

# EXECUTIVE SUMMARY

The intent of the No Child Left Behind (NCLB) Act of 2001 is to hold schools accountable for ensuring that all students in grades three through eight achieve proficiency in reading and math by 2014, 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 ambitious goals.

This study examines the NCLB accountability systems and the basic AYP rules for 28 states as they operate in practice. We did this by selecting 36 real schools from around the nation (half elementary, half middle)—schools that vary by size, achievement, diversity, and so on—and determining which of them would or would not make AYP when evaluated under each state’s accountability rules.<sup>1</sup> In other words, if a particular school that made AYP in Washington were relocated to North Dakota, or Ohio, or Texas, would that same school also make AYP there? And if not, what factors within NCLB, and its implementation by the various states, explain this? Based on this analysis, what can we learn about how AYP determinations vary across the country—and, at least by inference, about the effectiveness of NCLB in ensuring that *all* students attain proficiency?

NCLB imposes strict expectations for schools—100% of their students must achieve proficiency by 2014—but gives states wide latitude in terms of key variables. Under the act, states have leeway to:

1. Craft their own academic standards, select their own tests, and define proficiency in reading and math as they like; as a result, proficiency standards (which take the form of cut scores<sup>2</sup> on state tests) vary widely in their rigor and consistency.

2. Establish their own annual targets (also called annual measurable objectives or AMOs) for moving students to the proficient level by 2014. Some states require schools to follow a linear trajectory to the 100% proficiency goal, seeking similar gains each year; others use a back-loaded trajectory (meaning that little improvement is required during the early years and much is required during latter years) to achieve this result.
3. Apply confidence intervals, or margins of statistical error, to schools’ proficiency rates. When states use such intervals, it means that the percentage of students required to reach proficiency can actually be lower than the stated target. States also determine the confidence interval’s size and how it is used.
4. Determine when the size of a student subgroup within a school is large enough that it must meet AYP targets. In other words, states decide whether particular subgroups of minority, low-income, or limited English proficient (LEP) students, for instance, are large enough that their test results must be counted separately for determining their school’s AYP status, in addition to being counted within the general school population.

How do these multiple allowances for state discretion and variation affect AYP determinations from state to state? To find out, we evaluated the performance of students in 18 elementary schools and 18 middle schools relative to each state’s proficiency cut scores and 2008 annual targets. We also applied confidence intervals to results, according to each state’s rules, and evaluated the performance of all subgroups within a school that met or exceeded each state’s minimum pupil-count requirement. This allowed us to estimate whether a school would meet most of the requirements needed to make AYP.

<sup>1</sup> We did not examine the impact of NCLB’s “safe harbor” provision or other indicators such as attendance and test-participation rates. Nor were we able to consider the impact of the U.S. Department of Education’s recent growth model pilot program, which allows states to track individual student achievement over time. We used school data and proficiency cut score estimates from academic year 2005–2006 and applied them against state AYP rules for academic year 2007–2008 (shortened to “2008” in this report).

<sup>2</sup> A cut score is the minimum score a student must receive on the applicable state test in order to be considered proficient under that state’s accountability system.

