



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 Delaware’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 Delaware’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 Delaware’s system as well as under the systems 27 other states. We used school data and proficiency cut score¹ estimates from academic year 2005–2006, but applied them against Delaware’s AYP rules for academic year 2007–2008 (shortened to “2008” in this report).

Here are some key findings:

- We estimate that **13 of 18 elementary schools** and **16 of 18 middle schools** in our sample **failed to make AYP** in 2008 under Delaware’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.)

- Looking across the 28 state accountability systems examined in the study, **we find that the number of elementary schools making AYP in Delaware was exceeded in 11 other sample states, putting Delaware roughly in the middle of the sample distribution** (see Figure 1).²
- Nearly all the schools in our sample that failed to make AYP in Delaware are meeting expected targets for their overall populations but failed to make AYP because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.³
- One sample school (Alice Mayberry) that failed to make AYP in most other states made AYP in Delaware.

Looking across the 28 state accountability systems examined in the study, we find **Delaware** near the middle of the distribution in terms of how many sample schools make AYP. Delaware’s mix of rules means that several schools make AYP in Delaware that do not in most of the other 27 states. This is likely due to the fact that Delaware’s proficiency standards (or cut scores) are relatively easy compared to other states. However, Delaware’s annual targets (i.e., the percentage of students in various subgroups who have to meet proficiency) in reading are relatively difficult to achieve. Specifically, 68 percent of a given population in any school would have to be proficient on the state reading exam for the school to make AYP in 2008. Every single school with a limited English proficient (LEP) subgroup failed to make AYP in Delaware, in part because these students did not meet the state’s proficiency targets in reading or/math.

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 Delaware Student Testing Program.

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

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

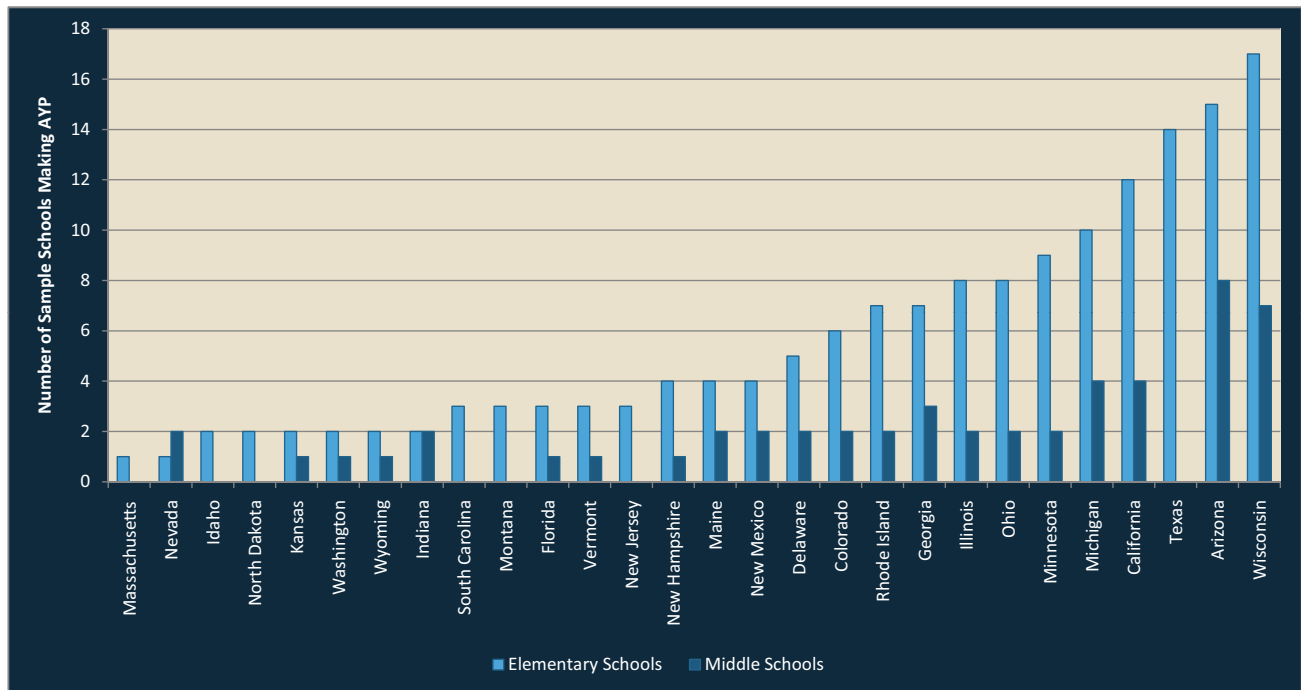


Figure 1. Number of sample schools making AYP by state

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.

