California

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 California'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 California'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 California'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 California's AYP rules for academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- We estimate that **6 of 18 elementary schools** and **14 of 18 middle schools** in our sample failed to make AYP in 2008 under California's accountability system. (This 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 only three states exceeded California in terms of the number of elementary schools making AYP (Texas, Arizona, and Wisconsin).**

[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 California Standards Test.

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

- In California, subgroups of students (such as minorities or low-income children[2]) must be quite large in order to be counted separately in AYP calculations. In this way, **the achievement scores of many minority, disabled, or limited English proficient students that do not count separately in California would count separately in most of the other states**.

- Furthermore, although the majority of states examined in the study apply confidence intervals (or margins of statistical error) to their student proficiency rates, California's 99% confidence interval gives schools greater leniency than the 95% confidence interval used by most other states. Such a lenient confidence interval might normally rescue otherwise failing schools, but because California's minimum subgroup size is rather large anyway and because the state places limitations on the use of intervals, it is seldom used.

More schools in the study make AYP in **California** than in most other states. There are several factors which contribute to this. First, though California has relatively high proficiency standards (or cut scores) in reading and math, the percentage of students required to meet those standards in 2008 is relatively low (roughly 35 percent proficient in English Language Arts and 37 percent proficient in math). An additional factor is that the minimum subgroup size for reporting purposes is relatively high. California's minimum subgroup size is generally 100 students (they also use a "sliding" *n* size depending on a school's enrollment). This is larger than the minimum subgroup size used by most other states examined in the study. Hence, the achievement scores of many minority, limited English proficient (LEP), and disabled students that are *not* counted separately for accountability purposes in California would be counted separately in most other states.
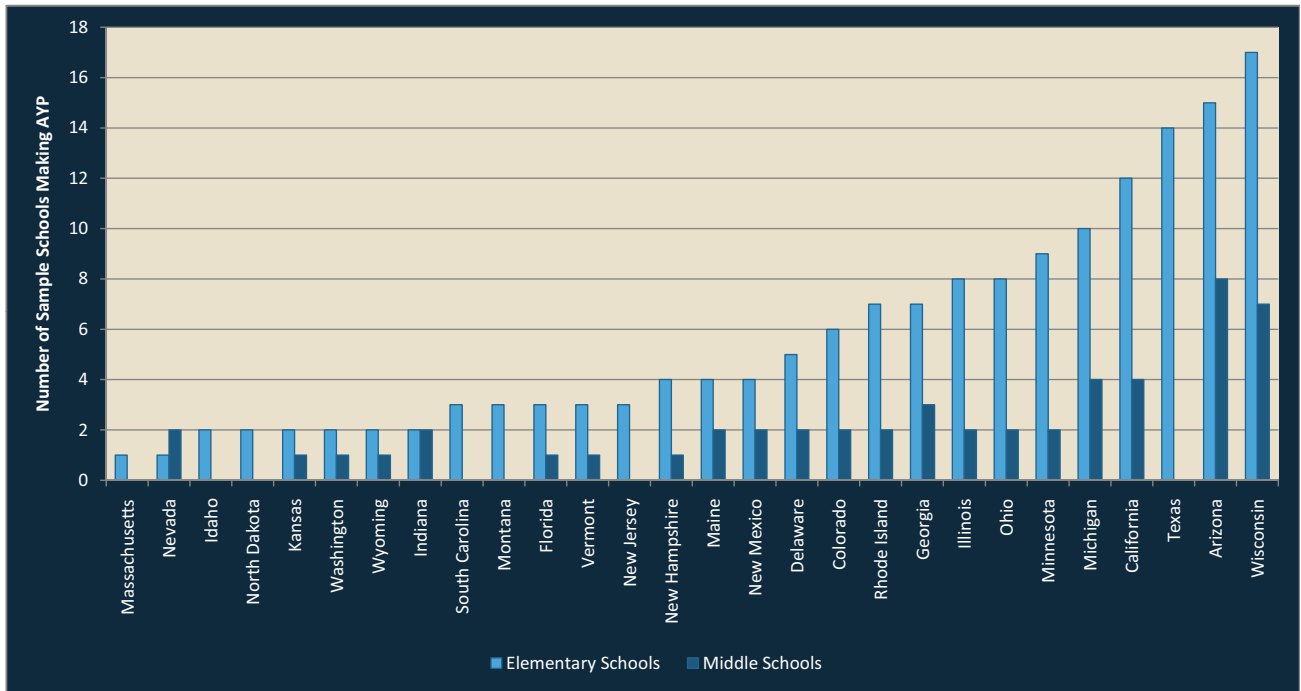
Figure 1. Number of sample schools making AYP by state

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.