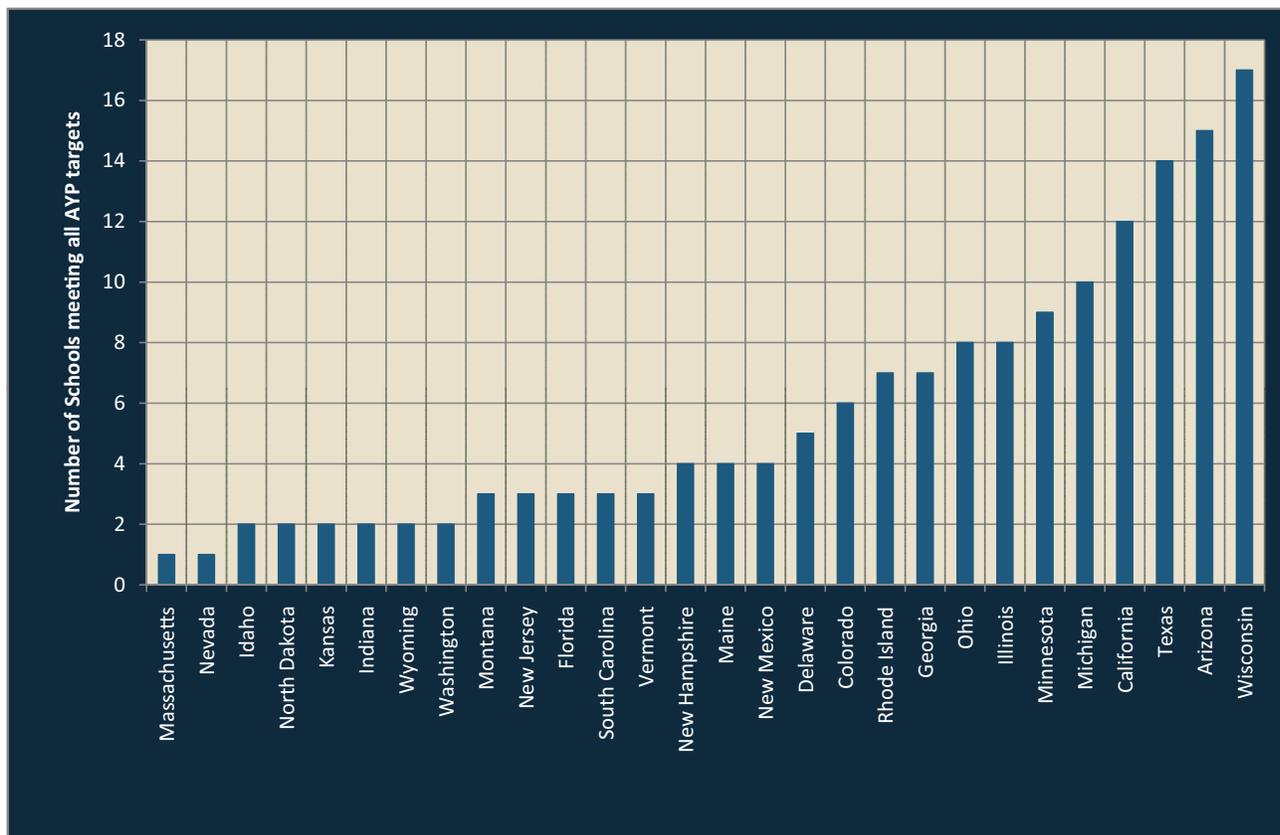


Figure ES-1. Number of sampled elementary schools that made AYP in 2008, by state

Here are the study’s key findings:

- **Within the elementary school sample, the number of schools that made AYP varied greatly by state.** Almost all our sampled schools **failed to make AYP** in some states, and nearly all of these **same schools made AYP** in others. In Massachusetts, for example, a state with high proficiency cut scores and relatively challenging annual targets and AYP rules, **only 1 of 18 elementary schools made AYP**; in Wisconsin 17 schools made AYP (Figure ES-1). Same kids, same academic performance, same schools—different states, different cut scores, different rules. And very different results.
- **There is more consistency across states with the middle school sample because so few of these schools made AYP in any states.** In 21 of the 26 states stud-

ied,<sup>3</sup> two or fewer middle schools made AYP. In no state did even half of the 18 middle schools meet the 2008 AYP requirements. This is mostly because the larger size of middle schools generally means that they have plenty of students with disabilities (SWDs) and minority, low-income,<sup>4</sup> and LEP pupils who are counted separately for accountability purposes. Although subgroups of minority students within our sample schools performed well enough to meet their annual targets in many states, almost all schools with a qualifying LEP or SWD subgroup failed to meet the targets for these groups in nearly every state.

- **When it comes to whether the performance of a subgroup will hurt a school’s chances of making AYP, the state’s decision relative to minimum subgroup size (called “*n* size”) is critical.** Consider Chaucer Middle School, for example, the highest performing middle

<sup>3</sup> Two states (Texas and New Jersey) are not included in the middle school analysis because 8th grade cut scores were not available.

<sup>4</sup> Low-income students are those who receive a free or reduced-price lunch.

