New Jersey

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 towards these goals.

This report examines New Jersey'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 Jersey's system is compared with other states. We selected 36 schools from around the nation, schools that vary by size, achievement, and diversity, among other factors, and determined whether or not each would make AYP under New Jersey'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 Jersey's AYP rules for academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- We estimate that **15 of 18 elementary schools** in our sample **failed to make AYP** in 2008 under New Jersey'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. **It's also likely due to New Jersey's low minimum *n* size of 20 (for most subgroups) and its fairly high annual targets, especially in reading**.

- Looking across the 28 state accountability systems examined in the study, we find that the number of elementary schools making AYP in New Jersey was exceeded by 15 other sample states (New Jersey ties with 4 other states that each have 3 elementary schools making AYP). This puts New Jersey in the lower part of the sample distribution (see Figure 1). (Note that middle schools were not examined in New Jersey, unlike other states, since eighth grade cut scores were not available.)

- Most of the schools in our sample that fail to make AYP in New Jersey are meeting expected targets for their overall populations but failing because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.[2]

- As is the case in other states, schools with fewer subgroups attain AYP more easily in New Jersey 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.

New Jersey falls near the middle of the state distribution in terms of the number of schools that make AYP. One particularly interesting thing about New Jersey is that a large group of Hispanic/Latino, African American, and low-income students met their targets in math. This is unusual because New Jersey's minimum subgroup size for these groups (20) is smaller than most other states', meaning that schools in New Jersey are held accountable for more subgroups than would similar schools in other states. However, New Jersey's definitions of proficiency generally ranked below average compared with the standards set by the other states, especially in grades 3-5 math. This likely accounts for the higher pass rate for traditionally disadvantaged groups.

---

[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 Jersey Assessment of Skills and Knowledge (NJ ASK).

[2] SWDs are defined as those students following individualized education plans.
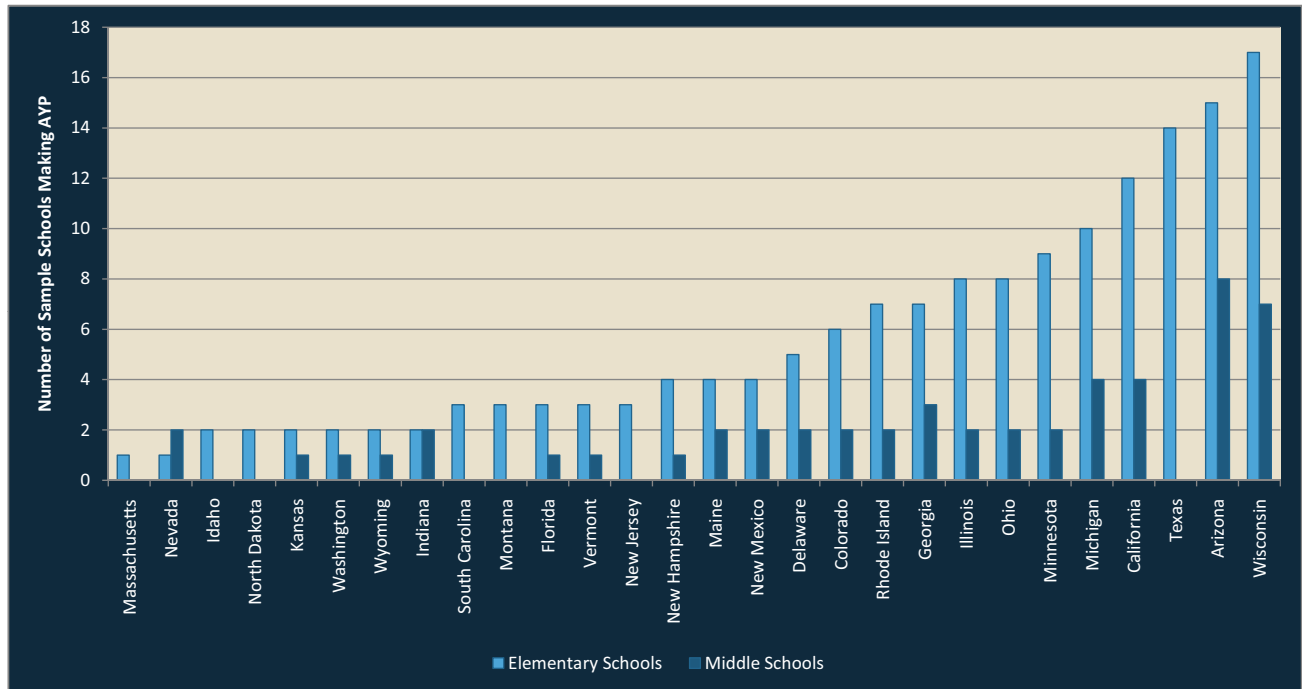
Thomas B. Fordham Institute

