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 New Hampshire'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 New Hampshire'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 New Hampshire'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 New Hampshire's 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 **17 of 18 middle schools** in our sample **failed to make adequate yearly progress** in 2008 under New Hampshire's accountability system. (This high failure 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 New Hampshire was exceeded in 12 other sample states. New Hampshire ties Maine and New Mexico with 4 elementary schools making AYP. In addition, New Hampshire is one of 6 states with just a single passing middle school in the sample (see Figure 1).

- Many of the schools in our sample that failed to make AYP in New Hampshire met expected targets for their overall populations[2] but failed because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.

New Hampshire is squarely in the middle of the state distribution in terms of the number of schools making AYP. This is not surprising given New Hampshire's complex rule set. First, New Hampshire's 99 percent confidence interval provides schools with greater leniency than the more commonly used 95 percent confidence interval. Second, the state awards students "partial credit" for performing at lower levels of proficiency. On the other hand, New Hampshire's annual targets require that schools reach a relatively high bar (e.g., in 2008, 86 percent of students in all subgroups must reach proficiency on the state's reading exam in order to make AYP). So, while the state's definitions of proficiency generally ranked about average compared with the standards set by other states, getting 86 percent of all students over that bar is relatively difficult. Finally, New Hampshire's minimum subgroup size is 11, which is much smaller than the subgroup size in most other states we examined. This means that more subgroups are held separately accountable for performance than would be in other jurisdictions.

---

[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 New England Common Assessment Program.

[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.
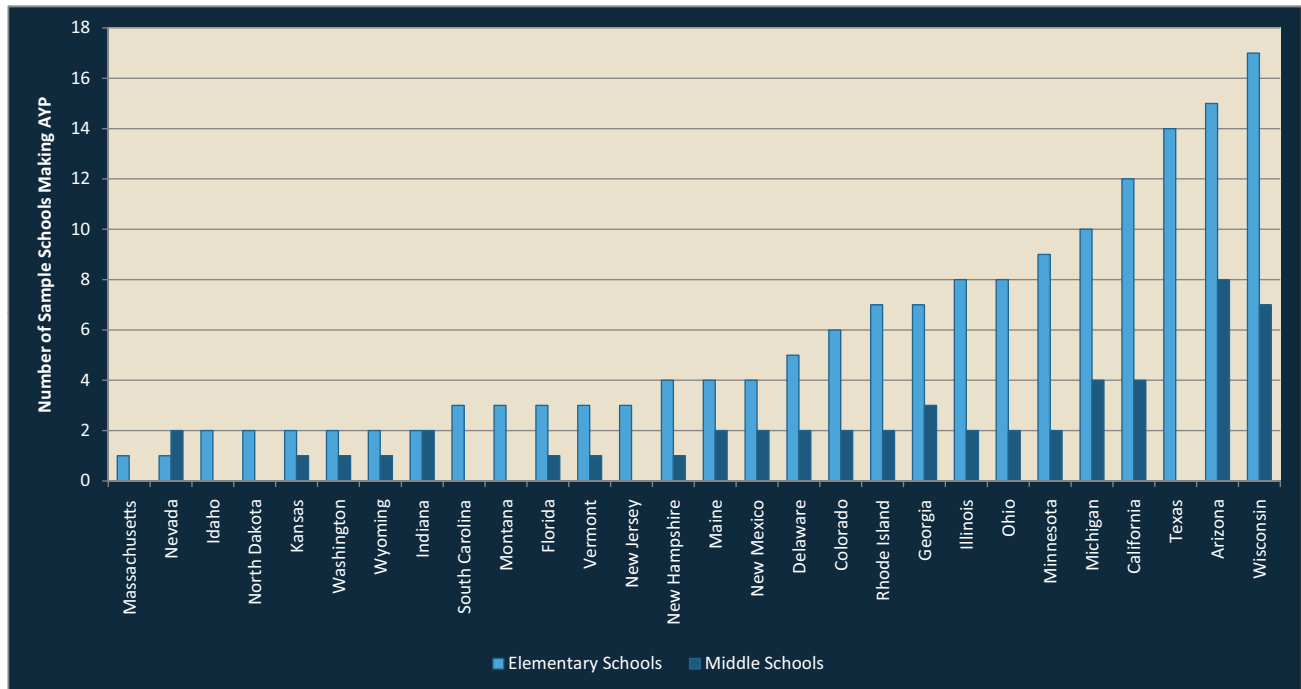
Figure 1. Number of sample schools making AYP by state

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.

