



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

Here are some key findings:

- We estimate that **16 of 18 elementary schools and 17 of 18 middle schools in our sample would fail to make adequate yearly progress** in 2008 under Kansas’s accountability system. This high failure rate is partly explained by our sample, which intentionally includes some schools with relatively large populations of low-performing students. But it’s also partially explained by Kansas’s demanding annual targets for students (roughly 75% of students were expected to meet proficiency targets in 2008).

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 Kansas Assessment Program.

² It’s important to note that students in subgroups not meeting the minimum *n* sizes are still included for accountability purposes in the overall student calculations; they simply are not treated as their own subgroup.

- Looking across the 28 state accountability systems examined in the study, **only two states passed fewer of the sample elementary schools than Kansas (Kansas ties 5 other states with only 2 elementary schools making AYP). In addition, Kansas is one of 6 states with a single passing middle school in the sample (see Figure 1).**
- Many of the schools in our sample that failed to make AYP in Kansas are meeting expected targets for their overall populations but failed because of the performance of individual subgroups, particularly students with disabilities (SWDs) and students with limited English proficiency (LEP).²

Under **Kansas’s** accountability system, 16 of 18 elementary schools and 17 of 18 middle schools in our sample fail to make AYP in 2008. This places Kansas near the low end of the state distribution in terms of the number of schools making AYP. Kansas’s definitions of proficiency generally ranked about average compared with the standards set by the other states. However, Kansas’s annual targets in reading (the percentage of students in various subgroups that have to meet proficiency) are relatively difficult to achieve. Specifically, 75.6 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. In addition, Kansas’s minimum subgroup size (30) is slightly lower than in many of the other states we examined. This means that more groups of students are held separately accountable than would be in many other states. In fact, every single school with a limited English proficient (LEP) or students with disabilities (SWD) subgroup failed to make AYP in Kansas.

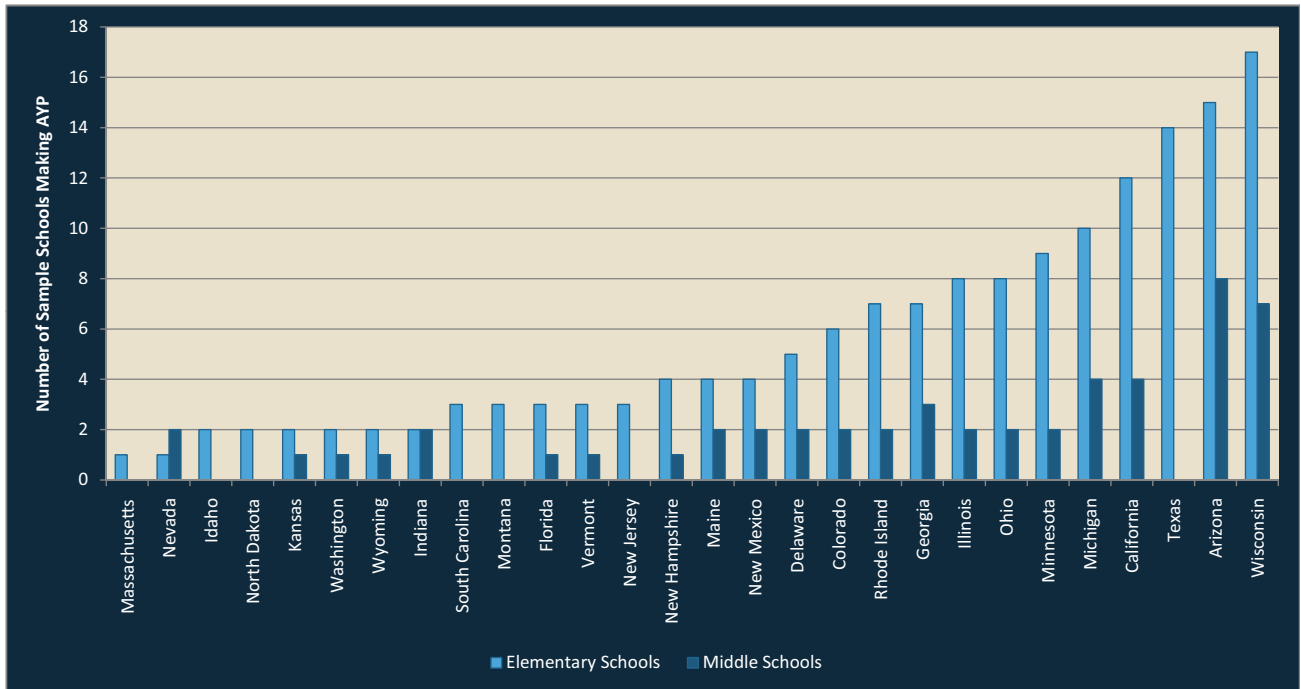


Figure 1. Number of sample schools making AYP by state

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.