Figure 1. Number of sample schools making AYP by state

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

- A strong predictor of a school making AYP under New Jersey's system is whether it has enough English language learners and SWDs to qualify as separate subgroups. Every single elementary school with limited English proficient and SWD subgroups failed to make AYP, in part because these students did not meet the state's targets in reading.[3]

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance on New Jersey's tests and 25 other state tests to the Northwest Evaluation Association's Measures of Academic Progress (MAP), a computerized adaptive test used in schools nationwide. This single common scale permitted cross-state comparisons of each state's reading and math proficiency standards to measure school performance under the No Child Left Behind (NCLB) Act

of 2001. That study revealed profound differences in states' proficiency standards (i.e., how difficult it is to achieve proficiency on the state test), and even across grades within a single state.

Our study expands on *The Proficiency Illusion* to examine 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 estimate 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-perform-

---

[3] It should be noted that our subgroup findings for Limited English Proficient (LEP) and students with disabilities may be slightly more negative than would be seen under real world conditions. This is mostly due to the differences in testing practices between how LEP students and students with disabilities are treated in the NWEA's Measures of Academic Progress (MAP), the assessment used in this study, and in the New Jersey Assessment of Skills and Knowledge (NJ ASK), the state standardized assessment. Specifically, the U.S. Department of Education has issued NCLB guidelines permitting schools to exclude small percentages of LEP or disabled students from taking state tests, or providing them alternate assessments. In the current study, however, no valid MAP scores were omitted from consideration.

ing 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 school year 2013–14. 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 (low income[4] 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 both 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.) As a result of this flexibility, states have set widely varying n sizes for their subgroups, from as few as ten 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. This means that a school making 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 2005-06 academic year from 18 elementary and 18 middle schools around the country. We also collected the NCLB subgroup designations for all students in those schools—in other words, whether they had been classified as members of a minority group such as English language learners,[5] 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]
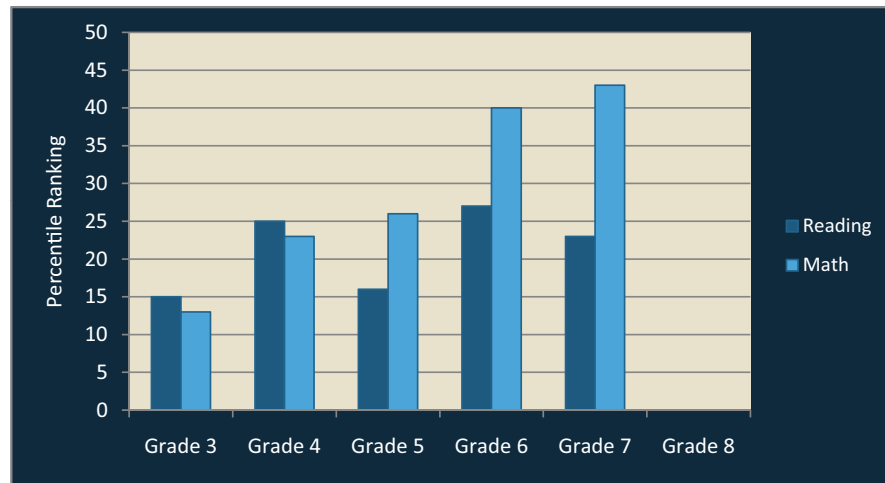
Proficiency cut score estimates for the New Jersey Assessment of Skills and Knowledge (NJ ASK) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that New Jersey's definitions of proficiency generally ranked below average compared with the standards set by the other 25 states in that study. These cut scores were used to estimate whether students would have scored as proficient or better on the New Jersey test, given their performance on MAP. Student test data and subgroup designations are then used to determine how these 18 elementary schools would have fared under New Jersey AYP rules for 2008. In other words, the school data and our proficiency cut score estimates are from academic year 2005–06, but we are applying them against New Jersey's 2008 AYP rules. Note that in New Jersey, unlike most of the other state reports, the 18 sample middle schools were not examined since New Jersey's eighth grade cut scores were not available.