This is probably because **Delaware's proficiency standards are relatively easy compared to other states.**

- In Delaware, as in most states, schools with fewer subgroups attain AYP more easily in Delaware than schools with more subgroups, even when their average student performance is much lower. In other words, schools with greater diversity and size face greater challenges in making AYP.
- As in other states, middle schools have greater difficulty reaching AYP in Delaware than do elementary schools, primarily because their student populations are larger and therefore have more qualifying subgroups—not because their student achievement is lower than in the elementary schools.
- A strong predictor of a school making AYP under

Delaware’s system is whether it has enough English language learners to qualify as a separate subgroup. Every school with a subgroup of students with limited English proficiency (LEP)⁴ failed to make AYP, in part because these students did not meet the state’s proficiency targets in reading and/or math. Likewise, many schools with enough qualifying students with disabilities (SWDs) failed to meet their AYP targets.⁵

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance on Delaware’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 compar-

⁴ Note that we use “LEP students” and “English language learners” interchangeably to refer to students in the same subgroup.

⁵ 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 Delaware Student Testing Program (DSTP), 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.

isons of each state’s reading and math proficiency standards to measure school performance under the No Child Left Behind (NCLB) Act of 2001. That study revealed profound differences in states’ proficiency standards (i.e., how difficult it is to achieve proficiency on the state test), and even across grades within a single state.

Our study expands on *The Proficiency Illusion* by examining other key factors of state NCLB accountability plans and how they interact with state proficiency standards to determine whether the schools in our sample made adequate yearly progress (AYP) in 2008. Specifically, we estimated how a single set of schools, drawn from around the country, would fare under the differing rules for determining AYP in 28 states (the original 25 in *The Proficiency Illusion* plus 3 others for which we now have cut score estimates). In other words, if we could somehow move these entire schools—with their same mix of characteristics—from state to state, how would they fare in terms of making AYP? Will schools with high-performing students consistently make AYP? Will schools with low-performing students consistently fail to make AYP? If AYP determinations for schools are not consistent across states, what leads to the inconsistencies?

NCLB requires every state, as a condition of receiving Title I funding, to implement an accountability system that aims to get 100% of its students to the proficient level on the state test by academic year 2013–2014. In the intervening years, states set annual measurable objectives (AMOs). This is the percentage of students in each school, and in each subgroup within the school (such as low income⁶ 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 re-

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

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

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

What We Studied

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

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

Proficiency cut score estimates for the Delaware Student Testing Program (DSTP) are taken from *The Proficiency*

⁶ Low-income students are those who receive a free or reduced-price lunch.

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

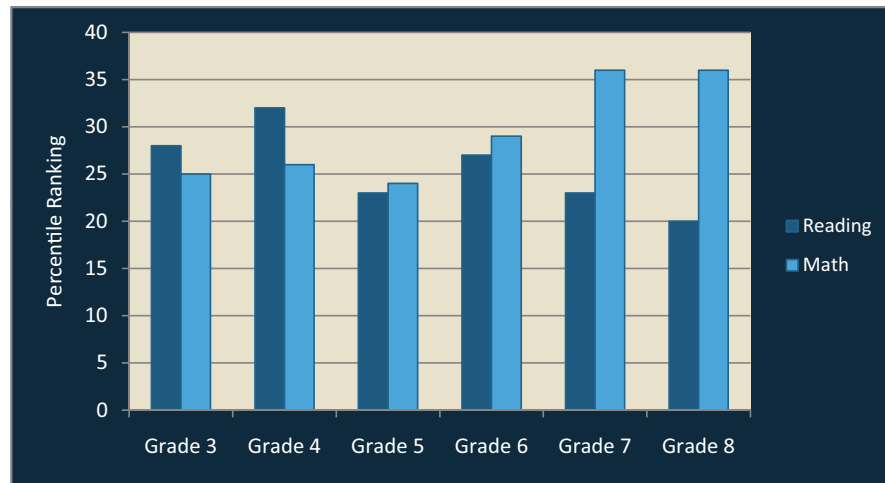


Figure 2. Delaware reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Delaware’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 Delaware’s cut scores are below the 40th percentile.