- As in most states, middle schools in New Hampshire had greater difficulty reaching AYP than elementary schools, possibly because their student populations are larger and therefore have more qualifying subgroups—not because their student achievement is lower.

- A strong predictor of whether or not a school would make AYP under New Hampshire's system is whether it has enough limited English proficient (LEP)[3] students or SWDs to qualify as a separate subgroup. Most schools with a LEP or SWD subgroup failed to make AYP.[4]

- **Although New Hampshire awards "partial credit" to students performing at lower levels and uses a fairly lenient confidence interval (margin of statis-** tical error), **most schools still failed to make AYP, partly because of New Hampshire's small minimum *n* size (which makes more subgroups accountable) and partly because of New Hampshire's fairly high annual targets or AMOs.**

Introduction

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

---

[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 slightly more negative than actual findings, mostly because of the differences in testing practices between the Measures of Academic Progress (MAP), the assessment we used in this study, and in the New England Common Assessment Program, the standardized state test. Specifically, the U.S. Department of Education has issued new NCLB 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.

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 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.[6]

Proficiency cut score estimates for the New England Common Assessment Program are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that New Hampshire's definitions of proficiency generally ranked about average compared with the standards set by

---

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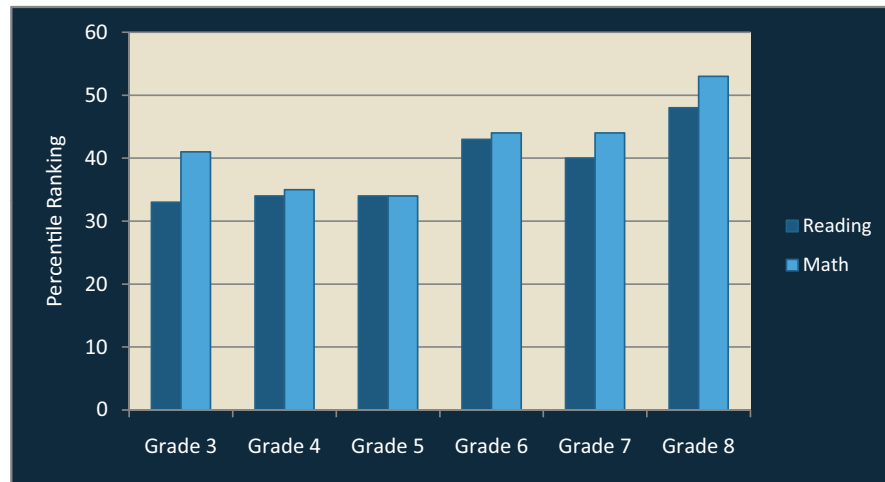
Figure 2. New Hampshire reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of New Hampshire'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 New Hampshire's cut scores are below the 55th percentile.

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 New Hampshire 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 New Hampshire AYP rules for 2008. So to clarify, the school data and our proficiency cut score estimates are from academic year 2005–2006, but we are applying them against New Hampshire's 2008 AYP rules.

Table 1 shows the pertinent New Hampshire AYP rules that we applied to elementary and middle schools in the current study. New Hampshire's minimum subgroup size is 11, which is much smaller than the ones in most other states we examined.[7] **This means that schools in New Hampshire have more accountable subgroups than do similar schools in other states.**

Most states also apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates. **New Hampshire's 99% confidence interval, however, gives schools greater leniency than the** **more commonly used 95% confidence interval**. This means that if the annual target requires a school to achieve, for example, 86% reading proficiency among its grade 3-8 students (and 86% reading proficiency among its grade 3-8 students in each subgroup), applying the confidence interval means that the real target can be lower, particularly with smaller groups. Finally, rather than simply measuring the percentage of students achieving a "proficient" or higher performance level, **New Hampshire employs a proficiency "index," which gives partial credit to students performing at levels less than proficient**. In the short term, the index makes it easier for schools to achieve their targets, though as the targets approach the 100% requirement of NCLB in 2014, the assistance of the index diminishes.[8]

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),

---

[7] It's also likely that New Hampshire has small schools so a small *n* size may be appropriate.