Table 1 shows the pertinent New Jersey AYP rules that were applied to elementary schools in the current study.

---

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

[5] Note that we use "students with limited English proficiency (LEP)" or "LEP students" and "English language learners" interchangeably to refer to students in the same subgroup.

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

Thomas B. Fordham Institute

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

Note: This figure illustrates the difficulty of New Jersey's cut scores (or proficiency passing scores) for its reading and math tests, as percentiles of the NWEA norm, in grades three through seven. Cut scores were not available for grade eight. Higher percentile ranks are more difficult to achieve. All of New Jersey's cut scores are below the 45th percentile. Cut score estimates for 8th grade were not available.

New Jersey's minimum subgroup size is 20 for all groups except for SWDs which is 35. While 35 is fairly consistent with the sizes used by most other states, 20 is smaller than most.[7] This means that schools in New Jersey will be accountable for more subgroups than would similar schools in other states.

Most states also apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates. The 95% confidence interval applied to proficiency rate calculations in New Jersey is comparable to the majority of states examined in the study. So, for instance, though schools are supposed to get 82% of their grade 3 students (as well as 82% of their students in each subgroup) to the proficient level on the state reading test, applying the confidence interval means that the real target can actually be lower, particularly with smaller groups.

Note that we were not able 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 is 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 school's subgroup—to participate in testing.

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

---

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

Table 1. New Jersey AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 20	
	SWDs: 35	
	Low-income students: 20	
	LEP students: 20	
CI	**Applied to proficiency rate calculations?**	
	Yes; 95% CI used	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	n/a	82
Grade 4	68	82
Grade 5	n/a	82
Grade 6	n/a	76
Grade 7	n/a	76
Grade 8	58	76
MATH		
Grade 3	n/a	73
Grade 4	53	73
Grade 5	n/a	73
Grade 6	n/a	62
Grade 7	n/a	62
Grade 8	39	62

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

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

Figure 3 illustrates the AYP performance of the sample elementary schools under New Jersey's 2008 AYP rules. **Only three elementary schools made AYP (Wayne Fine Arts, Winchester, and Roosevelt) while fifteen did not**. 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 the higher performing students were found at these schools.

Yet among these high performing schools, the only schools actually to make AYP are those with relatively few qualifying subgroups—and thus the fewest targets to meet (because each subgroup has separate targets). For example, Winchester passed, but has only nine targets. Among the eighteen elementary schools, this school has the fewest subgroups in New Jersey (along with Clarkson).