Illusion (as shown in Figure 2), which found that Delaware’s definitions of proficiency generally ranked below the 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 Delaware test, given their performance on MAP. Student test data and subgroup designations are then used to determine how these 18 elementary and 18 middle schools would have fared under Delaware AYP rules for 2008. In other words, the school data and our proficiency cut score estimates are from academic year 2005–2006, but we are applying them against Delaware’s 2008 AYP rules.

Table 1 shows the pertinent Delaware AYP rules that were applied to elementary and middle schools in this study. Delaware’s minimum subgroup size is 40, which is comparable to most other states we examined.⁸ Furthermore, although most states examined in the study apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates, **Delaware’s 98% confidence interval gives schools greater leniency than the 95% confidence interval used by most other states.** So, for instance, though schools are supposed to get 68% of their students (as well as 68% 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 unable to examine the effect of NCLB’s “safe harbor” provision. This provision permits a school to make AYP even if some of its subgroups fail, as long as it reduces the number of nonproficient students within any failing subgroup by at least 10% relative to the previous year’s performance. Because we had access to only a single academic year’s data (2005–2006), we were not able to include this in our analysis. As a result, it 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 inter-

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

Table 1. Delaware AYP rules for 2008

| Subgroup minimum <i>n</i> | Race/ethnicity: 40 | |
|------------------------------|--|------------------|
| | SWDs: 40 | |
| | Low-income students: 40 | |
| | LEP students: 40 | |
| CI | Applied to proficiency rate calculations? | |
| | Yes; 98% CI | |
| AMOs | Baseline proficiency levels as of 2002 (%) | 2008 targets (%) |
| READING/LANGUAGE ARTS | | |
| Grade 3 | 62 | 68 |
| Grade 4 | 62 | 68 |
| Grade 5 | 62 | 68 |
| Grade 6 | 62 | 68 |
| Grade 7 | 62 | 68 |
| Grade 8 | 62 | 68 |
| MATH | | |
| Grade 3 | 41 | 50 |
| Grade 4 | 41 | 50 |
| Grade 5 | 41 | 50 |
| Grade 6 | 41 | 50 |
| Grade 7 | 41 | 50 |
| Grade 8 | 41 | 50 |

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

vals) 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 Delaware’s AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Delaware’s 2008 AYP rules. **Only 5 schools made AYP and 13 failed to make AYP.** The triangles in Figure 3 show the average academic performance of students within the school, with negative

values indicating below-grade-level performance for the average student and positive values indicating above-grade-level performance. 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 almost without regard to average student performance, the only schools actually to make AYP were those with relatively few qualifying subgroups—and thus the fewest targets to meet (because each subgroup has separate targets). For example, Wayne Fine Arts and Winchester passed, but had only four targets each. Each school must make AYP for its overall student population in reading and math (two targets) and for its white population resulting in four total targets.