- In Kansas, as in most states, schools with fewer subgroups attained AYP more easily than schools with more subgroups, even when their average student performance was much lower. In other words, schools with greater diversity and size face greater challenges in making AYP.
- As in other states, middle schools had greater difficulty reaching AYP in Kansas than did 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.
- Most states examined apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates. However, **Kansas’s 99% confidence interval give schools greater leniency than**

the more commonly used 95% confidence interval. Although the confidence interval did help a handful of schools in Kansas meet overall reading and math targets, it had little or no impact on final AYP outcomes because individual subgroups still failed to meet their targets (*all of a school’s subgroups must have met their targets for the school to make AYP*).

- A strong predictor of whether or not a school will make AYP under Kansas’s system is whether it has enough SWDs or English language learners³ to qualify as a separate subgroup. Every single school with even one such subgroup failed to make AYP.⁴

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance on Kansas’s tests and those of 25 other

³ Note that we use “LEP students” and “English language learners” interchangeably to refer to students in the same subgroup.

⁴ SWDs are defined as those students following individualized education plans. We should also note that our subgroup findings for LEP students and SWDs may be slightly more negative than actual findings, mostly because of the differences in testing practices between the Measures of Academic Progress (MAP), the assessment we used in this study, and in the Kansas Assessment Program, the standardized state test. Specifically, the U.S. Department of Education has issued NCLB guidelines permitting schools to exclude small percentages of LEP students and SWDs from taking state tests, or providing them alternate assessments. In this study, however, no valid MAP scores were omitted from consideration.

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 income⁵ 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

⁵ Low-income students are those who receive a free or reduced-price lunch.

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

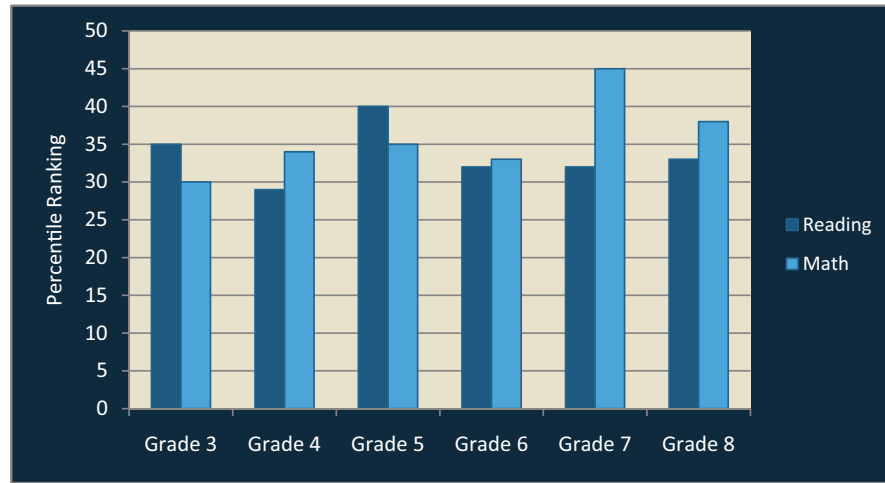


Figure 2. Kansas reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Kansas’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 Kansas’s cut scores are at or below the 45th percentile.

Proficiency cut score estimates for the Kansas Assessment System are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Kansas’s definitions of proficiency generally ranked about 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 Kansas 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 Kansas 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 Kansas’s 2008

Table 1 shows the pertinent Kansas AYP rules that were applied to elementary and middle schools in this study. Kansas’s minimum subgroup size is 30, which is slightly lower than in many of the other states we examined. This means that Kansas’s schools would have to account for more subgroups than would similar schools in other states. Furthermore, although most states also apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates, Kansas’s 99% confidence interval gives schools greater leniency than the more commonly used 95% confidence interval. So for instance,

while schools are supposed to get 75.6% of their grade 3–8 students to the “proficient” level on the state reading test, and 75.6% of the grade 3–8 students in each subgroup, applying the confidence interval means that the real target can be lower (particularly with smaller groups).⁷

Note that we were unable to examine the impact of NCLB’s “safe harbor” provision. This provision permits a school to make AYP even if some of its subgroups fail, as long as it reduces the number of nonproficient students within any failing subgroup by at least 10% relative to the previous year’s performance. Because we had access to only a single academic year’s data (2005–2006), 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.