[8] In six of the states studied (Massachusetts, Minnesota, Rhode Island, Vermont, and Wisconsin, as well as New Hampshire), an index is used that gives full credit to students who achieve proficient (or better) and partial credit to students performing at lower levels. Consequently, the resultant score in states using this "hybrid" model is always higher than the actual proficiency percentage (giving students partial credit for achieving lower proficiency levels is obviously better than no credit, at least for the schools' ratings). The index provides a fair amount of help when annual targets are below 50%; however, once targets rise above 75%, the index has far less impact.

Table 1. New Hampshire AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 11	
	SWDs: 11	
	Low-income students: 11	
	LEP students: 11	
CI	**Applied to proficiency rate calculations?**	
	Yes; 99% CI used	
AMOs	**Baseline proficiency levels as of 2002 (index)**	**2008 targets (index)**
READING/LANGUAGE ARTS		
Grade 3	82	86
Grade 4	82	86
Grade 5	82	86
Grade 6	82	86
Grade 7	82	86
Grade 8	82	86
MATH		
Grade 3	76	82
Grade 4	76	82
Grade 5	76	82
Grade 6	76	82
Grade 7	76	82
Grade 8	76	82

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

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 aca-

demic 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 New Hampshire's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under New Hampshire's 2008 AYP rules. **Only 4 elementary schools (Wayne Fine Arts, Win-**
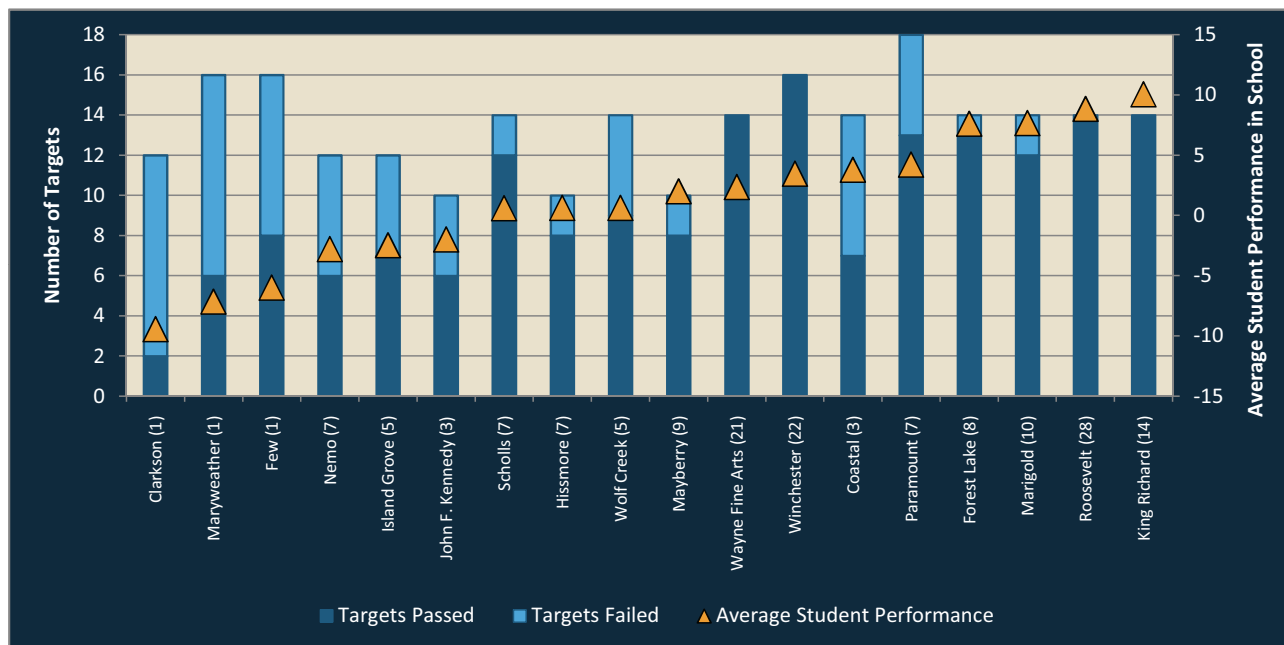
Figure 3. AYP performance of the elementary school sample under New Hampshire's 2008 AYP rules