- **California's accountability system, then, has high cut scores and high minimum subgroup sizes, but fairly low annual targets (hovering around proficiency levels of 35%).**

- Still, many of the schools in our sample that failed to make AYP in California did meet expected targets for their overall populations but failed because of the performance of individual subgroups.[3]

- In California, as in most states, schools with fewer subgroups attain AYP more easily 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 California 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 California's system is whether it has enough English language learners to qualify as a separate subgroup. Almost every single school with a subgroup of students with limited English proficiency (LEP)[4] failed to make AYP. Likewise, most of the schools with enough qualifying students with disabilities (SWDs) failed to meet their AYP targets.[5]

---

[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.

[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 California Standards Test, 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.

Introduction

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

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

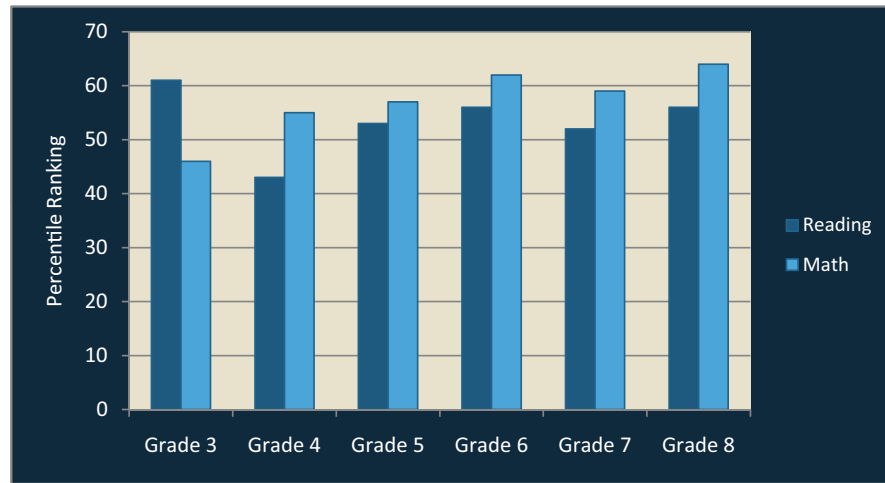[6] We gave all schools in our sample pseudonyms in this report.

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

Note: This figure illustrates the difficulty of California'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 California's cut scores are at or below the 65th percentile.

Proficiency cut score estimates for the California Standards Tests (CST) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that California's definitions of proficiency in reading and math were relatively difficult compared with the standards set by the other 25 states in that study. These cut scores were used to estimate whether students would have scored as proficient or better on the California 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 California 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 California's 2008 AYP rules.

Table 1 shows the pertinent California AYP rules that we applied to elementary and middle schools in this study. California's minimum subgroup size is 15% of the student population; however, the minimum subgroup size can't be less than 50 or more than 100.[7] This is larger than the minimum subgroup size used by most other states examined in the study.

Furthermore, although the majority of states examined in the study apply confidence intervals (or margins of statistical error) to their student proficiency rates, California's 99% confidence interval gives schools greater leniency than the 95% confidence interval used by most other states. So, for instance, although schools are supposed to get 35.2% of their grade 3-8 students (and 35.2% of their grade 3-8 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. Such a lenient confidence interval might normally rescue otherwise failing schools, but two factors prevent the interval from being used that often: 1) California's minimum n size is rather large anyway, so fewer subgroups are held separately accountable in the first place; and 2) it is only used if the school population is fewer than 100 students.[8] **California's accountability system, then, has high cut scores and high minimum n sizes, but lenient confidence intervals and fairly low annual targets (hovering around proficiency levels of 35%).**