⁷ We also 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. Kansas AYP rules for 2008

Subgroup minimum <i>n</i>	Race/ethnicity: 30	
	SWDs: 30	
	Low-income students: 30	
	LEP students: 30	
CI	Applied to proficiency rate calculations?	
	Yes; 99% CI	
AMOs	Baseline proficiency levels as of 2002 (%)	2008 targets (%)
READING/LANGUAGE ARTS		
Grade 3	67.7	75.6
Grade 4	67.7	75.6
Grade 5	67.7	75.6
Grade 6	67.7	75.6
Grade 7	67.7	75.6
Grade 8	67.7	75.6
MATH		
Grade 3	62.5	73.4
Grade 4	62.5	73.4
Grade 5	62.5	73.4
Grade 6	62.5	73.4
Grade 7	62.5	73.4
Grade 8	62.5	73.4

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

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.

How Did the Sample Schools Fare under Kansas’s AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Kansas’s 2008 AYP rules. **Only 2 elementary schools out of 16 made 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. 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 only schools actually to make AYP were those with relatively few qualifying subgroups—and thus the fewest targets to meet. For example, Wayne Fine Arts passed, but had only eight targets – two in reading and math for the overall population, two in reading and math for its low-income population, two in reading and math for its Asian/Pacific Islander population, and two for its white population.

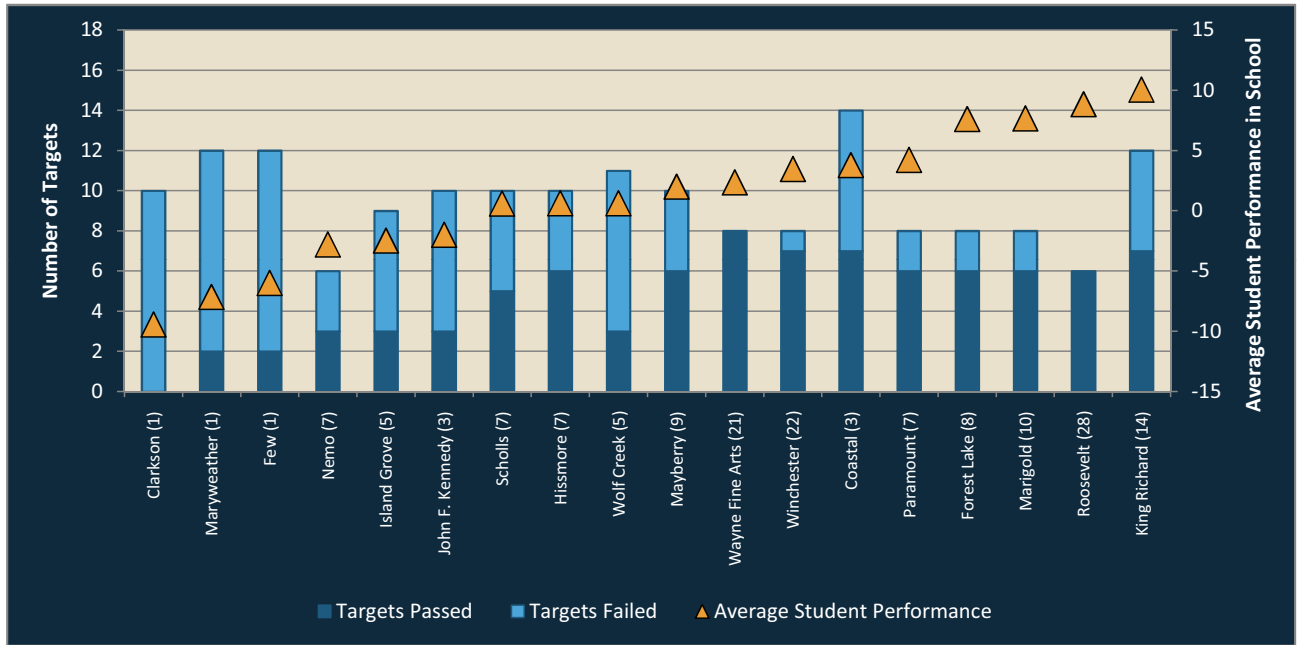


Figure 3. AYP performance of the elementary school sample under Kansas's 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under Kansas'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. Winchester Elementary, for example, met seven 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). This is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance.