Figures 4 and 5 indicate the degree to which elementary schools' reading and math proficiency rates are aided by New Jersey's confidence interval. On these figures, the dark blue bars show the actual proficiency rates at each school, and the light blue bars show the degree to which these proficiency rates were increased by applying the
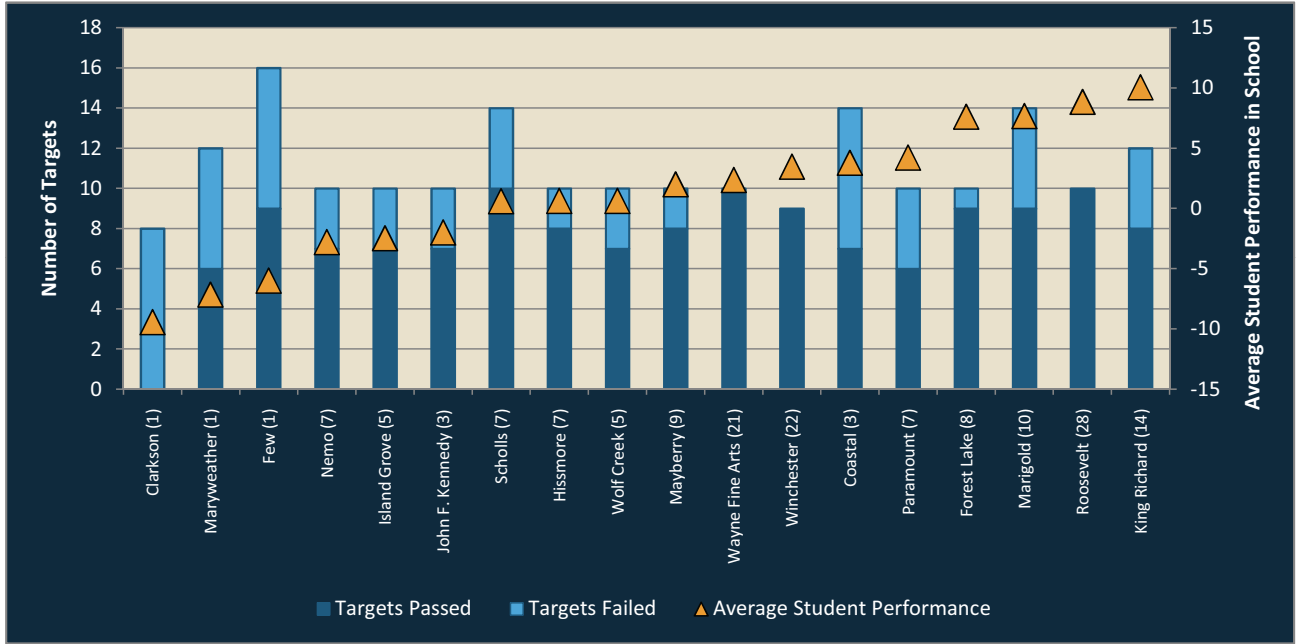
THOMAS B. FORDHAM INSTITUTE

Figure 3. Performance of the elementary school sample under New Jersey's 2008 AYP Rules