Note that we were unable to examine the impact of NCLB's "safe harbor" provision. This provision per-

---

[7] In California, the minimum subgroup size is 15% of the total school population. Generally, this means that the subgroup size grows with the school size. However, there's also a clause that specifies that the minimum subgroup size can't be less than 50 or more than 100. For example, a school with a total population of 500 would have a minimum subgroup size of 75 (i.e, 15%), but a school with only 300 students would have a minimum subgroup size of 50 since 15% of 300 (i.e., 45) is below the required minimum. Similarly, a school with 800 students would have a minimum subgroup size of 100, since 15% of 800 (i.e., 120) is greater than the maximum size of 100.

[8] We conducted an analysis to show the effect of confidence intervals on the reading and math proficiency rates for elementary and middle schools. We describe those results later in the report.

Table 1. California AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 15% of the student population but with a minimum of 50 and maximum of 100	
	SWDs: 15% of the student population but with a minimum of 50 and maximum of 100	
	Low-income students: 15% of the student population but with a minimum of 50 and maximum of 100	
	LEP students: 15% of the student population but with a minimum of 50 and maximum of 100	
CI	**Applied to proficiency rate calculations?**	**Additional notes**
	Yes; 99% CI	Used only if school population is fewer than 100 students; not used otherwise
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	13.6	35.2
Grade 4	13.6	35.2
Grade 5	13.6	35.2
Grade 6	13.6	35.2
Grade 7	13.6	35.2
Grade 8	13.6	35.2
MATH		
Grade 3	16.0	37.0
Grade 4	16.0	37.0
Grade 5	16.0	37.0
Grade 6	16.0	37.0
Grade 7	16.0	37.0
Grade 8	16.0	37.0

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

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

Furthermore, attendance and test participation rates are beyond the scope of the study. 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 academic year. For each school, we calculated reading and math proficiency rates (along with any confidence intervals) to determine whether the overall school population and any qualifying subgroups achieved the AMOs. We deemed that a school made AYP if its overall student body and all its qualifying subgroups met or exceeded its AMOs. Again, Appendix 1 supplies further methodological detail.
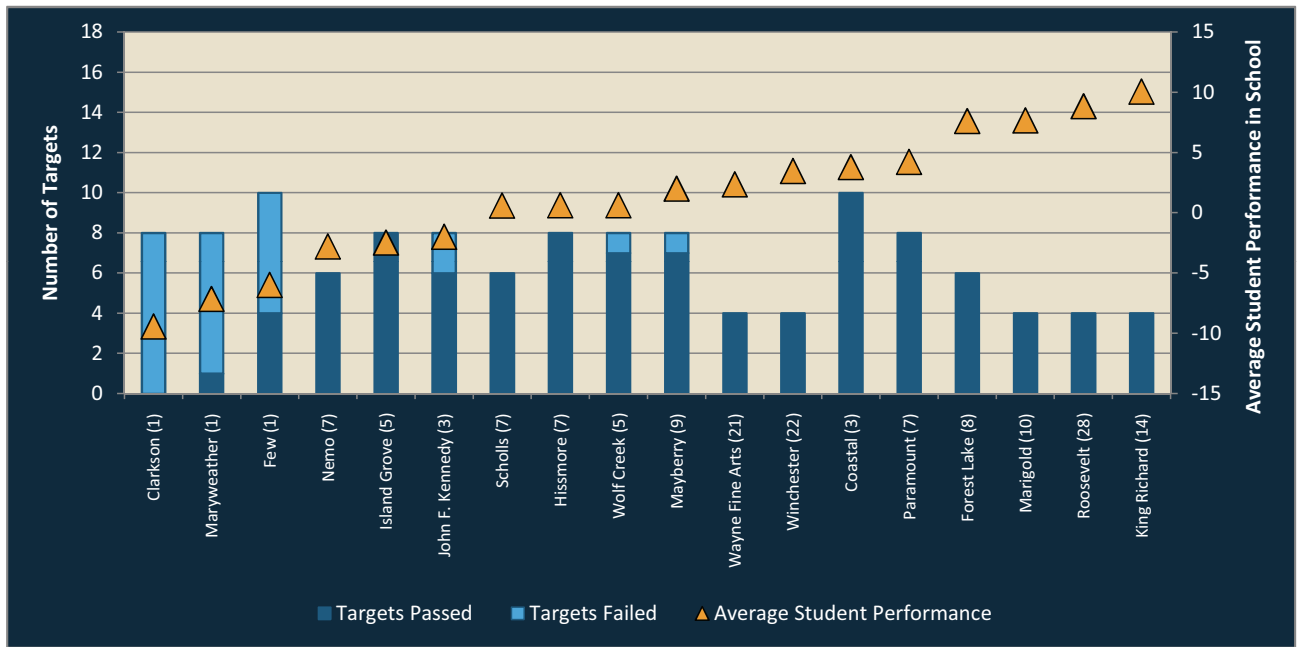
Figure 3. AYP performance of the elementary school sample under California's 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under California'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. Mayberry Elementary, 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, 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.