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Kansas AYP rules. **Of 18 middle schools in our sample, only one made AYP** (Walter Jones), which has relatively few qualifying subgroups.

are assisted by the confidence interval (note how the orange line falls within the light blue band). We know from Figures 3 and 4, however, that all of these schools failed to make AYP because of subgroup performance.

Figures 5 and 6 indicate the degree to which schools' overall math proficiency rates are aided by Kansas'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 AMO needed to meet AYP. Figures 5 and 6 show that four of the sample elementary schools (Nemo, Island Grove, JFK, and Wolf Creek) and three middle schools (Kekata, Hoyt, and Lake Joseph)

The effect of confidence intervals on the reading proficiency rates for elementary and middle schools is much the same (not shown). In reading, four elementary schools (Hissmore, Mayberry, Coastal, and Paramount) and three middle schools (Pogesto, Hoyt, and Zeus) met the overall target with the confidence interval, although these schools still failed to meet all their subgroup targets (see Figures 3 and 4). **So, though the confidence interval does help some schools to meet overall reading and math targets, it has little or no impact on final AYP outcomes since individual subgroups failed to meet targets.**⁸

⁸ In the current analyses, confidence intervals were applied to both the overall school population and to all eligible subgroups in our sample schools. Thus, the ultimate impact of the confidence interval is likely larger than the impact depicted in Figures 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.

