Illinois

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 the NCLB accountability system in Illinois—particularly how its various rules, criteria, and practices result in schools either making AYP or not making AYP. It also gauges how tough the Illinois 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 the Illinois 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 the Illinois AYP rules for academic year 2007–2008 (shortened to "2008" in this report).

Here are some key findings:

- We estimate that **10 out of 18 elementary schools** and **16 of 18 middle schools** in our sample **failed to make adequate yearly progress** in 2008 under the Illinois 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 six states exceed Illinois in terms of the number of elementary schools making AYP (Minnesota, Michigan, California, Texas, Arizona, and Wisconsin)(see Figure 1)**. Illinois ties with Ohio, each with eight (out of 18) elementary schools making AYP.
- Nearly all the schools in our sample that failed to make AYP in Illinois are meeting expected targets for their overall populations but failing because of the performance of individual subgroups.[2]
- Two sample schools made AYP in Illinois that failed to make AYP in most other states. This is likely because the **proficiency standards in Illinois are relatively easy compared to those in other states**; these schools also have fewer accountable subgroups (**the minimum *n* size in Illinois is a bit higher than in many other states, meaning it takes more students within a subgroup category for that group to be held separately accountable**).

There are several contributing factors to the relatively high number of sample schools that make AYP in **Illinois**. The first is the state's low proficiency standards (or cut scores). In reading, cut scores range from the 35th to the 22nd percentile, and in math, cut scores range from the 20th to the 15th percentile. Moreover, schools have fewer accountable subgroups in Illinois than in most other states because Illinois's minimum subgroup size (45) is a bit higher than the minimum in many other states. In other words, Illinois schools must have more students within a subgroup category in order for that group to be held separately accountable. With fewer accountable subgroups and relatively easy proficiency standards, it's possible for more study schools to make AYP in Illinois.

[1] A cut score is the minimum score a student must receive on NEWA's Measures of Academic Progress (MAP) that is equivalent to performing proficient on the Illinois Standards Achievement Test.

[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.