Note: This figure indicates how each of the elementary schools within the sample fared under the New Jersey 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. Forest Lake, for example, meets nine 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.
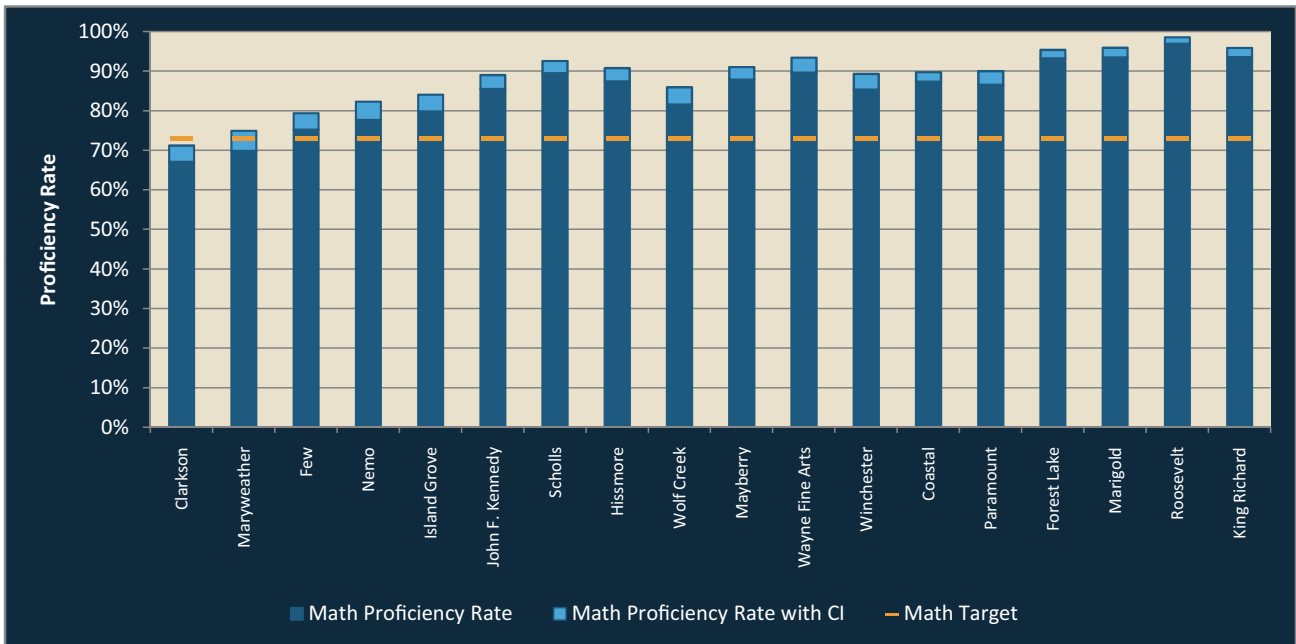

Figure 4. Impact of the confidence interval on elementary school math proficiency rates for 2008

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that one of the sample elementary schools (Maryweather) 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.
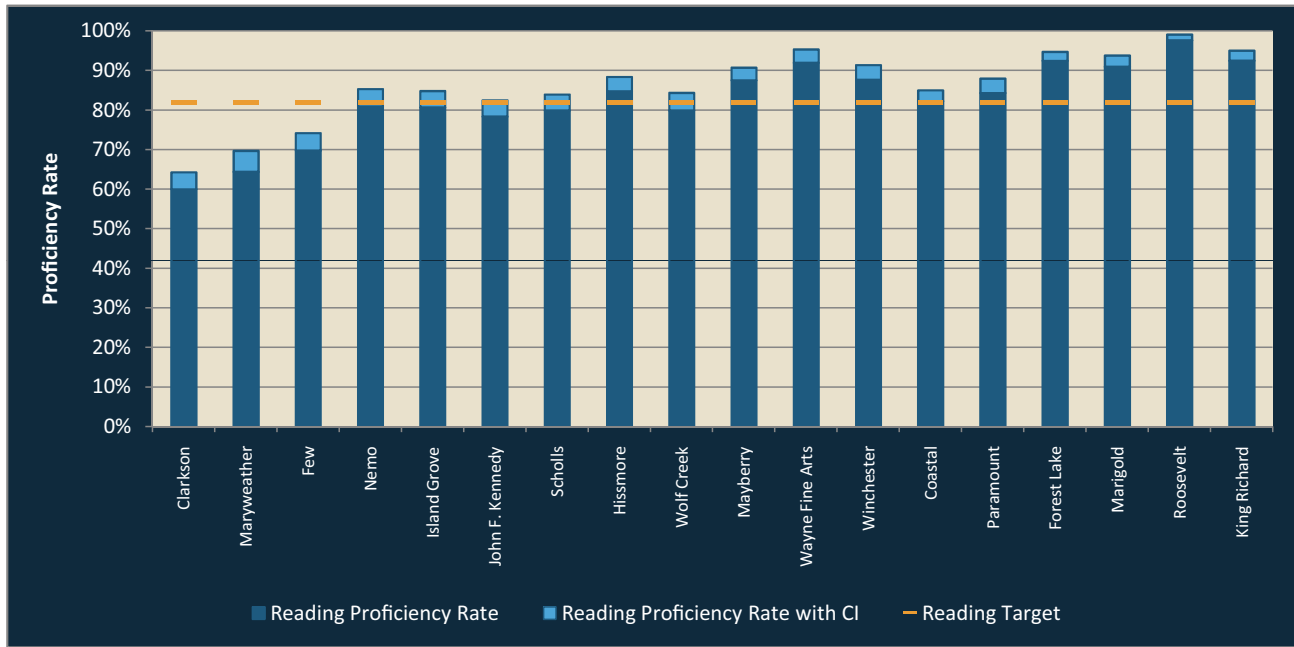
Figure 5. Impact of the confidence interval on elementary school reading proficiency rates for 2008