How Did the Sample Schools Fare under California's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under California's 2008 AYP rules. **Twelve elementary schools made AYP and six failed to make it**. 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. The majority of the schools making AYP are in the right half of the figure, meaning that the highest performing students were found at these schools.

Yet almost without regard to average student performance, the schools that made AYP were those with relatively few qualifying subgroups—and thus the fewest targets to meet (since each subgroup has its own 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 4 illustrates the AYP performance of the sample middle schools under the 2008 California AYP rules. **Of 18 middle schools in our sample, only 4 made AYP**—one low-performance school (Pogesto), and three high-performance schools (Walter Jones, Artemus, and Chaucer), most of which have relatively few qualifying subgroups.

Figures 5 and 6 indicate the degree to which schools' math proficiency rates are aided by California's confidence interval for elementary and middle schools, respectively. 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 are increased by the application of the confidence interval. The orange lines show the annual measurable objective
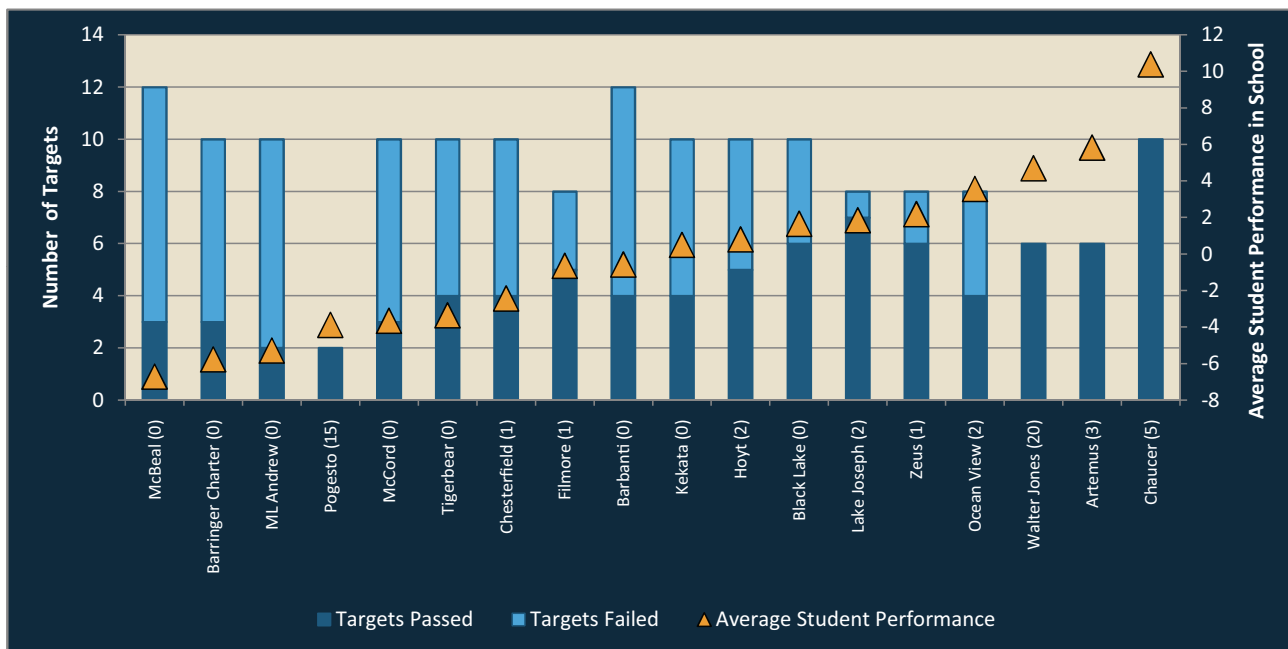
Figure 4. AYP performance of the middle school sample under California's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under California'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. Lake Joseph Middle School, 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, and its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