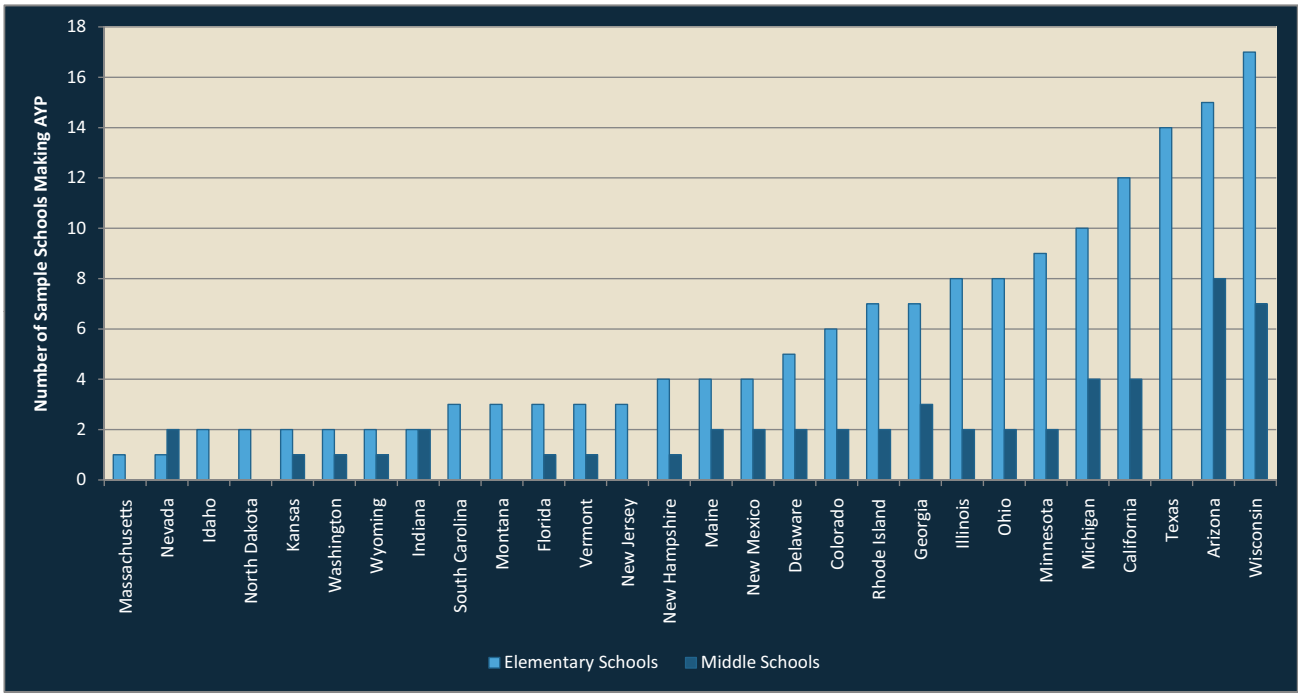

Figure 1. Number of sample schools making AYP by state

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.

- As in other states, schools with fewer subgroups attain AYP more easily in Illinois 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.

- Middle schools have greater difficulty reaching AYP in Illinois than do elementary schools, primarily because their student populations are larger and they therefore have more qualifying subgroups—not because their student achievement is lower than in the elementary schools.

- A strong predictor of whether or not a school will make AYP under the Illinois system is whether it has enough students with disabilities (SWD) or English language learners to qualify as a separate subgroup. All but one school with limited English proficient (LEP)[3] or SWD subgroups failed to make AYP, in part because these students did not meet the state's annual proficiency targets, especially in reading.[4]

Introduction

The Proficiency Illusion (Cronin et al. 2007a) linked student performance on Illinois' 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

[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 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 Illinois Standards Achievement 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.

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[5] or African American among others), that must reach the proficient level in order for the school to make AYP in a given year. These AMOs vary by state (as do, of course, the difficulty of the proficiency standards).

States also determine the minimum number of students that must constitute a subgroup in order for its scores to be analyzed separately (also called the minimum *n* [number of students in sample] size). The rationale is that reporting the results of very small subgroups—fewer than ten pupils, for example—could 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 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]

Proficiency cut score estimates for the Illinois Standards Achievement Test (ISAT) were taken from *The Proficiency Illusion* (as shown in Figure 2), which found that the definitions of reading and math proficiency in Illi-

[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.

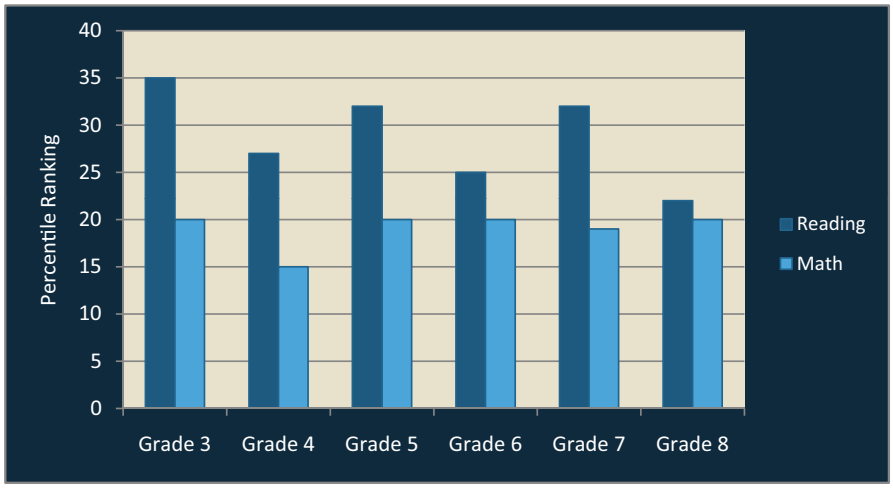

Figure 2. Illinois reading and math cut score estimates, expressed as percentile ranks (2006))

Note: This figure illustrates the difficulty of the cut scores (or proficiency passing scores) in Illinois for the state's 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 the state's cut scores are at or below 35th percentile.

nois were generally below average, or among the less difficult of the 26 states examined in that study. These cut scores were used to estimate whether students would have scored as proficient or better on the Illinois 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 Illinois AYP rules for 2008. In other words, the school data and our proficiency cut score estimates are from 2005–2006, but we are applying them against the Illinois 2008 AYP rules.