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 sample elementary schools (Nemo, Island Grove, Scholls, 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.

confidence interval. The orange lines show the annual measurable objective needed to meet AYP. Figure 4 shows that one of the sample elementary schools (Maryweather) met its overall math target with the assistance of the confidence interval (note how the orange bar falls in the light blue band). In reading (Figure 5), four schools (Nemo, Island Grove, Scholls, and Wolf Creek) were able to achieve their overall targets when assisted by the confidence interval. All of these schools, however, still fail to make AYP because of low subgroup performance (shown in Figure 3). **Overall, the application of the confidence interval had no effect on whether the sample schools met their overall reading or math targets in New Jersey.**[8]

Where do schools fail?

Figure 3 illustrates how the number of subgroups can impact the AYP decisions for our sample schools, but it conveys no information about which subgroups failed or passed in which school. Table 2 lists information on individual subgroup performance.

Table 2 shows 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 purposes 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. Table 2 also shows whether a school made AYP under the New Jersey rules, and the total number of states within the study in which that school met AYP.

---

[8] In the current analyses, confidence intervals were applied to both the overall school population and to all eligible subgroups in our sample schools. Thus, the ultimate impact of the confidence interval is likely larger than the impact depicted in Figures 4 and 5. However, we chose not to show how the confidence interval impacted subgroup performance because it would have added greatly to the report's length and complexity.

Table 2. Elementary subgroup performance of sample schools under the 2008 New Jersey AYP rules

SCHOOL PSEUDONYM	Overall Proficiency Rate (Math)	Overall Proficiency Rate (Reading)	Overall M	Overall R	SWDs M	SWDs R	LEP Students M	LEP Students R	Low-income Students M	Low-income Students R	AA M	AA R	Asian M	Asian R	Hispanic M	Hispanic R	AI/AN M	AI/AN R	White M	White R	AYP Targets Required	Targets MET	% of Targets Met	School Met AYP?	Number of states in which school met AYP?
Clarkson	67.1%	60.0%	N	N			N	N	N	N					N	N					8	0	0%	N	1
Maryweather	69.9%	64.4%	Y	N			N	N	Y	N	Y	Y			N	N			Y	Y	12	6	50%	N	1
Few	75.3%	69.8%	Y	N	N	N	N	N	Y	N	Y	Y			Y	N	Y	Y	Y	Y	16	9	56%	N	1
Nemo	77.7%	80.9%	Y	Y					Y	N	N	N			Y	Y			Y	Y	10	7	70%	N	7
Island Grove	79.8%	80.7%	Y	Y			N	N	Y	Y					Y	N			Y	Y	10	7	70%	N	4
JFK	85.5%	78.4%	Y	Y	Y	N			Y	N	Y	N							Y	Y	10	7	70%	N	3
Scholls	89.6%	79.9%	Y	Y	Y	N	Y	N	Y	N	Y	N			Y	Y			Y	Y	14	10	71%	N	7
Hissmore	87.5%	84.7%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Wolf Creek	81.7%	79.9%	Y	Y			Y	N	Y	N					Y	N			Y	Y	10	7	70%	N	5
Alice Mayberry	87.9%	87.5%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	9
Wayne Fine Arts	89.7%	92.0%	Y	Y					Y	Y	Y	Y			Y	Y			Y	Y	10	10	100%	Y	21
Winchester	85.4%	87.7%	Y	Y					Y	Y				Y	Y	Y			Y	Y	9	9	100%	Y	22
Coastal	87.4%	82.3%	Y	Y	N	N	N	N	Y	N	Y	N			Y	N			Y	Y	14	7	50%	N	3
Paramount	86.6%	84.3%	Y	Y			N	N	Y	N					Y	N			Y	Y	10	6	60%	N	7
Forest Lake	93.3%	92.5%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	8
Marigold	93.5%	91.0%	Y	Y	Y	N	Y	N	Y	N			Y	Y	N	N			Y	Y	14	9	64%	N	10
Roosevelt	97.0%	97.6%	Y	Y					Y	Y	Y	Y			Y	Y			Y	Y	10	10	100%	Y	28
King Richard	93.6%	92.5%	Y	Y	Y	N	Y	N	Y	N					Y	N			Y	Y	12	8	67%	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.