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

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. Since California only makes use of the confidence interval in schools with fewer than 100 students, confidence intervals are not shown in Figure 5 (all schools have more than 100 students). If confidence intervals were used, however, they would be depicted in a lighter shade of blue on top of the dark blue bar. Annual targets are indicated by the orange lines.
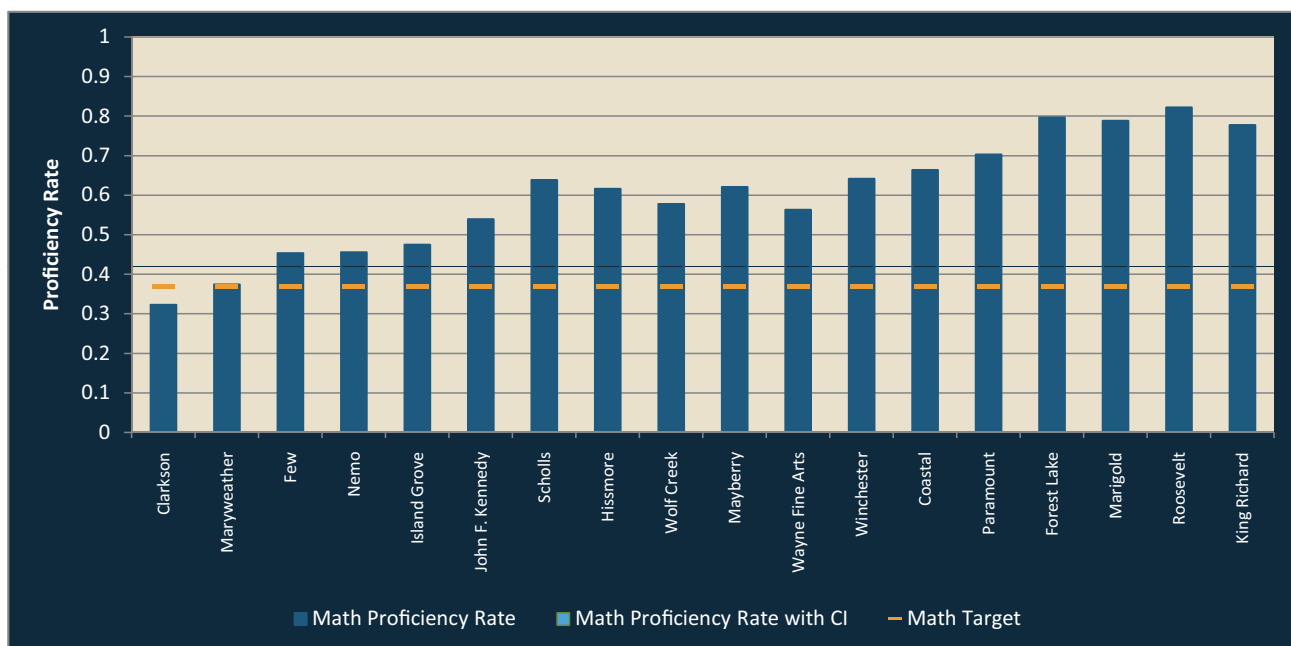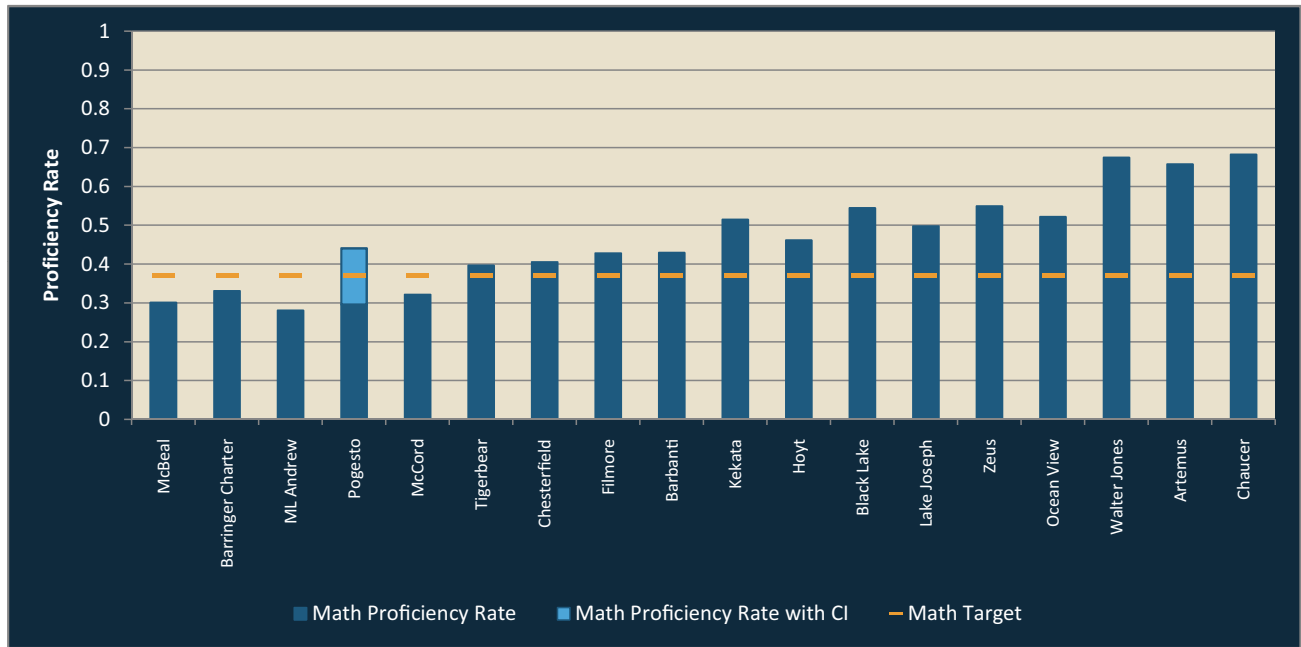
Figure 6. Impact of the confidence interval on middle school math proficiency rates