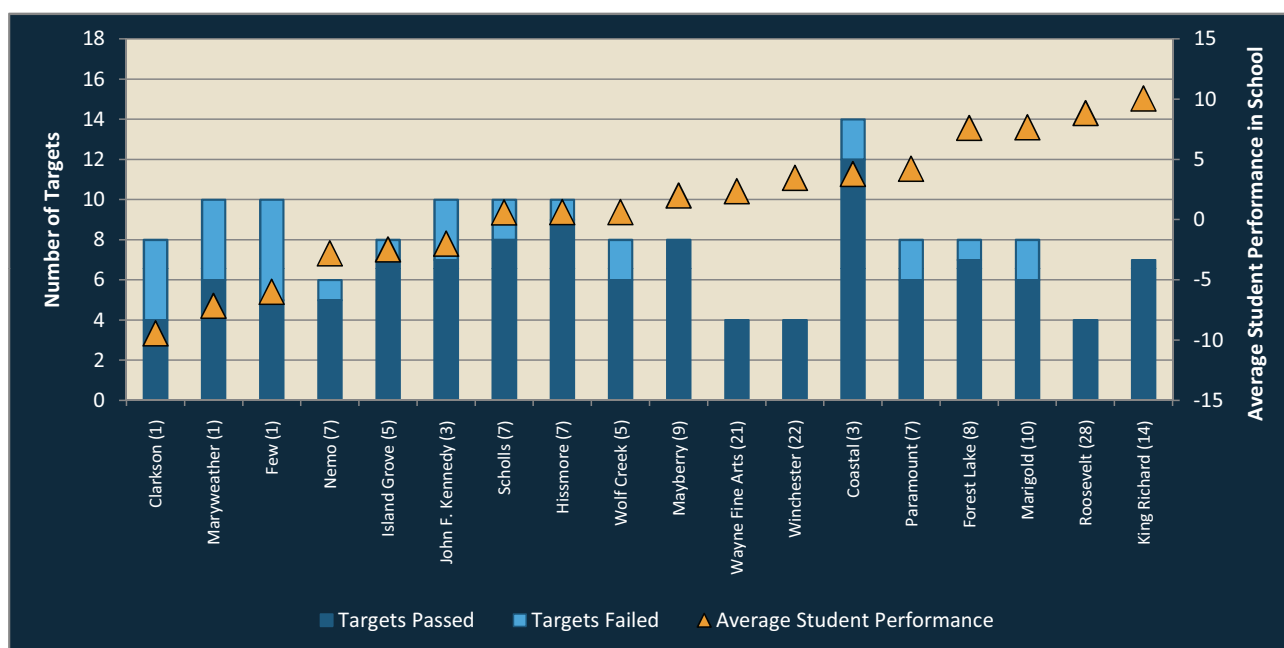


Figure 3. AYP performance of the elementary school sample under Delaware's 2008 AYP rules

Note: This figure indicates how each elementary school within the sample fared under Delaware's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet 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. Wolf Creek Elementary, for example, meets six of its eight targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles), which is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

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

Figure 5 indicates the degree to which elementary schools' math proficiency rates are aided by the confidence interval. On this figure, 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 confidence interval. The orange lines show the annual measurable objective needed to meet AYP. The figure shows that none of the sample elementary schools was assisted by the confidence intervals, because

the annual mathematics targets in Delaware are already low (i.e., 50%, see Table 1) relative to schools' overall performance. The effect of confidence intervals on middle school math proficiency rates and the reading proficiency rates for elementary and middle schools is much the same (not shown). In reading, none of the sample elementary or middle schools is assisted by the confidence intervals. **In short, applying the confidence interval (even a generous one like the 98% confidence interval used in Delaware) has little or no effect on whether schools meet their overall reading and math targets in Delaware, mostly because of the state's low annual targets.**⁹

Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low or mid-dling performance can still make AYP when the school

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 Figure 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.

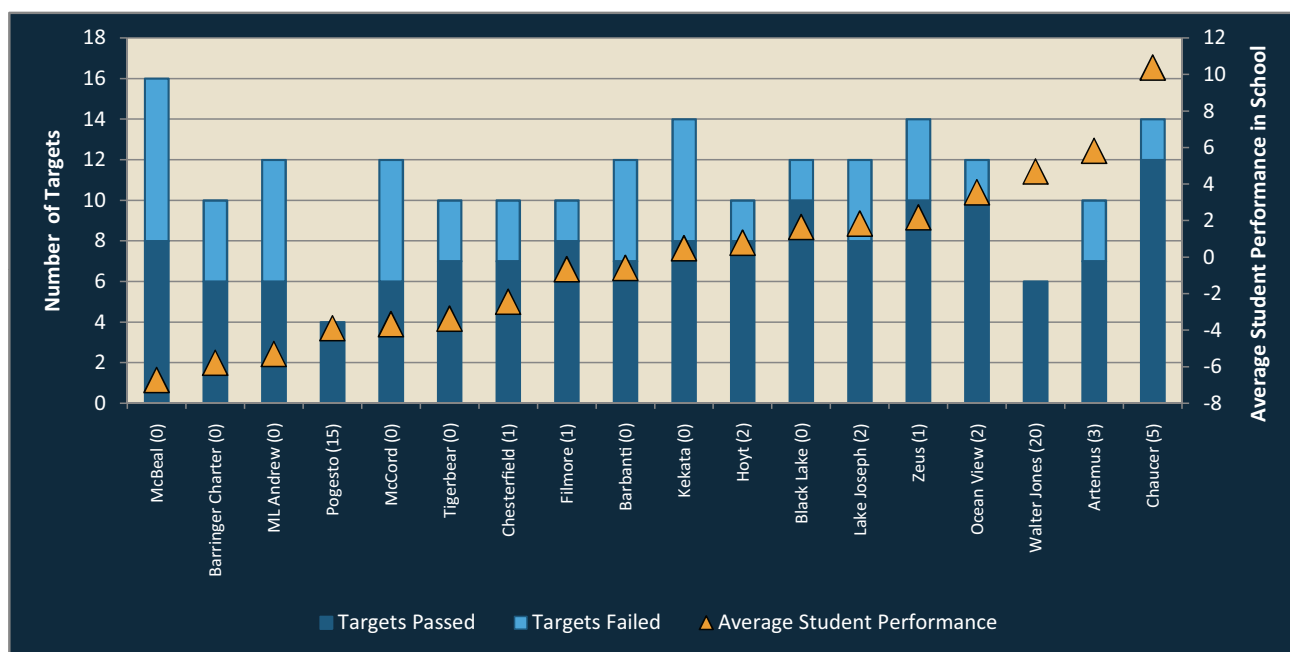


Figure 4. AYP performance of the middle school sample under Delaware's 2008 AYP rules