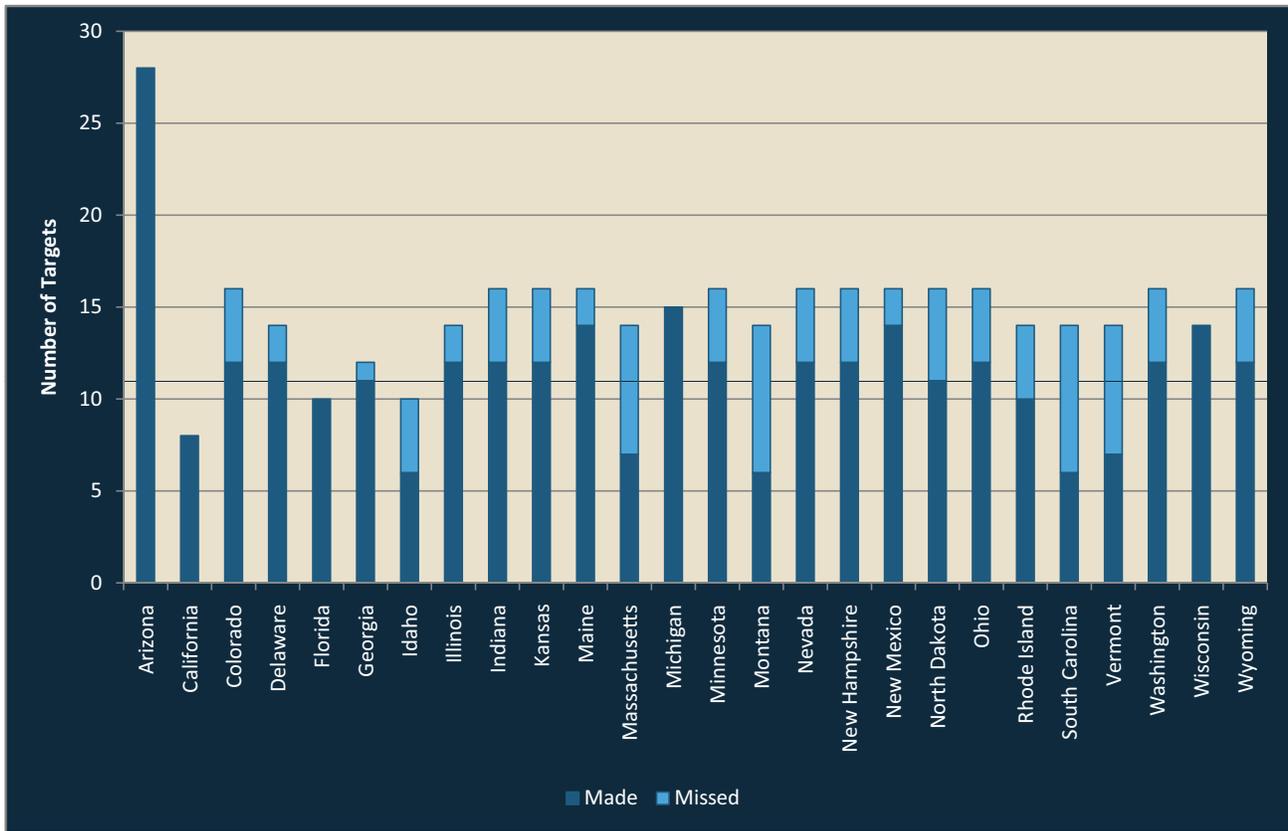


Figure ES-2. Number of subgroup targets met by Chaucer Middle School in 2008, by state

Note: Arizona has more targets because each grade level is considered a group unto itself. For instance, a middle school in Arizona with three grades and four subgroups has  $3 \times 4 \times 2$  (subjects) or 24 targets.

school in our sample (see Figure ES-2). Though it achieved strong performance overall and added greater value to its students’ performance over time than most other schools in the country (and virtually all schools in the sample), it failed to make AYP in 21 of the 26 states because of the performance of its subgroups (if even one target is missed, as indicated by the light blue bars, the school does not make AYP in that state). In the states with relatively small *n* sizes, where Chaucer is held accountable for numerous subgroups (e.g., Nevada, New Hampshire, and North Dakota), it did not make AYP.<sup>5</sup> On the other hand, in states with large *n* sizes, where Chaucer is held accountable for fewer subgroups (e.g., Florida and California), it made AYP. Generally, the lower the state’s *n* size, the more subgroups for which the typical school is accountable, and the more separate targets that school must hit.

## Implications

For an accountability system to be effective, educators must believe that it is fair, consistent, and understandable. Unfortunately, the way NCLB rates schools appears to be idiosyncratic—even random—and opaque. Schools that make AYP in one state fail to make AYP in another. Those that are considered failures in one part of the country are deemed to be doing fine in another. Although schools are being told that they need to improve student achievement in order to make AYP under the law, the truth is that many would fare better if they were just allowed to move across state lines.

One of the adages of the NCLB era is that a child’s zip code shouldn’t determine her life chances. Indeed. But neither should a school’s zip code determine whether or

<sup>5</sup> Arizona is an exception, but the number of subgroups in Arizona is large primarily because they treat each grade level as a subgroup. Grade levels are not subgroups in the same sense as low-income students, or LEP students would be considered a subgroup because they have no defining achievement related characteristic that distinguishes them from others.

not it makes AYP. Yet regrettably it often does. And so the success or failure of a given school under NCLB is driven as much by the way the law is implemented by its home state as it is by the performance of its students and

the amount of progress they've made over the course of a year.

This is the Accountability Illusion.