Note: This figure shows how each of the elementary schools within the sample fared under New Hampshire'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. Marigold Elementary, for example, met 12 of its 14 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; 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.

chester, Roosevelt, and King Richard) made AYP and 14 failed. 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. All schools that made AYP are in the right half of the figure, meaning that relatively high performing students were found at these schools.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 New Hampshire AYP rules. **Of 18 middle schools in our sample, only 1 made AYP**—a high-performance school (Walter Jones) that has relatively few qualifying subgroups compared to other schools.

Figures 5 and 6 indicate the degree to which math proficiency rates are aided by New Hampshire's confidence interval for elementary and middle schools, respectively. On these figures, the darker portion of the bars

show the actual proficiency rates at each school, and the lighter portion of the bars show the degree to which these proficiency rates are increased by the application of the confidence interval. The orange lines show the AMO needed to meet AYP. These figures show that four elementary schools (Few, Island Grove, Nemo, and Wolf Creek) and two middle schools (Hoyt and Lake Joseph) were assisted by the confidence intervals to meet their overall targets in math (note how the orange line falls within the light blue band); all of these schools, however, still failed to make AYP because of low subgroup performance (see Figures 3 and 4).

The effect of the confidence intervals on reading proficiency rates at the elementary and middle school levels is much the same (not shown). In reading, six elementary schools (Nemo, Island Grove, JFK, Scholls, Wolf