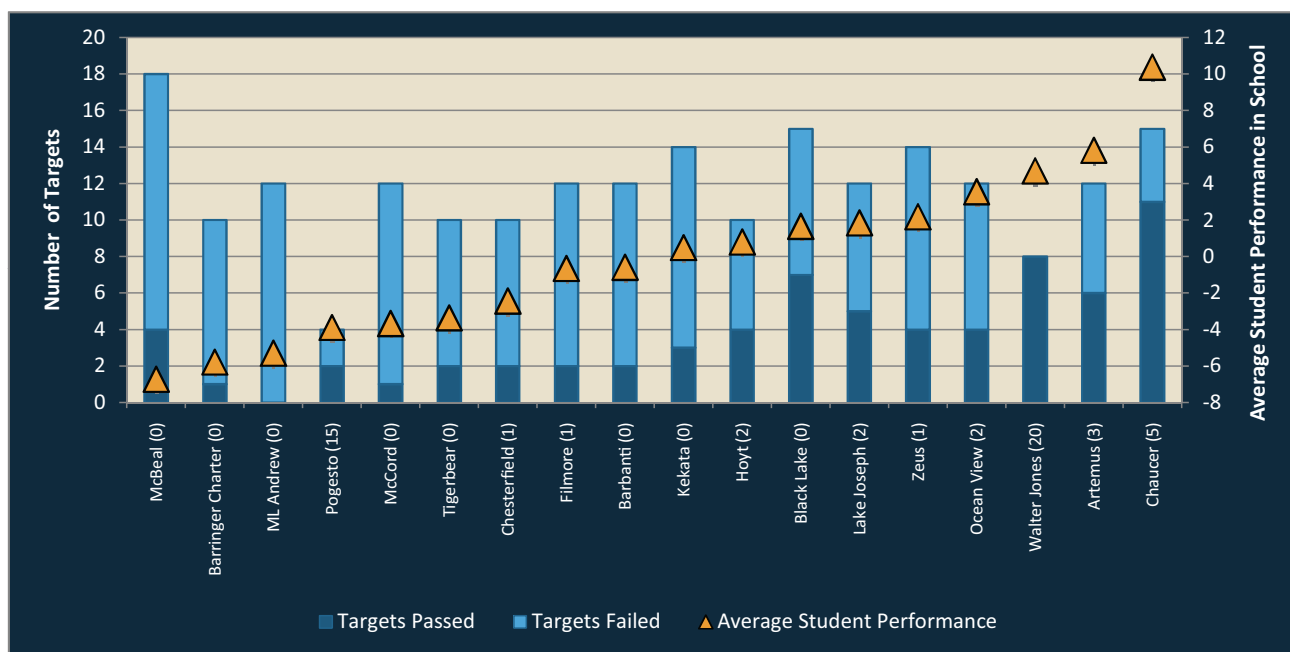


Figure 4. AYP performance of the middle school sample under Kansas's 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under Kansas'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. Chaucer, for example, met 11 of its 15 targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). This is measured by the average MAP performance of students within the school; its scale is shown on the right side of the figure. Scores below zero (which is the grade level median) denote below-grade-level performance and scores above zero denote above-grade-level performance. One unit does not equal a grade level; however, the higher the number, the better the average performance and the lower the number, the worse the average performance. The number in parentheses after each school name indicates the number of states (out of 28) in which that school would have made AYP.

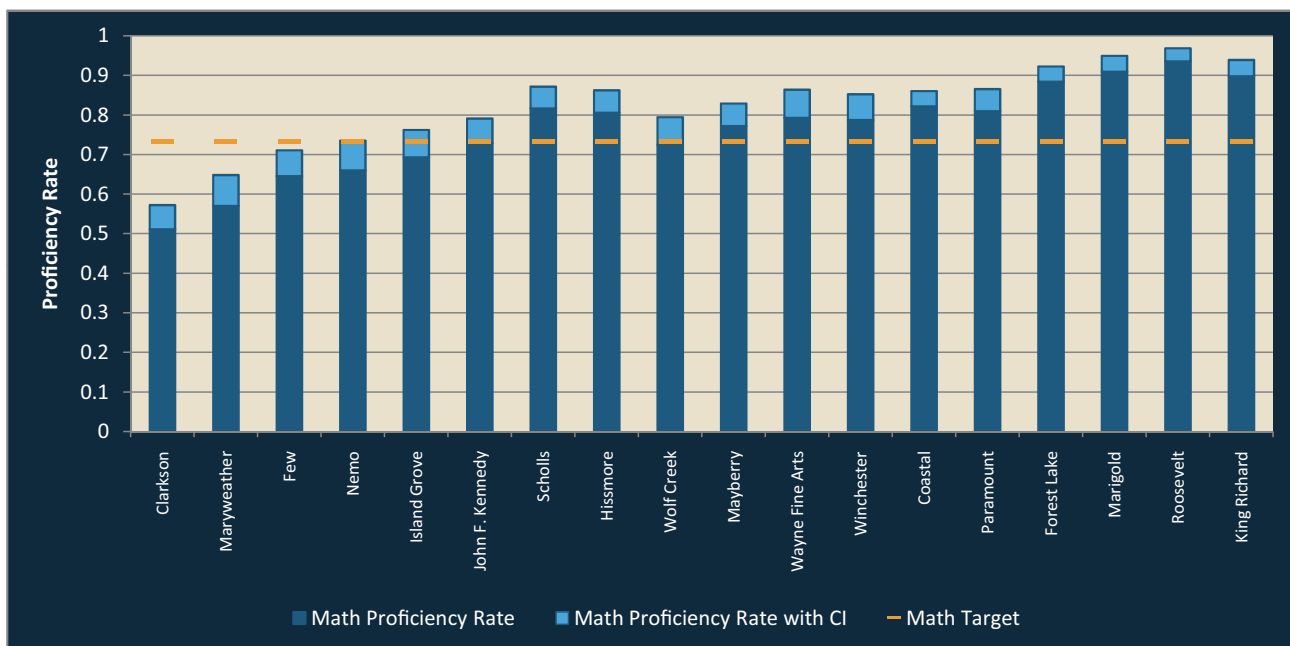


Figure 5. Impact of the confidence interval on elementary school math proficiency rates

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that four schools (Nemo, Island Grove, JFK, and Wolf Creek) were assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.