NNote: 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. Since California only makes use of the confidence interval in schools with fewer than 100 students, confidence intervals are not shown for the most part. Pogesto, in fact, is the only eligible school and the only one that meets its overall math target via 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.

needed to meet AYP. These figures show that no sample elementary schools and one middle school (Pogesto) were assisted by the confidence interval. It's important to keep in mind, however, that Pogesto was the only school *eligible* to make use of the confidence interval (California rules allow confidence intervals to be used only in schools with fewer than 100 students.)

The effect of confidence intervals on reading proficiency rates for elementary and middle schools is much the same (not shown). In reading, no elementary schools and only one middle school (Pogesto again) met the overall targets with the confidence interval. In short, **the application of the confidence interval had little or no impact on whether the sample elementary and middle schools met California's overall reading and math targets.**[9] **So, even though we would expect California's generous confidence interval to rescue otherwise failing schools, we see that the state's high minimum n size and low school enrollment requirement prevent the interval from serving that function.**

Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low or middling performance can still make AYP when the school has fewer targets to meet because it has fewer subgroups. These figures do not, however, indicate which subgroups failed or passed in which school. Information on individual subgroup performance appears in Tables 2 and 3 for elementary and middle schools, respectively.

Tables 2 and 3 show which subgroups qualified for evaluation at each school (i.e., whether the number of students within that subgroup exceeded the state's minimum n), and whether that subgroup passed or failed. Although all schools are evaluated on the proficiency rate of their overall population, potential subgroups that are separately evaluated for AYP include SWDs, students with LEP, low-income students, and the following race/ethnic categories: African American, Asian/Pacific Islander, Hispanic/Latino, American In-

---

[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 may be 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 subgroup performance of sample schools under the 2008 California 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	32.3%	18.6%	N	N			N	N	N	N					N	N					8	0	0%	N	1
Maryweather	37.4%	32.9%	Y	N			N	N	N	N					N	N					8	1	13%	N	1
Few	45.4%	32.6%	Y	N	N	N	Y	N	Y	N					Y	N					10	4	40%	N	1
Nemo	45.6%	44.7%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	7
Island Grove	47.5%	50.6%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	4
JFK	53.9%	42.2%	Y	Y					Y	N	Y	N							Y	Y	8	6	75%	N	3
Scholls	63.8%	48.0%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	7