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

- Three elementary schools (Clarkson, Maryweather, and Few) failed to meet the reading targets for their overall school population. Only one elementary school (Clarkson) failed to meet its overall target in math.

- Three elementary schools (Hissmore, Alice Mayberry, and Forest Lake) met all their reading and math targets for all subgroups except for their SWDs.

- Most low-income students met their math but not their reading targets (perhaps because reading cut scores are generally higher than math in the lower grades, as are annual targets in reading).

Table 3 summarizes the performance of the various subgroups. As shown, the performance of SWDs is particularly challenging within our sample schools. Every school within the sample with sufficient numbers of students with disabilities to qualify as a subgroup failed to meet its reading targets (this was also true for students with limited English proficiency.)

Table 3. Summary of subgroup performance of sample elementary schools under the 2008 New Jersey 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	9	4	9
Students with limited English proficiency	10	6	10
Low-income students	18	1	11
African-American students	11	1	4
Asian/Pacific Islander students	1	0	0
Hispanic students	14	3	9
American Indian/Alaska Native students	1	0	0
White students	17	0	0

Other state reports contain a section comparing some of the characteristics of the sample schools that made AYP versus those that did not. In New Jersey, there were no striking differences between schools that did and didn't make AYP at the elementary level, 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 schools across the country to see how they would fare under New Jersey's AYP rules (and AMOs) for 2008. We found that only three elementary schools would have made AYP in New Jersey. Looking across the 28 state accountability systems examined in the study, this puts New Jersey in the lower middle of the sample distribution in terms of schools making AYP (see Figure 1). Part of this may be due to New Jersey's low minimum n of 20 (for non-SWD subgroups) and its fairly high annual performance targets, especially in reading.

The overriding goal of the No Child Left Behind act (NCLB) is to eliminate educational disparities within and across states; it's important to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, New Jersey's No Child Left Behind accountability system is working exactly as Congress intended: identifying as "needing attention" schools with relatively high test score averages that mask low performance for particular groups of students, such as low-income or Hispanic students. Many of the sample schools make AYP in New Jersey for their student populations as a whole, i.e., 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 pupils weren't meeting state standards. Disaggregating data by race, income, and so on has made those students visible. That is surely a positive step.

Yet NCLB's design flaws are also readily apparent. Does it make sense that having fewer subgroups enhances the likelihood of making AYP? Even if actual participation guidelines for English language learners and students with disabilities are more generous under the current state assessment system,[9] doesn't the mas-

---

[9] See footnote 3.

sive failure of these students to meet New Jersey'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 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 re-authorization 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.