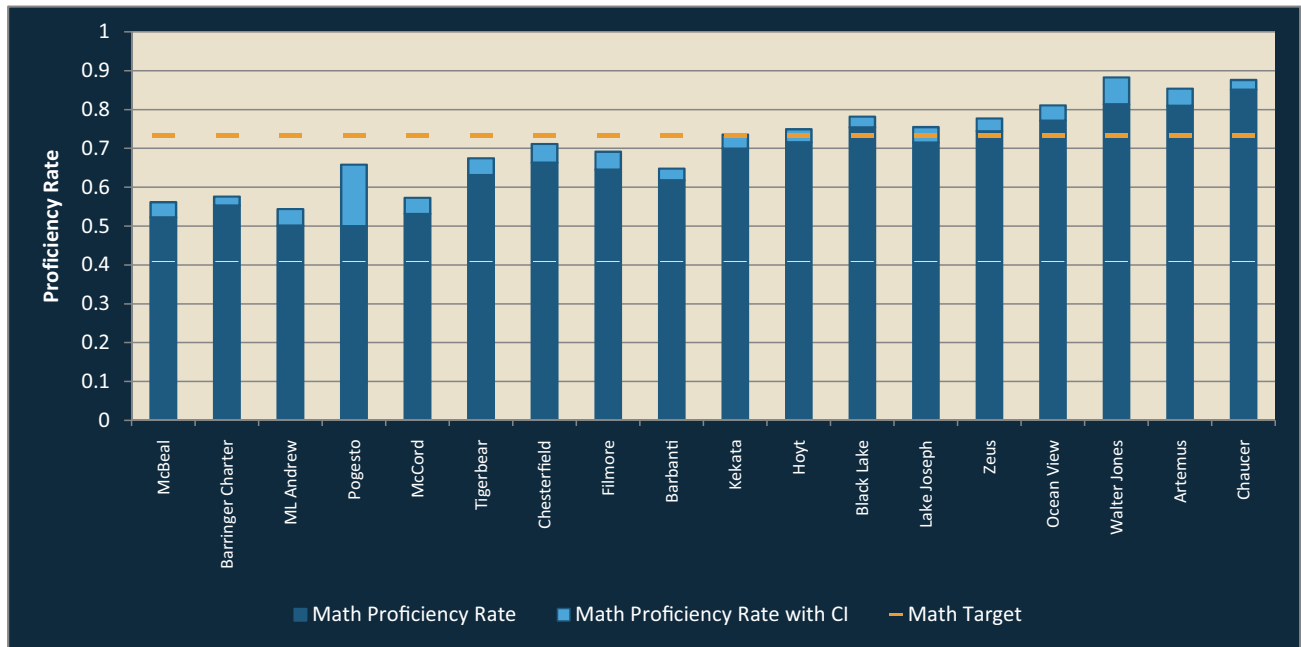


Figure 6. Impact of the confidence interval on middle school math proficiency rates under the 2008 Kansas AYP rules

Note: This figure shows the reported proficiency rate for the student population as a whole and the impact of the confidence interval on meeting annual targets. The darker portions of the bars show the actual proficiency rate achieved, while the lighter (upper) portions of the bars show the margin of error as computed by the confidence interval. The figure shows that three schools (Kekata, Hoyt, and Lake Joseph) were assisted by the confidence interval. Annual targets (the orange lines) are considered to be met by the confidence interval if they fall within the light blue portion.

Where Do Schools Fail?

Figures 3 and 4 illustrate that schools with low or mid-dling 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 purposes include SWDs, LEP 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 Kansas 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 both math and reading targets for their overall school population.
- Five other elementary schools (Nemo, Island Grove, JFK, Scholls, and Wolf Creek) in the sample failed to meet their reading targets for their overall populations.
- Eight of the 17 failing middle schools in the sample (McBeal, Barringer, ML Andrew, McCord, Tigerbear, Chesterfield, Filmore, and Barbanti) failed for both reading and math for their overall populations.
- Most schools did not make AYP because of more than one subgroup.

Tables 4 and 5 summarize the performance of the various subgroups for elementary and middle schools, respectively. We see that the performance of SWDs is proving especially challenging under the Kansas accountability system. In fact, every SWD group at the middle