Creek, and Coastal) and two middle schools (Pogesto and Lake Joseph) met their overall targets with the help of the confidence interval. However, we know from Fig-
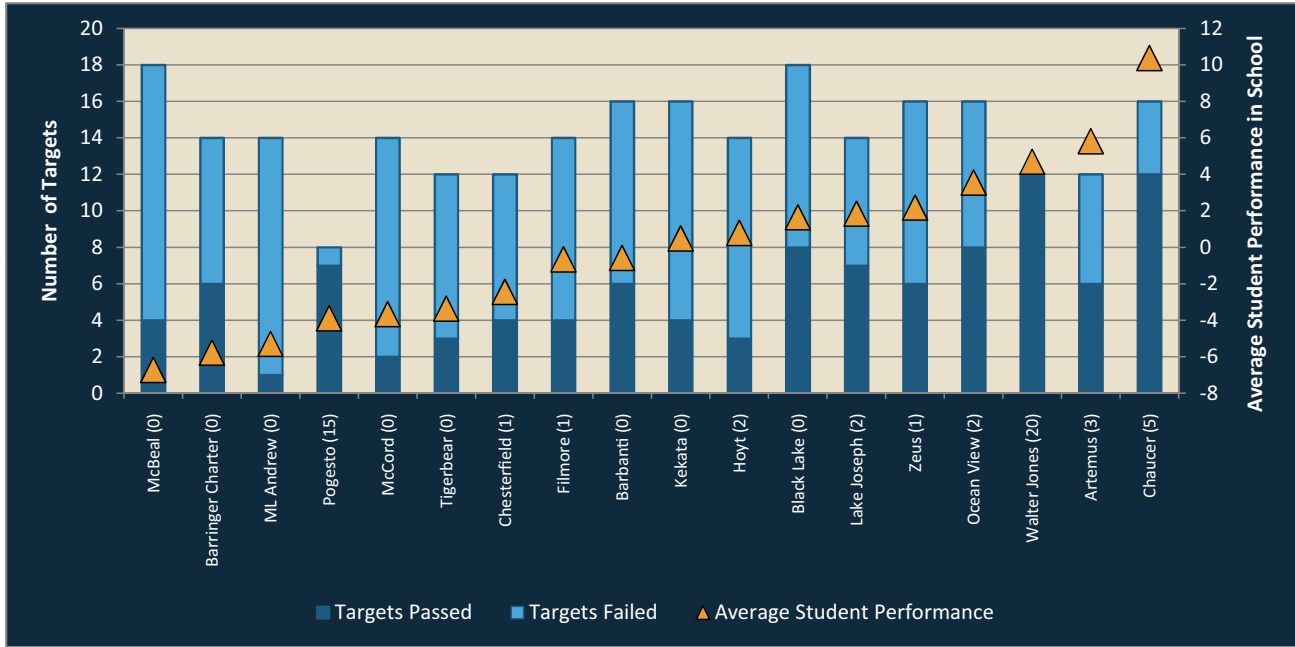
Figure 4. AYP performance of the middle school sample under New Hampshire's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under New Hampshire'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. Pogesto, for example, met 7 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; 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 New Hampshire'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 four of the elementary schools (Few, Island Grove, Nemo, and Wolf Creek) were assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.
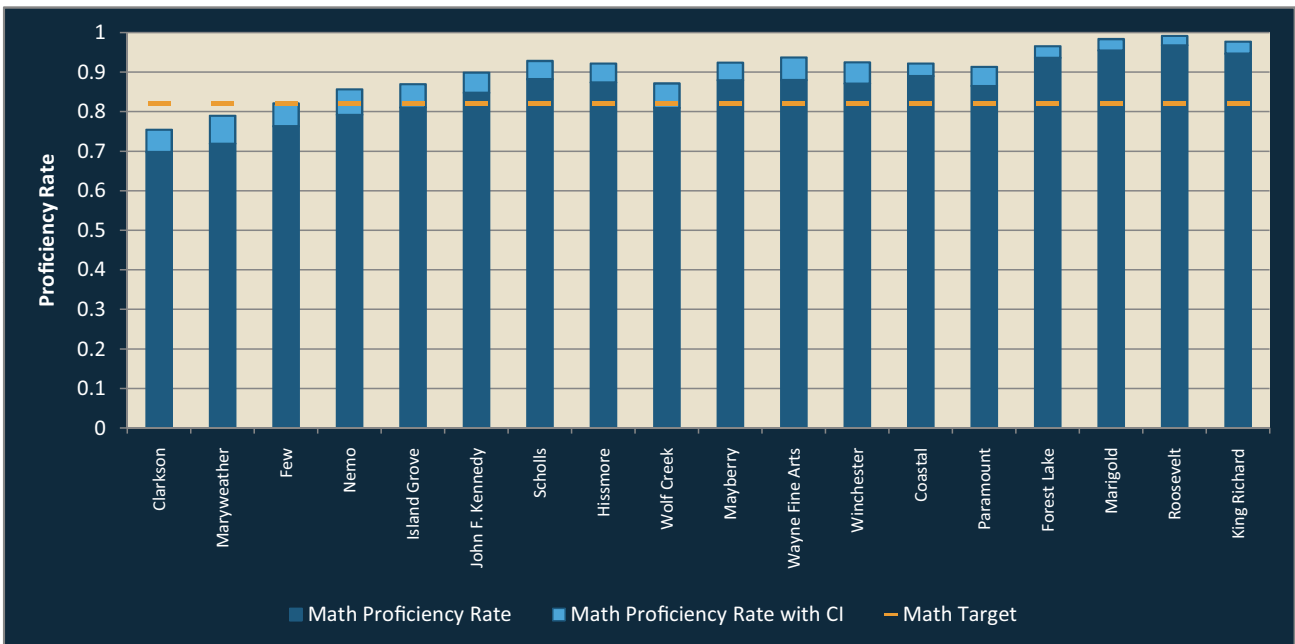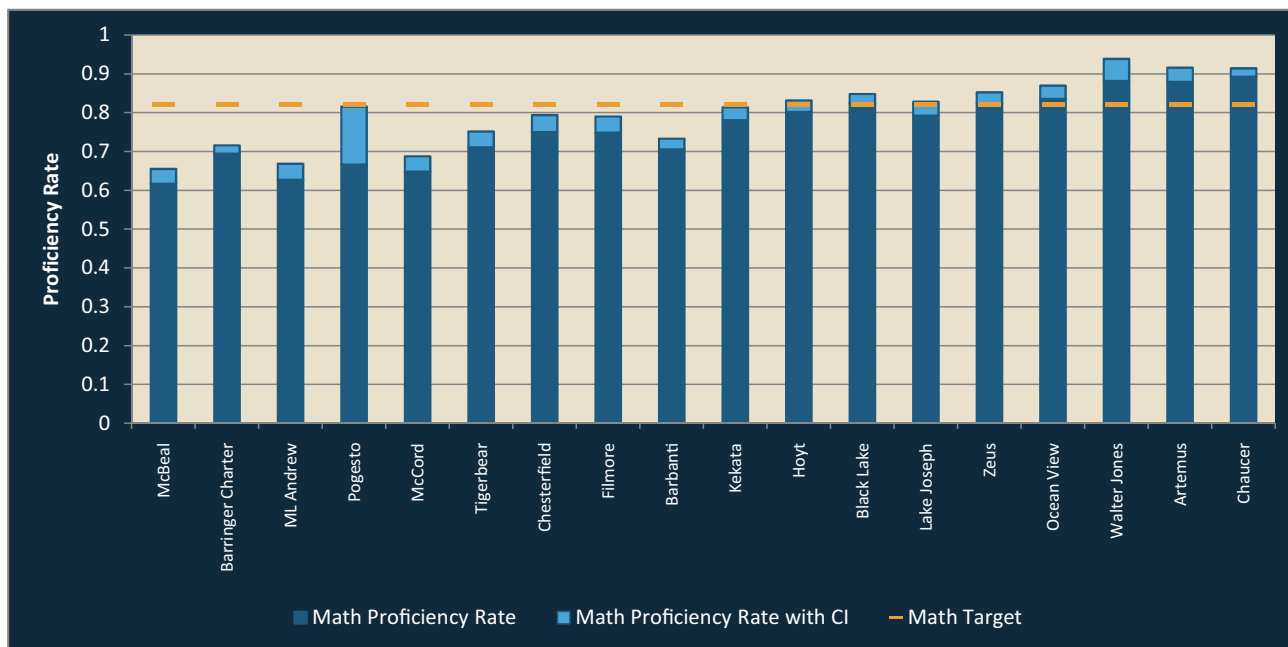
Figure 6. Impact of the confidence interval on middle school math proficiency rates under New Hampshire's 2008 AYP rules