Table 1 shows the pertinent Illinois AYP rules that were applied to elementary and middle schools in the current study. The state's minimum subgroup size is 45, which is slightly larger than most other states we examined. Most states examined also apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates. So, although schools are supposed to get 62.5% of their students to the proficient level on the state reading test, as well as 62.5% of their students in each subgroup, applying the confidence interval means that the real targets can actually be lower (particularly with smaller groups). In other words, using a confidence interval makes it easier for Illinois schools to meet the targets as defined by their AMOs.[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.

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

[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. Illinois AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 45	
	SWDs: 45	
	Low-income students: 45	
	LEP students: 45	
CI	**Applied to proficiency rate calculations?**	
	Yes; 95% CI	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	47.5	62.5
Grade 4	47.5	62.5
Grade 5	47.5	62.5
Grade 6	47.5	62.5
Grade 7	47.5	62.5
Grade 8	47.5	62.5
MATH		
Grade 3	47.5	62.5
Grade 4	47.5	62.5
Grade 5	47.5	62.5
Grade 6	47.5	62.5
Grade 7	47.5	62.5
Grade 8	47.5	62.5

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

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 the Illinois' AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under the 2008 AYP rules in Illinois. **Eight elementary schools made AYP and ten 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. Six of the eight passing schools are on the right half of the figure, meaning that students with the highest average academic performance 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.

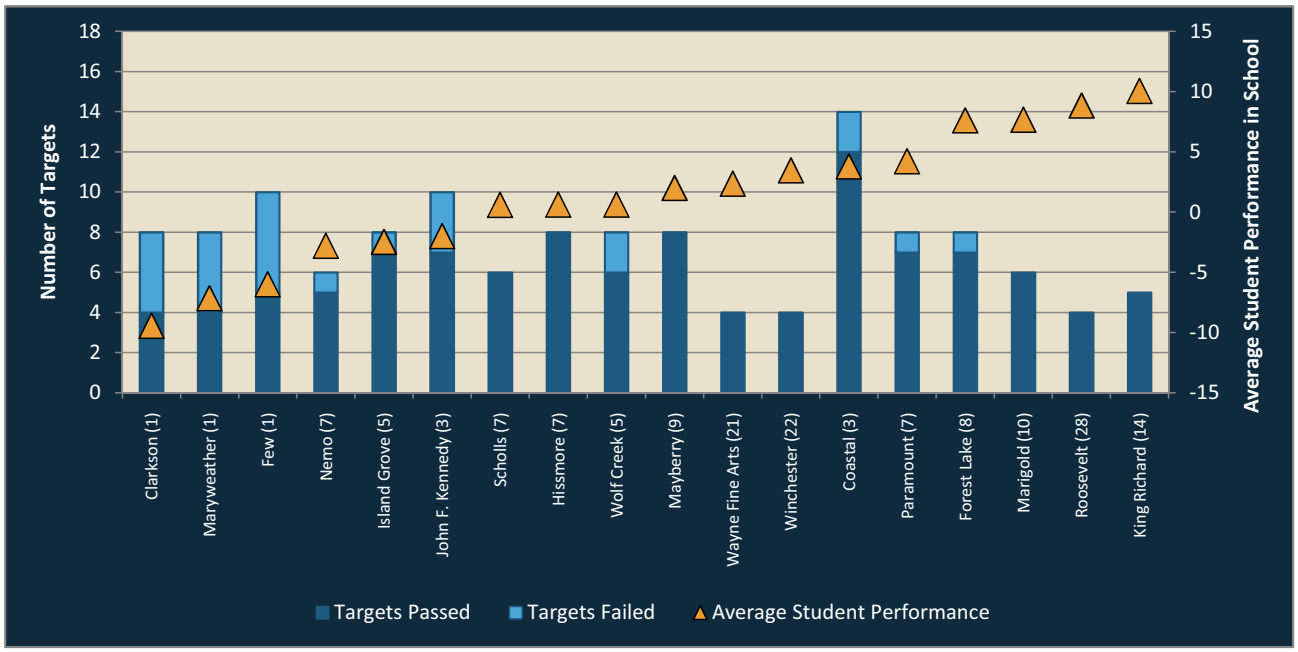

Figure 3. AYP performance of the elementary school sample under the Illinois 2008 AYP rules