Hissmore	61.6%	50.0%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	7
Wolf Creek	57.8%	54.3%	Y	Y					Y	Y					Y	N			Y	Y	8	7	88%	N	5
Alice Mayberry	62.1%	50.9%	Y	Y					Y	Y	Y	N							Y	Y	8	7	88%	N	9
Wayne Fine Arts	56.3%	61.5%	Y	Y															Y	Y	4	4	100%	Y	21
Winchester	64.2%	63.0%	Y	Y															Y	Y	4	4	100%	Y	22
Coastal	66.3%	63.0%	Y	Y	Y	Y			Y	Y	Y	Y							Y	Y	10	10	100%	Y	3
Paramount	70.3%	62.7%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	7
Forest Lake	79.6%	70.4%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	8
Marigold	78.8%	76.5%	Y	Y															Y	Y	4	4	100%	Y	10
Roosevelt	82.2%	79.0%	Y	Y															Y	Y	4	4	100%	Y	28
King Richard	77.7%	82.3%	Y	Y															Y	Y	4	4	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.

dian/Alaska Native, and White. Tables 2 and 3 also show whether a school met AYP under the 2008 California 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

- One elementary school (Clarkson) and four middle schools (McBeal, Barringer Charter, ML Andrew, and McCord Charter) failed to meet math targets for their overall school populations.

- One elementary school (Few) and nine middle schools failed the AMOs for their SWDs.

- All elementary schools (Clarkson, Maryweather, and Few) and middle schools (McBeal, Barbanti) with qualified LEP subgroups failed to make AYP.

- Four elementary schools and nine middle schools failed to meet the AMOs for low-income students.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. As shown, **California's minimum *n* of 100 means that the schools in the sample have essentially five subgroups—SWDs, low-income, Hispanic/Latino, African American, and**

Table 3. Middle school subgroup performance of sample schools under the 2008 California 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	30.1%	36.2%	N	Y	N	N	N	N	N	N					N	N			Y	Y	12	3	25%	N	0
Barringer Charter	33.1%	35.3%	N	Y	N	N			N	N	N	N			Y	Y					10	3	30%	N	0
ML Andrew	28.0%	38.1%	N	Y					N	N	N	N			N	N			N	Y	10	2	20%	N	0
Pogesto	29.6%	33.3%	Y	Y																	2	2	100%	Y	15
McCord Charter	32.2%	43.6%	N	Y					N	N	N	N			N	N			Y	Y	10	3	30%	N	0
Tigerbear	39.6%	39.0%	Y	Y	N	N			N	N	N	N							Y	Y	10	4	40%	N	0
Chesterfield	40.5%	38.2%	Y	Y	N	N			N	N	N	N							Y	Y	10	4	40%	N	1
Filmore	42.8%	47.2%	Y	Y					N	Y					N	N			Y	Y	8	5	63%	N	1
Barbanti	42.9%	45.3%	Y	Y	N	N	N	N	N	N					N	N			Y	Y	12	4	33%	N	0
Kekata	51.5%	47.3%	Y	Y	N	N			N	N	N	N							Y	Y	10	4	40%	N	0
Hoyt	46.2%	49.8%	Y	Y	N	N			N	Y	N	N							Y	Y	10	5	50%	N	2
Black Lake	54.4%	48.7%	Y	Y	N	N			Y	Y	N	N							Y	Y	10	6	60%	N	0
Lake Joseph	49.8%	53.7%	Y	Y					Y	Y					N	Y			Y	Y	8	7	88%	N	2
Zeus	54.9%	53.2%	Y	Y	N	N			Y	Y									Y	Y	8	6	75%	N	1
Ocean View	52.2%	63.6%	Y	Y					N	N					N	N			Y	Y	8	4	50%	N	2
Walter Jones	67.4%	66.9%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	20
Artemus	65.7%	66.2%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	3
Chaucer	68.2%	73.9%	Y	Y					Y	Y			Y	Y	Y	Y			Y	Y	10	10	100%	Y	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.

