	Y	14	10	71%	N	2
Zeus	63.5%	63.4%	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	N			Y	Y	16	11	69%	N	1
Ocean View	64.6%	74.5%	Y	Y	Y	N	N	N	Y	N			Y	Y	Y	N			Y	Y	14	9	64%	N	2
Walter Jones	71.5%	74.9%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	20
Artemus	71.8%	72.9%	Y	Y	Y	N			Y	N			Y	Y	Y	N			Y	Y	12	9	75%	N	3
Chaucer	74.4%	80.4%	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	16	14	88%	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.

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, thanks to fewer subgroups. These figures do not, however, indicate which subgroups failed in which school. Information on individual subgroup performance appears in Tables 2 and 3 for elementary and middle schools, respectively.

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 Maine rules, and the total number of states within the study in which that school met AYP.

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Maine 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	5	13
Students with limited English proficiency	10	3	8
Low-income students	18	0	5
African-American students	11	0	2
Asian/Pacific Islander students	1	0	0
Hispanic students	14	0	6
American Indian/Alaska Native students	1	0	1
White students	17	0	0

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Maine 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	13	16
Students with limited English proficiency	11	9	11
Low-income students	18	4	11
African-American students	12	3	9
Asian/Pacific Islander students	6	0	0
Hispanic students	15	2	11
American Indian/Alaska Native students	2	1	1
White students	18	0	0

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

- Three elementary schools (Clarkson, Maryweather, and Few) and two middle schools (McBeal and Barringer) failed to meet the reading targets for their overall school population.
- No school failed to meet their overall targets in math.
- Four of the fourteen failing elementary schools (Scholls, Hissmore, Alice Mayberry, and Forest Lake) missed only for the SWD subgroup.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively. First, elementary students did better in math than reading, perhaps because Maine's proficiency targets are lower in math than in reading at the elementary grades

(32% and 49%, respectively, as shown in Table 1). The performance of SWD students is also proving challenging for schools under Maine's system, particularly in middle schools, where this subgroup tends to have enough students to meet the state's minimum *n* of 20. In fact, all but two elementary and all middle schools in the study with a qualifying SWD subgroup failed to make AYP. Students with LEP are also struggling to meet the state's targets; all but two elementary schools with a large enough LEP population to qualify as a separate subgroup failed to meet their reading targets for these students.

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

Other state reports contain a section comparing some of the characteristics of the sample schools that made AYP versus those that did not. In Maine, there were no striking differences between schools that made and didn't make AYP, other than the (expected) finding that the former had students with higher average student performance than the latter, as measured by NWEA reading and math tests.

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 have fared under the Maine AYP rules (and AMOs) for 2008. We found that only 4 elementary schools and 2 middle schools–6 in all from a sample of 36–would have made AYP in Maine. Looking across the 28 state accountability systems examined in the study, this puts Maine in the middle of the distribution in terms of the number of schools making AYP (as shown in Figure 1).

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, the NCLB accountability system in Maine is working exactly as Congress intended: identifying as needing attention those schools with relatively high test score averages that mask low performance for particular groups of students, such as low-income or Hispanic students. Almost all the sample schools met the Maine AMO targets for their student populations as a whole, i.e., not considering subgroup results. In the pre-NCLB era, such schools might have been considered effective or at least not in need of improvement, even though sizable numbers of their pupils were not meeting state standards. Disaggregating data by race, income, and so on has made those students visible. That is surely a positive step.

Yet NCLB's design flaws are also readily apparent. Does it make sense that 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,[8] does the massive failure of middle school students to meet Maine's targets indicate that a new approach is needed for holding schools accountable for the performance of these students? Yes, schools should redouble their efforts to boost achievement for ELL students and students with disabilities, as for other 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.

[8] See footnote 4.

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.