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

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Illinois AYP rules. **Of 18 middle schools in our sample, only 2 passed**—1 low-performance school (Pogesto) and 1 high-performance school (Walter Jones), both of which have relatively few qualifying subgroups and consequently few targets to meet.

Figure 5 indicates the degree to which elementary schools' overall math proficiency rates are aided by the confidence interval. On this figure, the darker portions of the bars show the actual proficiency rates at each school and the lighter portions of the bars show the degree to which the proficiency rates were increased by applying the confidence interval. The orange lines show the AMOs needed to meet AYP. This figure shows that none of the sample elementary schools was assisted by the confidence interval, because **the math targets in Illinois are low relative to the schools' overall performance**. The pattern is much the same for middle school math proficiency rates and for reading proficiency rates at the elementary and middle school levels (not shown). In fact, only one elementary school, John F. Kennedy, received assistance from the confidence interval in meeting its reading target. As shown in Figure 3, however, this school still failed to meet all its subgroup targets. **Overall, applying the confidence interval had little or no effect on whether schools made AYP in Illinois**.[8]

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, thanks to fewer subgroups. These figures do not, however, indicate which subgroups

[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 may be 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.

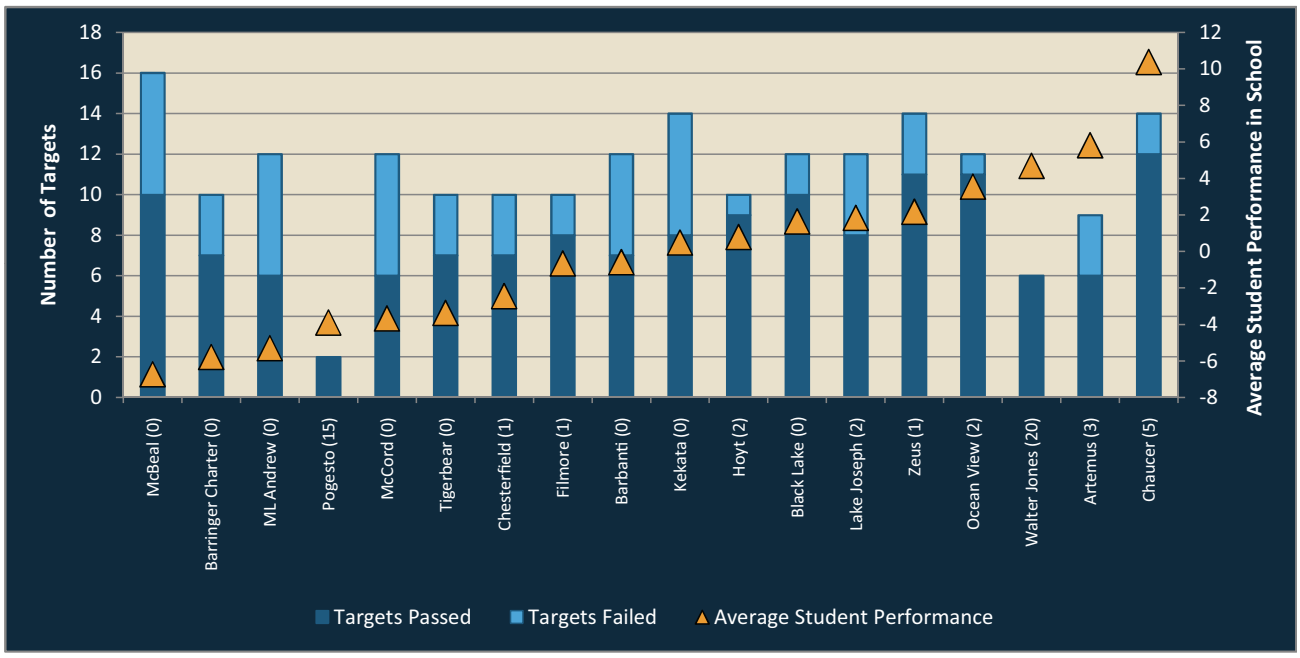

Figure 4. AYP performance of the middle school sample under the Illinois 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under the AYP rules in Illinois (as described in Table 1). The bars show the number of targets that each school had to meet in order to make AYP under the state's NCLB rules, and whether they met them (dark blue) or did not meet them(light blue). The more subgroups in a school, the more targets it must meet. Under the study conditions, a school that failed to meet the AMO for even a single subgroup didn't make AYP, so any light blue means the school failed. Artemus Middle School, for example, met six of its nine 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.

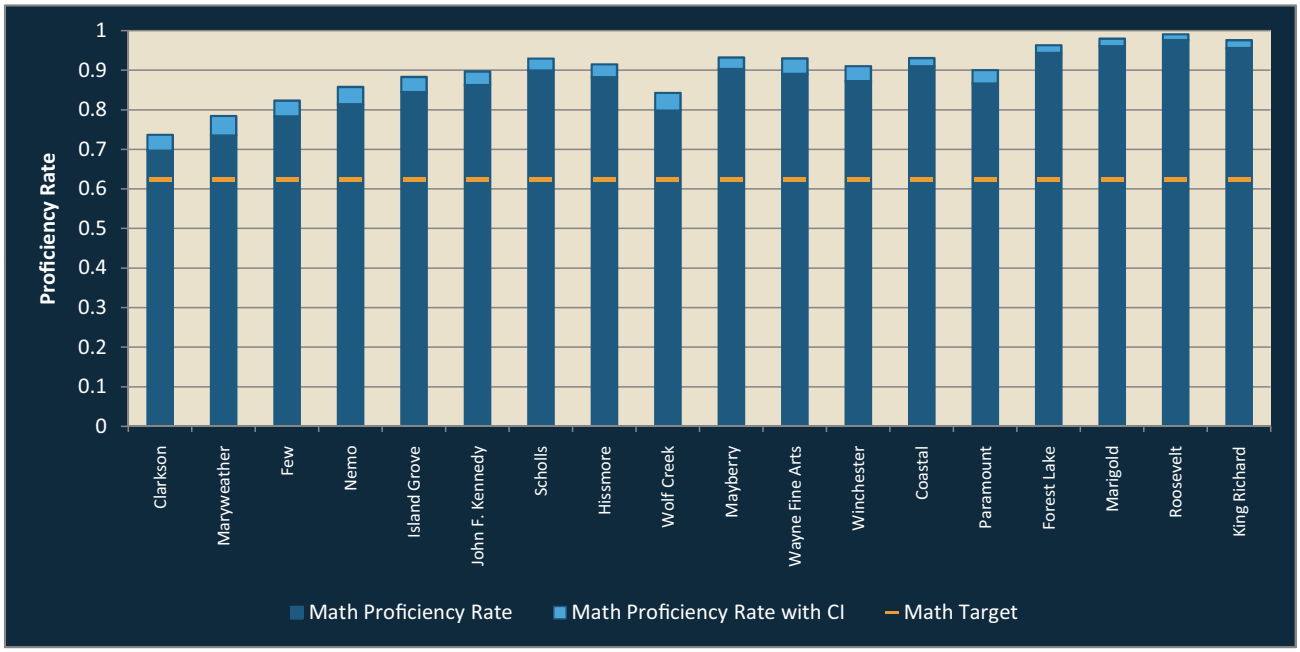

Figure 5. Impact of the confidence interval on elementary school math proficiency rates