Table 2. Elementary school subgroup performance of sample schools under the 2008 Kansas 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	AYP	Targets				
Clarkson	51.1%	36.1%	N	N	N	N	N	N	N	N					N	N						10	0	0%	N	1
Maryweather	57.1%	47.0%	N	N	N	N	N	N	N	N					N	N				Y	Y	12	2	17%	N	1
Few	64.6%	48.5%	N	N	N	N	N	N	N	N					Y	N				Y	N	12	2	17%	N	1
Nemo	66.0%	63.7%	Y	N					N	N									Y	Y	6	3	50%	N	7	
Island Grove	69.3%	65.4%	Y	N				N	N	N					N	N				Y	Y	9	3	33%	N	4
JFK	72.9%	57.5%	Y	N	N	N			Y	N	N	N							Y	N	10	3	30%	N	3	
Scholls	81.7%	66.5%	Y	N	N	N			Y	N	Y	N							Y	Y	10	5	50%	N	7	
Hissmore	80.6%	69.8%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	7	
Wolf Creek	72.5%	66.7%	Y	N	N	N		N	N	N					N	N				Y	Y	11	3	27%	N	5
Alice Mayberry	77.2%	71.3%	Y	Y	N	N			Y	N	Y	N							Y	Y	10	6	60%	N	9	
Wayne Fine Arts	79.3%	77.6%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	21	
Winchester	78.8%	77.3%	Y	Y	Y	N									Y	Y				Y	Y	8	7	88%	N	22
Coastal	82.2%	72.9%	Y	Y	N	N	N	N	Y	N	Y	N			Y	N				Y	Y	14	7	50%	N	3
Paramount	81.0%	73.9%	Y	Y					Y	N					Y	N				Y	Y	8	6	75%	N	7
Forest Lake	88.5%	83.3%	Y	Y	N	N			Y	Y										Y	Y	8	6	75%	N	8
Marigold	91.0%	84.1%	Y	Y	Y	N			Y	N										Y	Y	8	6	75%	N	10
Roosevelt	93.6%	90.5%	Y	Y					Y	Y										Y	Y	6	6	100%	Y	28
King Richard	89.9%	86.4%	Y	Y	N	N	Y	N	Y	N					Y	N				Y	Y	12	7	58%	N	14

Abbreviations: M = math; R = reading; N = no; Y = yes; SWDs = students with disabilities; AA = African American; Asian/Pacific Islander = Asian; Hispanic/Latino = Hispanic; American Indian/Alaska Native = AI/AN.

Note: Schools are ordered from lowest (Clarkson) to highest (King Richard) average student performance as measured by combined and weighted math and reading performance on the MAP assessment (not shown in table). A blank space underneath a subgroup means that subgroup contained fewer than the minimum number of students required for evaluation, so it wasn't counted. A "Y" in blue means that the group met the AMOs and an "N" in peach means that the group did not meet the AMOs. The two rightmost columns show (1) whether that school met AYP (i.e., it met the targets for its overall population and all required subgroups); and (2) the total number of states in the study for which that school met AYP.

school level failed to meet targets in both reading and math. A similar problem exists for students with limited English proficiency. All of those subgroups failed to meet their targets, save for one passing (in math) at the elementary level (King Richard).

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Kansas's NCLB accountability system is, in some respects, behaving like those in other states. For example, among the

elementary schools in our sample, Roosevelt and Wayne Fine Arts, made AYP in the greatest number of states—28 and 21, respectively. And these schools made AYP in Kansas, too. Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make AYP in Kansas.

But Kansas is also home to an anomaly. Winchester Elementary (see Figure 3) made AYP in 22 of the 28 states in our sample, but not in Kansas. In examining Table 2, we can see that Winchester missed only one target in reading for its SWD subgroup. This may be because

Table 3. Middle school subgroup performance of sample schools under the 2008 Kansas 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	52.3%	55.9%	N	N	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N	N	Y	Y	18	4	22%	N	0
Barringer Charter	55.3%	57.2%	N	N	N	N			N	N	N	N			Y	N						10	1	10%	N	0
ML Andrew	50.1%	59.8%	N	N	N	N			N	N	N	N			N	N				N	N	12	0	0%	N	0
Pogesto	50.0%	68.5%	N	Y																N	Y	4	2	50%	N	15
McCord Charter	53.2%	63.0%	N	N	N	N			N	N	N	N			N	N				N	Y	12	1	8%	N	0
Tigerbear	63.2%	61.0%	N	N	N	N			N	N	N	N								Y	Y	10	2	20%	N	0
Chesterfield	66.3%	63.0%	N	N	N	N			N	N	N	N								Y	Y	10	2	20%	N	1
Filmore	64.6%	71.1%	N	N	N	N	N	N	N	N	N				N	N				Y	Y	12	2	17%	N	1
Barbanti	61.8%	66.2%	N	N	N	N	N	N	N	N	N				N	N				Y	Y	12	2	17%	N	0
Kekata	69.9%	69.0%	Y	N	N	N	N	N	N	N	N	N			N	N				Y	Y	14	3	21%	N	0
Hoyt	71.6%	72.4%	Y	Y	N	N			N	N	N	N								Y	Y	10	4	40%	N	2
Black Lake	75.4%	72.4%	Y	N	N	N	N		N	N	N	N	Y	Y	Y	Y				Y	Y	15	7	47%	N	0
Lake Joseph	71.5%	76.9%	Y	Y	N	N	N	N	N	N	Y				N	N				Y	Y	12	5	42%	N	2
Zeus	74.4%	74.4%	Y	Y	N	N	N	N	N	N	N	N			N	N				Y	Y	14	4	29%	N	1
Ocean View	77.2%	83.4%	Y	Y	N	N	N	N	N	N					N	N				Y	Y	12	4	33%	N	2
Walter Jones	81.4%	81.1%	Y	Y					Y	Y					Y	Y				Y	Y	8	8	100%	Y	20
Artemus	81.0%	81.8%	Y	Y	N	N			N	N			Y	Y	N	N				Y	Y	12	6	50%	N	3
Chaucer	85.2%	88.1%	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y				Y	Y	16	12	75%	N	5