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 (Hoyt and Lake Joseph) were assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.

ures 3 and 4 that all these schools failed to meet their targets for some subgroups. **Overall, the application of the confidence interval, despite the fact that it is lenient, seems to have little or no effect on AYP outcomes for the sample elementary and middle schools in New Hampshire.**[9]

Where Do Schools Fail?

Figures 3 and 4 illustrate the number of subgroup targets at the sample elementary and middle schools and the number of targets met in New Hampshire. However, these figures do not indicate which subgroups passed or failed in each 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 New Hampshire 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:

- Only two elementary schools (Clarkson and Maryweather) failed to meet both the reading and the math targets for their overall school population.

- About half of the middle schools failed in both reading and math for their overall student populations.

---

[9] 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 is likely larger than the impact depicted in Figures 5 and 6. However, we chose not to show how the confidence interval impacted subgroup performance because it would have added greatly to the report's length and complexity.

Table 2. Elementary school subgroup performance of sample schools under the 2008 New Hampshire 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	69.8%	66.5%	N	N	N	N	N	N	N	N					N	N			Y	Y	12	2	17%	N	1
Maryweather	72.0%	69.6%	N	N	N	N	N	N	N	N	Y	Y			N	N	Y	Y	Y	Y	16	6	38%	N	1
Few	76.4%	72.9%	Y	N	N	N	N	N	N	N	Y	Y			Y	N	Y	Y	Y	Y	16	8	50%	N	1
Nemo	79.3%	83.7%	Y	Y	N	N			N	N	N	N			Y	Y			Y	Y	12	6	50%	N	7
Island Grove	81.1%	82.2%	Y	Y	N	N	N	N	Y	Y					Y	N			Y	Y	12	7	58%	N	4
JFK	84.8%	81.1%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	3
Scholls	88.3%	84.2%	Y	Y	N	N	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	12	86%	N	7
Hissmore	87.5%	86.3%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Wolf Creek	81.0%	83.6%	Y	Y	N	N	N	N	Y	Y			Y	Y	N	Y			Y	Y	14	9	64%	N	5
Alice Mayberry	88.0%	88.3%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	9
Wayne Fine Arts	88.0%	93.9%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	14	100%	Y	21
Winchester	87.2%	90.1%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	16	16	100%	Y	22
Coastal	89.1%	85.1%	Y	Y	N	N	N	N	Y	N	Y	N			Y	N			Y	Y	14	7	50%	N	3
Paramount	86.5%	86.5%	Y	Y	N	Y	N	N	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	18	13	72%	N	7
Forest Lake	93.7%	93.3%	Y	Y	Y	N			Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	14	13	93%	N	8
Marigold	95.5%	92.5%	Y	Y	Y	Y	Y	N	Y	Y			Y	Y	Y	N			Y	Y	14	12	86%	N	10
Roosevelt	96.8%	96.9%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	14	100%	Y	28
King Richard	94.7%	94.5%	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	Y	Y			Y	Y	14	14	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.

■ Four elementary schools (Scholls, Hissmore, Alice Mayberry, and Forest Lake) met every target except for their SWDs.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively. We see that the performance of SWDs is proving very challenging for schools under New Hampshire's system, particularly in middle schools, where this subgroup tends to have enough students to meet the state's minimum *n* of 11. The same is true for students with limited English proficiency. In fact, all but one middle school (Walter Jones) in the study with qualifying

SWD and two middle schools (Barringer Charter and McCord Charter) with qualifying LEP subgroups failed to meet their targets for these subgroups in reading or math. Low-income students are also struggling to meet the state's targets. Most middle schools with a large enough low-income population to qualify as a separate subgroup failed to meet their reading and math targets for these students (recall that proficiency cut scores in math and reading are generally lower at the elementary than the middle school level).