Note: This figure shows how each middle school within the sample would have fared under Delaware's AYP rules (as described in Table 1). The bars show the number of targets that each school had to meet to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not meet them (light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMO for even a single subgroup did not make AYP, so any light blue means that the school failed. Artemus Middle School, for example, met 7 of its 10 targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles), which is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

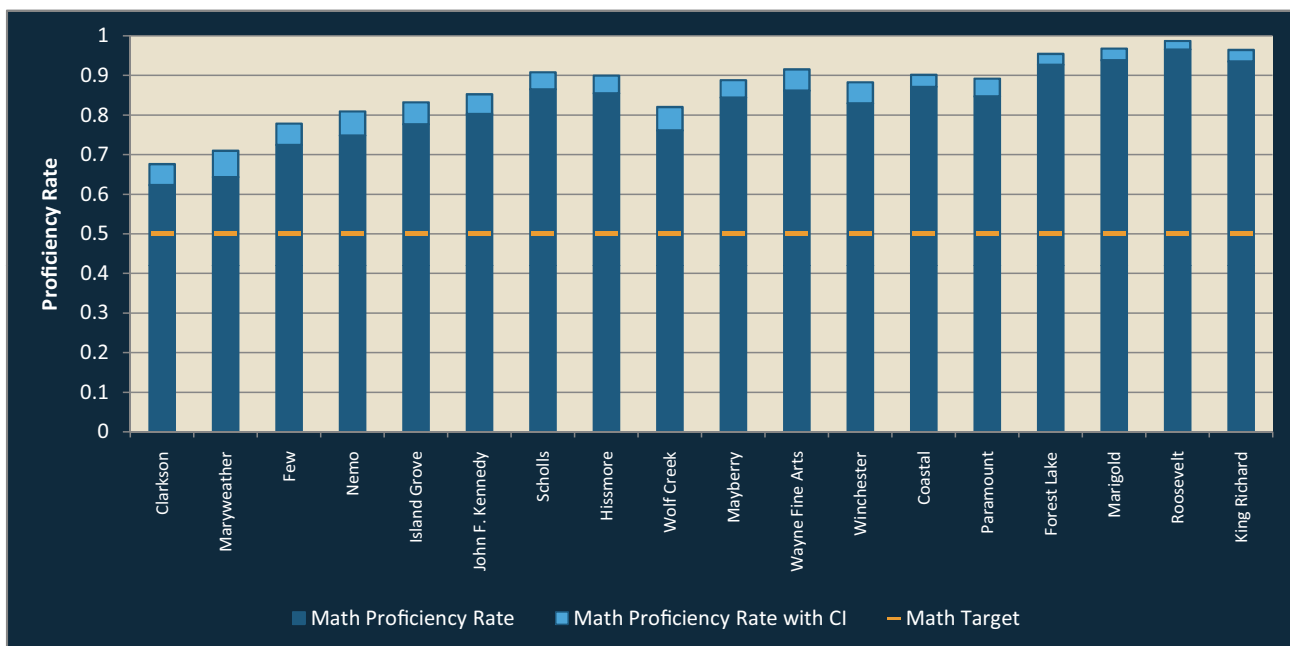


Figure 5. Impact of the confidence interval on elementary school math proficiency rates