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, 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 Illinois 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	69.7%	40.8%	Y	N			Y	N	Y	N					Y	N					8	4	50%	N	1
Maryweather	73.5%	52.1%	Y	N			Y	N	Y	N					Y	N					8	4	50%	N	1
Few	78.4%	51.9%	Y	N	Y	N	Y	N	Y	N					Y	N					10	5	50%	N	1
Nemo	81.4%	68.4%	Y	Y					Y	N									Y	Y	6	5	83%	N	7
Island Grove	84.5%	68.3%	Y	Y					Y	Y					Y	N			Y	Y	8	7	88%	N	4
JFK	86.2%	61.6%	Y	Y	Y	N			Y	N	Y	N							Y	Y	10	7	70%	N	3
Scholls	89.9%	69.9%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	7
Hissmore	88.2%	73.3%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	7
Wolf Creek	79.8%	68.9%	Y	Y					Y	N					Y	N			Y	Y	8	6	75%	N	5
Alice Mayberry	90.3%	74.0%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	9
Wayne Fine Arts	89.1%	81.0%	Y	Y															Y	Y	4	4	100%	Y	21
Winchester	87.3%	79.6%	Y	Y															Y	Y	4	4	100%	Y	22
Coastal	91.0%	76.5%	Y	Y	Y	N	Y	N	Y	Y	Y	Y			Y	Y			Y	Y	14	12	86%	N	3
Paramount	86.6%	76.5%	Y	Y					Y	N					Y	Y			Y	Y	8	7	88%	N	7
Forest Lake	94.4%	84.9%	Y	Y	Y	N			Y	Y									Y	Y	8	7	88%	N	8
Marigold	96.0%	86.3%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	10
Roosevelt	97.6%	92.5%	Y	Y															Y	Y	4	4	100%	Y	28
King Richard	95.6%	89.1%	Y	Y	Y														Y	Y	5	5	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.

failed 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 Indian/Alaska Native, and white. Tables 2 and 3 also show whether a school met AYP under the 2008 Illinois 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 overall school population.

Table 3. Middle school subgroup performance of sample schools under the 2008 Illinois 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	69.9%	62.6%	Y	Y	N	N	N	N	Y	N	Y	Y			Y	N	Y	Y	Y	Y	16	10	63%	N	0
Barringer Charter	76.2%	63.2%	Y	Y	N	N			Y	Y	Y	N			Y	Y					10	7	70%	N	0
ML Andrew	73.3%	66.9%	Y	Y	N	N			N	N	N	N			Y	Y			Y	Y	12	6	50%	N	0
Pogesto	74.1%	74.1%	Y	Y																	2	2	100%	Y	15
McCord Charter	74.7%	69.6%	Y	Y	N	N			N	N	N	N			Y	Y			Y	Y	12	6	50%	N	0
Tigerbear	79.1%	66.8%	Y	Y	N	N			Y	Y	Y	N							Y	Y	10	7	70%	N	0
Chesterfield	83.6%	70.7%	Y	Y	N	N			Y	Y	Y	N							Y	Y	10	7	70%	N	1
Filmore	82.9%	76.8%	Y	Y	N	N			Y	Y					Y	Y			Y	Y	10	8	80%	N	1
Barbanti	77.8%	71.7%	Y	Y	N	N	N	N	Y	N					Y	Y			Y	Y	12	7	58%	N	0
Kekata	86.0%	73.9%	Y	Y	N	N	N	N	Y	Y	Y	N			Y	N			Y	Y	14	8	57%	N	0
Hoyt	88.4%	77.6%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	2
Black Lake	89.2%	78.6%	Y	Y	N	N			Y	Y	Y	Y			Y	Y			Y	Y	12	10	83%	N	0
Lake Joseph	86.6%	81.3%	Y	Y	N	N	N	N	Y	Y					Y	Y			Y	Y	12	8	67%	N	2
Zeus	89.9%	79.3%	Y	Y	Y	N	Y	N	Y	Y	Y	Y			Y	N			Y	Y	14	11	79%	N	1
Ocean View	90.6%	87.5%	Y	Y	Y	Y	Y	N	Y	Y					Y	Y			Y	Y	12	11	92%	N	2
Walter Jones	91.3%	85.1%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	20
Artemus	92.0%	84.4%	Y	Y		N			Y	N					Y	N			Y	Y	9	6	67%	N	3
Chaucer	94.3%	90.8%	Y	Y	Y	N	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.