Other state reports contain a section comparing some of

THOMAS B. FORDHAM INSTITUTE

Table 3. Middle school subgroup performance of sample schools under the 2008 New Hampshire 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	61.8%	68.7%	N	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N	N	Y	Y	18	4	22%	N	0
Barringer Charter	69.4%	76.9%	N	N	N	N	Y	Y	N	N	N	N			Y	Y			Y	Y	14	6	43%	N	0
ML Andrew	62.8%	75.6%	N	N	N	N	N	N	N	N	N	N			N	N			N	Y	14	1	7%	N	0
Pogesto	66.7%	78.9%	N	Y							Y	Y			Y	Y			Y	Y	8	7	88%	N	15
McCord Charter	64.8%	77.8%	N	N	N	N	N	Y	N	N	N	N			N	N			N	Y	14	2	14%	N	0
Tigerbear	71.1%	72.6%	N	N	N	N					N	N	N	N	Y	N			Y	Y	12	3	25%	N	0
Chesterfield	75.0%	76.8%	N	N	N	N					N	N	N	N	Y	Y			Y	Y	12	4	33%	N	1
Filmore	74.9%	82.0%	N	N	N	N	N	N	N	N			Y	Y	N	N			Y	Y	14	4	29%	N	1
Barbanti	70.5%	77.3%	N	N	N	N	N	N	N	N	Y	Y	Y	Y	N	N			Y	Y	16	6	38%	N	0
Kekata	78.1%	79.7%	N	N	N	N	N	N	N	N	N	N	Y	Y	N	N			Y	Y	16	4	25%	N	0
Hoyt	80.2%	82.1%	Y	N	N	N	N	N	N	N	N	N			N	N			Y	Y	14	3	21%	N	2
Black Lake	82.4%	81.8%	Y	N	N	N	N	N	N	N	N	N	Y	Y	Y	N	Y	Y	Y	Y	18	8	44%	N	0
Lake Joseph	79.3%	84.8%	Y	Y	N	N	N	N	N	Y	Y	Y			N	N			Y	Y	14	7	50%	N	2
Zeus	82.4%	83.1%	Y	N	N	N	N	N	N	N	Y	N	Y	Y	N	N			Y	Y	16	6	38%	N	1
Ocean View	83.5%	89.3%	Y	Y	N	N	N	N	N	N	Y	Y	Y	Y	N	N			Y	Y	16	8	50%	N	2
Walter Jones	88.1%	89.9%	Y	Y	Y	Y			Y	Y	Y	Y			Y	Y			Y	Y	12	12	100%	Y	20
Artemus	87.9%	87.7%	Y	Y	N	N			N	N			Y	Y	N	N			Y	Y	12	6	50%	N	3
Chaucer	89.3%	92.5%	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.

the characteristics of the sample schools that made AYP versus those that did not. In New Hampshire, there were no striking differences between schools that made AYP and those that didn't, either at the elementary or middle school level. The one exception (rather expected) was that schools that made AYP had students with higher average performance than did schools that didn't make it, as measured by NWEA reading and math tests.[10]

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 New Hampshire's AYP rules (and AMOs) for 2008. We found that only 4 elementary schools and 1 middle school—just 5 out of a sample of 36—would have made AYP in New Hampshire. Looking across the 28 state ac-

---

[10] There were also no "anomalies" in New Hampshire. All the sample schools that made AYP in New Hampshire made it in the other states examined; similarly, sample schools that failed to make AYP in New Hampshire tended to fail in most other states as well.

Table 4. Summary of subgroup performance of sample elementary schools under 2008 New Hampshire 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	18	12	12
Students with limited English proficiency	13	7	8
Low-income students	18	4	7
African-American students	13	1	3
Asian/Pacific Islander students	6	0	0
Hispanic students	15	3	7
American Indian/Alaska Native students	3	0	0
White students	18	0	0

Table 5. Summary of subgroup performance of sample middle schools under the 2008 New Hampshire 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	17	16	16
Students with limited English proficiency	13	12	11
Low-income students	18	15	14
African-American students	15	9	10
Asian/Pacific Islander students	9	0	0
Hispanic students	18	11	13
American Indian/Alaska Native students	2	1	1
White students	18	2	0

countability systems examined in the study, this puts New Hampshire roughly in the middle of the sample distribution in terms of the number of schools making AYP (see Figure 1). So, although New Hampshire awards "partial credit" to students performing at lower levels and uses a fairly lenient confidence interval (margin of error), most schools still failed to make AYP, partly because New Hampshire's small minimum n size (which makes more

subgroups accountable) and partly because of New Hampshire's fairly high annual targets or AMOs.

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, New Hampshire's NCLB account-

THOMAS B. FORDHAM INSTITUTE

ability 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 minority youngsters. Some of the sample schools met the New Hampshire reading and math targets for their student populations as a whole, that is, without considering subgroup results. In the pre-NCLB era, such schools might have been considered effective or at least not in need of improvement, even though sizable numbers of their students aren'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 like-lihood of making AYP? Is it "fair" that, in New Hampshire and in a handful of other states, students are awarded "partial" credit even though they do not achieve proficiency? Even if actual participation guidelines for English language learners and SWDs are more generous under the current state assessment system,[11] doesn't the massive failure of these students to meet New Hampshire'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 pupils, 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, 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.

[11] See footnote 4.