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

Table 2. Elementary school subgroup performance of sample schools under the 2008 Delaware 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 | 62.4% | 47.3% | Y | N | | | Y | N | Y | N | | | | | Y | N | | | | | 8 | 4 | 50% | N | 1 |
| Maryweather | 64.4% | 53.4% | Y | N | | | Y | N | Y | N | | | | | Y | N | | | Y | Y | 10 | 6 | 60% | N | 1 |
| Few | 72.5% | 59.1% | Y | N | Y | N | Y | N | Y | N | | | | | Y | N | | | | | 10 | 5 | 50% | N | 1 |
| Nemo | 74.9% | 71.2% | Y | Y | | | | | Y | N | | | | | | | | | Y | Y | 6 | 5 | 83% | N | 7 |
| Island Grove | 77.7% | 70.4% | Y | Y | | | | | Y | Y | | | | | Y | N | | | Y | Y | 8 | 7 | 88% | N | 4 |
| JFK | 80.3% | 66.8% | Y | Y | Y | N | | | Y | N | Y | N | | | | | | | Y | Y | 10 | 7 | 70% | N | 3 |
| Scholls | 86.6% | 72.1% | Y | Y | Y | N | | | Y | Y | Y | N | | | | | | | Y | Y | 10 | 8 | 80% | N | 7 |
| Hissmore | 85.6% | 75.2% | Y | Y | Y | N | | | Y | Y | Y | Y | | | | | | | Y | Y | 10 | 9 | 90% | N | 7 |
| Wolf Creek | 76.1% | 72.1% | Y | Y | | | | | Y | N | | | | | Y | N | | | Y | Y | 8 | 6 | 75% | N | 5 |
| Alice Mayberry | 84.5% | 79.2% | Y | Y | | | | | Y | Y | Y | Y | | | | | | | Y | Y | 8 | 8 | 100% | Y | 9 |
| Wayne Fine Arts | 86.2% | 85.6% | Y | Y | | | | | | | | | | | | | | | Y | Y | 4 | 4 | 100% | Y | 21 |
| Winchester | 83.0% | 82.9% | Y | Y | | | | | | | | | | | | | | | Y | Y | 4 | 4 | 100% | Y | 22 |
| Coastal | 87.2% | 78.2% | Y | Y | Y | N | Y | N | Y | Y | Y | Y | | | Y | Y | | | Y | Y | 14 | 12 | 86% | N | 3 |
| Paramount | 84.8% | 78.4% | Y | Y | | | | | Y | N | | | | | Y | N | | | Y | Y | 8 | 6 | 75% | N | 7 |
| Forest Lake | 92.8% | 87.4% | Y | Y | Y | N | | | Y | Y | | | | | | | | | Y | Y | 8 | 7 | 88% | N | 8 |
| Marigold | 93.9% | 88.1% | Y | Y | Y | N | | | Y | N | | | | | | | | | Y | Y | 8 | 6 | 75% | N | 10 |
| Roosevelt | 96.6% | 93.9% | Y | Y | | | | | | | | | | | | | | | Y | Y | 4 | 4 | 100% | Y | 28 |
| King Richard | 93.6% | 91.2% | Y | Y | Y | Y | | | Y | | | | | | | | | | Y | Y | 7 | 7 | 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.

has fewer targets to meet because it has fewer subgroups. These figures do not, however, indicate which subgroups failed or passed in which school. Tables 2 and 3 list information on individual subgroup performance 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 sub-

groups that are separately evaluated for AYP include SWDs, students with LEP, low-income students, and the following race/ethnic categories: African American, Asian/Pacific Islander, Hispanic/Latino, American Indian/Alaska Native, and White. Tables 2 and 3 also show whether a school met AYP under the 2008 Delaware 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:

- Three elementary schools (Clarkson, Maryweather, and Few) failed to meet reading targets for their