Abbreviations: M = math; R = reading; N = no; Y = yes; SWDs = students with disabilities; AA = African American; Asian/Pacific Islander = Asian; Hispanic/Latino = Hispanic; American Indian/Alaska Native = AI/AN.

Note: Schools are ordered from lowest (McBeal) to highest (Chaucer) average student performance as measured by combined and weighted math and reading performance on the MAP assessment (not shown in table). A blank space underneath a subgroup means that subgroup contained fewer than the minimum number of students required for evaluation, so it wasn't counted. A "Y" in blue means that the group met the AMOs and an "N" in peach means that the group did not meet the AMOs. The two rightmost columns show (1) whether that school met AYP (i.e., it met the targets for its overall population and all required subgroups); and (2) the total number of states in the study for which that school met AYP.

Kansas's minimum subgroup size is somewhat smaller than in most other states examined, meaning that school may have more accountable subgroups under Kansas rules than it would in other states.

This is consistent with the patterns shown in Table 6, which compares 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 (especially at the middle school level),

fewer subgroups (and thus fewer targets to meet), and lower percentages of low-income and nonwhite students.

Concluding Observations

This study examined the test performance data of students from 18 elementary and 18 middle schools across the country to see how these schools would fare under Kansas's AYP rules (and AMOs) for 2008. We found that only 2 elementary schools and 1 middle school—3 out of a sample of 36— make AYP in Kansas. Looking across the 28 state accountability systems examined in the

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Kansas 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	13	11	13
Students with limited English proficiency	7	4	7
Low-income students	17	6	14
African-American students	6	1	5
Asian/Pacific Islander students	0	0	0
Hispanic students	9	4	8
American Indian/Alaska Native students	0	0	0
White students	17	0	2

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Kansas 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	16	16
Students with limited English proficiency	9	9	8
Low-income students	17	15	14
African-American students	11	10	10
Asian/Pacific Islander students	4	0	0
Hispanic students	14	10	11
American Indian/Alaska Native students	1	1	1
White students	17	3	1

study, this puts Kansas at the low end of the sample distribution in terms of the number of schools making AYP (see Figure 1). Part of the reason that Kansas has so many schools not making AYP is that its annual targets are somewhat high (roughly 75% of students were expected to meet targets in 2008).

The overriding goal of the NCLB is to eliminate education 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, Kansas'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 minority youngsters. Many of the sample schools met the Kansas reading and math targets for their student populations as

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	2	16	1	17
Average student body size	243	312	165	900
Average % low income	18	50	38	45
Average % nonwhite	25	43	33	45
Average performance [†]	5.61	0.68	4.69	-0.33
Average % growth [‡]	100	117	111	97
Average number of targets to meet	7	10	8	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.

a whole, that is, without considering subgroup results. In the pre-NCLB era, such schools might have been considered effective or at least not in need of improvement, even though sizable numbers of their students 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 the student population 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,⁹ doesn't the massive failure of middle school students to meet Kansas's targets indicate that a new approach is needed for holding schools accountable for the performance of these students? Yes, schools should redouble their efforts to boost achievement for LEP students and SWDs, as for other pupils, 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

⁹ See footnote 4.

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.