White—with sufficient numbers of students for reporting purposes. Of these subgroups, the performance of low-income students (and to a lesser extent, SWDs) is proving most challenging for schools under California's system. This is especially true in middle schools, which are generally larger and more likely to have enough students to meet the minimum *n* in the subgroups.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that California's NCLB accountability system is, in many 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 number of states—28, 22, and 21, respectively. And these schools all made AYP in California, too. Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make AYP in California.

But California is also home to a few anomalies. First, consider Coastal Elementary (see Figure 3). It failed to

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 California 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	2	1	1
Students with limited English proficiency	3	2	3
Low-income students	13	2	4
African-American students	4	0	2
Asian/Pacific Islander students	0	0	0
Hispanic students	6	2	4
American Indian/Alaska Native students	0	0	0
White students	15	0	0

Table 5. Summary of subgroup performance of sample middle schools under the 2008 California 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	9	9
Students with limited English proficiency	2	2	2
Low-income students	17	11	9
African-American students	8	8	8
Asian/Pacific Islander students	1	0	0
Hispanic students	9	7	6
American Indian/Alaska Native students	0	0	0
White students	16	1	0

make AYP in 25 of the 28 states in our sample, yet made AYP in California. In examining Table 2, we can see that Coastal didn't meet the minimum numbers for the LEP or Hispanic subgroups, which created difficulty for many schools in the study. Without those particular subgroups counting, Coastal was able to meet AYP, even when it failed under the standards of most other states.

This is consistent with the patterns shown in Table 6, which compares the schools that did and didn't make AYP on a number of academic and demographic dimensions. Within the sample, schools that make AYP do indeed show higher average student performance, but they also differ in the following ways: they have much smaller student populations, fewer subgroups (and thus fewer targets to meet), and much lower percentages of low-in-

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

	Elementary Schools		Middle Schools	
	Made AYP	**Failed to make AYP**	**Made AYP**	**Failed to make AYP**
Number of schools in sample	12	6	4	14
Average student body size	262	390	520	956
Average % low income	35	70	28	50
Average % nonwhite	30	62	27	49
Average performance†	3.67	-3.66	4.25	−1.29
Average % growth‡	121	103	121	92
Average number of targets to meet	6	8	6	9

† 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.

come and nonwhite students. Similarly, middle schools that made AYP have slightly higher performing students, on average, than middle schools that didn't make it, but have smaller total enrollments, smaller nonwhite populations, and fewer subgroups (and thus targets to meet).

Concluding Observations

This study examined the test performance data of students in 18 elementary and 18 middle schools across the country to see how those schools would fare under California's AYP rules (and AMOs) for 2008. We found that 12 elementary schools and 4 middle schools—16 in all, from a sample of 36—would have made AYP in California. Looking across the 28 state accountability systems examined in the study, this places California at the high end of the distribution in terms of the number of schools making AYP (see Figure 1).

Because the overriding goal of NCLB is to eliminate educational disparities within and across states, it's important to consider whether states' annual decisions about

the progress of individual schools are consistent with this aim. In some respects, California'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 of the sample schools made AYP in California 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 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? In the case of California, does it make sense that high cut scores can be

"tamed" by low annual targets,[10] or that large minimum *n* sizes mean that the achievement scores of students with disabilities or limited English proficiency are not counted separately? 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.

---

[10] There is some evidence that California is now rapidly increasing its annual targets. So even though the current accountability system has its drawbacks, California appears to be trying to remedy and align its various components..