Table 3. Middle school subgroup performance of sample schools under the 2008 Delaware 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 | 57.5% | 65.2% | Y | Y | N | N | N | N | N | N | Y | Y | | | N | N | Y | Y | Y | Y | 16 | 8 | 50% | N | 0 |
| Barringer Charter | 63.2% | 66.6% | Y | Y | N | N | | | Y | N | Y | N | | | Y | Y | | | | | 10 | 6 | 60% | N | 0 |
| ML Andrew | 55.8% | 71.9% | Y | Y | N | N | | | N | N | N | N | | | Y | Y | | | Y | Y | 12 | 6 | 50% | N | 0 |
| Pogesto | 53.7% | 77.8% | Y | Y | | | | | | | | | | | | | | | Y | Y | 4 | 4 | 100% | Y | 15 |
| McCord Charter | 58.6% | 73.3% | Y | Y | N | N | | | N | N | N | N | | | Y | Y | | | Y | Y | 12 | 6 | 50% | N | 0 |
| Tigerbear | 67.2% | 69.7% | Y | Y | N | N | | | Y | Y | Y | N | | | | | | | Y | Y | 10 | 7 | 70% | N | 0 |
| Chesterfield | 70.7% | 73.6% | Y | Y | N | N | | | Y | Y | Y | N | | | | | | | Y | Y | 10 | 7 | 70% | N | 1 |
| Filmore | 71.2% | 80.2% | Y | Y | N | N | | | Y | Y | | | | | Y | Y | | | Y | Y | 10 | 8 | 80% | N | 1 |
| Barbanti | 65.2% | 75.6% | Y | Y | N | N | N | N | Y | N | | | | | Y | Y | | | Y | Y | 12 | 7 | 58% | N | 0 |
| Kekata | 73.3% | 76.8% | Y | Y | N | N | N | N | Y | Y | Y | N | | | Y | N | | | Y | Y | 14 | 8 | 57% | N | 0 |
| Hoyt | 76.8% | 80.4% | Y | Y | N | N | | | Y | Y | Y | Y | | | | | | | Y | Y | 10 | 8 | 80% | N | 2 |
| Black Lake | 79.5% | 81.0% | Y | Y | N | N | | | Y | Y | Y | Y | | | Y | Y | | | Y | Y | 12 | 10 | 83% | N | 0 |
| Lake Joseph | 75.1% | 84.9% | Y | Y | N | N | N | N | Y | Y | | | | | Y | Y | | | Y | Y | 12 | 8 | 67% | N | 2 |
| Zeus | 79.0% | 81.7% | Y | Y | Y | N | N | N | Y | Y | Y | Y | | | Y | N | | | Y | Y | 14 | 10 | 71% | N | 1 |
| Ocean View | 81.5% | 89.1% | Y | Y | Y | Y | N | N | Y | Y | | | | | Y | Y | | | Y | Y | 12 | 10 | 83% | N | 2 |
| Walter Jones | 85.5% | 86.3% | Y | Y | | | | | Y | Y | | | | | | | | | Y | Y | 6 | 6 | 100% | Y | 20 |
| Artemus | 85.0% | 85.1% | Y | Y | Y | N | | | Y | N | | | | | Y | N | | | Y | Y | 10 | 7 | 70% | N | 3 |
| Chaucer | 87.4% | 92.6% | Y | Y | N | Y | Y | N | Y | Y | | | Y | Y | Y | Y | | | Y | Y | 14 | 12 | 86% | 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.

overall school population.

- All elementary schools met math targets for their overall population, as did all middle schools for both reading and math.
- Two of the 13 elementary schools (Hissmore and Forest Lake) and 3 of the 16 middle schools (Filmore, Hoyt, and Black Lake) that didn't make AYP only for their SWDs.
- One elementary school (Nemo) failed to make AYP only because of its low-income subgroup, and one elementary school (Island Grove) passed in every subgroup except for Hispanic students.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively.¹⁰ As shown, the performance of students with disabilities is proving most challenging for schools under Delaware's system,

¹⁰ Recall that elementary students do better on Delaware's math test than middle school students, perhaps because Delaware's cut scores are lower in math than in reading in grades 3 and 4 (see Figure 2).

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Delaware AYP rules

| SUBGROUP | Number of schools with qualifying subgroups | Number of schools where subgroup failed to meet math target | Number of schools where subgroup failed to meet reading target |
|---|---|---|--|
| Students with disabilities | 8 | 0 | 7 |
| Students with limited English proficiency | 4 | 0 | 4 |
| Low-income students | 15 | 0 | 8 |
| African-American students | 5 | 0 | 2 |
| Asian/Pacific Islander students | 0 | 0 | 0 |
| Hispanic students | 7 | 0 | 6 |
| American Indian/Alaska Native students | 0 | 0 | 0 |
| White students | 16 | 0 | 0 |

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Delaware 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 | 14 |
| Students with limited English proficiency | 7 | 6 | 7 |
| Low-income students | 17 | 3 | 6 |
| African-American students | 10 | 2 | 6 |
| Asian/Pacific Islander students | 1 | 0 | 0 |
| Hispanic students | 13 | 1 | 4 |
| American Indian/Alaska Native students | 1 | 0 | 0 |
| White students | 17 | 0 | 0 |

particularly in middle schools, where this subgroup tends to have enough students to meet the state’s minimum n of 40. In fact, all but one elementary school in the study with qualifying SWD subgroups failed to make AYP. Students with LEP are also struggling to meet the state’s targets; every school with a large enough LEP population to qualify as a separate subgroup failed to meet its reading targets for these students.