- No elementary schools failed to meet math targets.
- All sample middle schools met overall targets in both reading and math.
- One elementary school (Forest Lake) and three middle schools (Filmore, Hoyt, and Black Lake) missed only because of the SWD subgroup.
- One middle school (Ocean View) failed only because of its LEP subgroup.
- Two elementary schools (Nemo and Paramount) failed only because of their low-income subgroup.
- 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. First, elementary students did well on the state math test, perhaps because the Illinois proficiency cut scores are easier in math than in reading at the elementary grades (see Figure 2). Second, the performance of SWDs and LEP students are proving most challenging for schools under the Illinois system, particularly in middle schools, where this subgroup tends to have enough students to meet the state's

Table 4. Summary of subgroup performance of sample elementary schools under the 2008 Illinois 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	5	0	4
Students with limited English proficiency	4	0	4
Low-income students	14	0	7
African-American students	4	0	1
Asian/Pacific Islander students	0	0	0
Hispanic students	7	0	5
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 Illinois 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	15	11	15
Students with limited English proficiency	7	4	7
Low-income students	17	2	5
African-American students	10	2	6
Asian/Pacific Islander students	1	0	0
Hispanic students	13	0	4
American Indian/Alaska Native students	1	0	0
White students	16	0	0

minimum *n* of 45. Students with LEP are also struggling to meet the state's targets; every single school with a large enough LEP population to qualify as a separate subgroup fails to meet its reading targets for these students.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that the NCLB accountability system in Illinois 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 Illinois, too. Likewise, the elementary and middle schools that failed to make AYP in the greatest number of states also failed to make AYP in Illinois.

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	8	10	2	16
Average student body size	255	344	124	951
Average % low income	34	56	42	45
Average % nonwhite	32	48	27	46
Average performance†	4.45	-1.36	0.40	-0.11
Average % growth‡	113	117	109	97
Average number of targets to meet	6	9	4	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.

But Illinois 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 Illinois. In examining Table 2, one can see that Mayberry didn't meet the state's minimum numbers for the LEP or SWD subgroups, which created difficulty for so many other schools within the sample. This is likely a reflection of the fact that the Illinois minimum subgroup size, 45, is a bit higher than in many other states. **In other words, Illinois schools must have more students within a subgroup category for that group to be held separately accountable. With fewer accountable subgroups, and with relatively easy proficiency standards (Figure 2), Mayberry made AYP, even when other schools with higher average performance failed.**

Second, look at Pogesto Middle School (Figure 4). Even with its relatively low average performance, it makes AYP in Illinois, but fails to do so in 13 of 28 states. Like Mayberry, its AYP success in Illinois is likely attributable to the relatively small number of targets (two) it has to meet, as shown in Table 3, along with the relatively easy proficiency standards in Illinois, compared to other states.

This is consistent with the patterns shown in Table 6, which compares the schools that did and didn't make AYP on several 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 lower percentages of low-income students. Similarly, middle schools that made AYP have slightly higher performing students, on average, than middle schools that didn't, but have dramatically 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 from 18 elementary and 18 middle schools across the country to see how these schools would have fared under the Illinois AYP rules (and AMOs) for 2008. We found that 8 elementary schools and 2 middle schools—10 in all from a sample of 36—would have made AYP in Illinois. Looking across the 28 state accountability systems examined in the study, this puts Illinois towards

the upper end of the distribution in terms of the number of schools making AYP (as shown in 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, the NCLB accountability system in Illinois is working exactly as Congress intended: identifying as needing attention those 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 met the Illinois AMO targets for their student populations as a whole, i.e., not 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 were not 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,[9] does the massive failure of middle school students to meet Illinois targets indicate that a new approach is needed for holding schools accountable for the performance of these students? Yes, schools should redouble their efforts to boost achievement for ELL students and students with disabilities, as for other students, but when 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 not by themselves measure the effectiveness of the entire state accountability system, of which there are many parts.

[9] See footnote 4.

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.