Characteristics of Schools that Did and Didn’t Make AYP

A close look at Figures 2 and 3 indicates that Delaware’s NCLB accountability system is, in most respects, behaving like those in other states. For example, among the elementary schools in our sample, Roosevelt, Winchester, and Wayne Fine Arts all made AYP in the greatest

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

| | Elementary Schools | | Middle Schools | |
|-----------------------------------|--------------------|--------------------|----------------|--------------------|
| | Made AYP | Failed to make AYP | Made AYP | Failed to make AYP |
| Number of schools in sample | 5 | 13 | 2 | 16 |
| Average student body size | 265 | 320 | 124 | 951 |
| Average % low income | 24 | 55 | 42 | 45 |
| Average % nonwhite | 30 | 45 | 27 | 46 |
| Average performance† | 5.35 | -0.36 | 0.40 | -0.11 |
| Average % growth‡ | 113 | 115 | 109 | 97 |
| Average number of targets to meet | 5 | 9 | 5 | 12 |

† Student performance is measured by NWEA's MAP assessment and is expressed as an index of grade level normative performance. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance.

‡ Average growth refers to improvement from fall to spring on the NWEA MAP assessments, averaged across all students within the school. Growth is expressed as an index value relative to NWEA norms and is scaled as a percentage. Thus, 100% means that students at the school are achieving normative levels of growth for their age and grade. Less than 100% growth means that the average student is increasing *by less* than normative amounts, while percentages over 100 mean that the average student is *exceeding* normative growth expectations.

number of states—28, 22, and 21, respectively. And these schools all made AYP in Delaware, too. Likewise, the elementary and middle schools that fail to make AYP in the greatest number of states also failed to make AYP in Delaware.

But Delaware is also home to a few anomalies. First, consider Mayberry Elementary (see Figure 3). It failed to make AYP in 19 of the 28 states in our sample, yet made AYP in Delaware. In examining Table 2, we can see that Mayberry didn't meet the minimum numbers for the students with LEP or SWD subgroups, which create difficulty for so many other schools in the study. With fewer accountable subgroups and relatively easy proficiency standards (Figure 2), Mayberry made AYP even when other schools with higher average performance didn't. Second, look at Pogesto Middle School (Figure 4). Even with its relatively low average performance, it made AYP in Delaware, but failed to do so in 13 of 28 states. Like Mayberry, its AYP success in Delaware is most likely attributable to its relatively small number of targets (four) along with Delaware's relatively easy proficiency standards compared to other states.

This is consistent with the patterns shown in Table 6, which compares schools making and not making AYP on a number of academic and demographic dimensions. Within the sample, elementary schools that made AYP did indeed show higher average student performance, but they also differed in the following ways: they had smaller student populations, fewer subgroups (and thus fewer targets to meet), and lower percentages of low-income and minority students. Similarly, middle schools that made AYP had slightly higher performing students, on average, than middle schools that failed, but they also had dramatically smaller total enrollments, smaller non-white populations, and fewer subgroups (and thus targets to meet).

Concluding Observations

The 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 Delaware's AYP rules (and AMOs) for 2008. We found that only 5 elementary schools and 2 middle schools—7 in all, from a sample of 36—would have made AYP in Delaware. Looking across the 28 state accountability sys-

tems examined in the study, this puts Delaware roughly in the middle of the sample distribution, as shown in Figure 1. In addition, Delaware uses a generous 98% confidence interval, but it appears to have little or no effect on whether schools meet their overall reading and math targets because the state already has such low annual targets compared to other states.

The overriding goal of the federal NCLB is to eliminate educational disparities within and across states, it's important to consider whether states' annual decisions about the progress of individual schools are consistent with this aim. In some respects, Delaware's NCLB accountability system is working exactly as Congress intended: identifying as "needing attention" schools with relatively high test score averages that mask low performance for particular groups of students such as low-income or Hispanic students. Almost all the sample schools made AYP in Delaware 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 im-

provement, even though sizable numbers of their pupils weren't meeting state standards. Disaggregating data by race, income, and so on has made those students visible. That is surely a positive step.

Yet NCLB's design flaws are also readily apparent. Does it make sense that the size of a school's enrollment has so much influence over making AYP? Does it make sense that having fewer subgroups enhances the likelihood of making AYP? Even if actual participation guidelines for English language learners and SWDs are more generous under the current state assessment system,¹¹ doesn't the failure of these students to meet Delaware's targets (especially at the middle school level) 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 LEP students and SWDs, 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.

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

¹¹ See footnote 5.

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.