Michigan

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

Here are some key findings:

- We estimate that **8 of 18 elementary schools and 14 of 18 middle schools** in our sample **failed to make AYP** in 2008 under Michigan'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 the number of elementary schools that made AYP in Michigan is exceeded in just 4 other sample states (California, Texas, Arizona, Wisconsin). In addition, Michigan is one of just a handful of states where four or more middle schools made AYP (see Figure 1).**[2]
- Every school in our sample that failed to make AYP in Michigan met expected targets for their overall population but failed because of the performance of individual subgroups, particularly students with disabilities (SWDs) and English language learners.[3]
- Seven sample schools that made AYP in Michigan failed to make AYP in most other states. This is likely because Michigan's proficiency standards are relatively easy, compared to other states, and these schools generally have fewer accountable subgroups.

Compared with other states in the study, **Michigan** is at the high end of the distribution in terms of how many sample schools make AYP. One could attribute this to a number of factors. First, Michigan's proficiency standards (or cut scores) are relatively easy compared to other states in the study (none are above the 35th percentile according to NWEA norms). An additional factor is that unlike most states, which apply a confidence interval (margin of error) to measurements of group proficiency rates, Michigan applies a standard error to individual student scores. This increases the number of students whose scores are considered passing. A final contributing factor to the large number of schools making AYP in Michigan is that the state applies different annual targets for different grades and subjects (e.g., 54% of grade 8 students in reading are expected to reach proficiency in 2008; that number changes to 65% for grade 3 math students).

[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 Michigan Educational Assessment Program (MEAP).

[2] It's important to note that Michigan received full and immediate approval from the U.S. Department of Education in 2008 to implement a student growth model in 2007–2008. This analysis, which draws on data from 2005–2006, does not in any way use or incorporate Michigan's student growth model calculations.

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

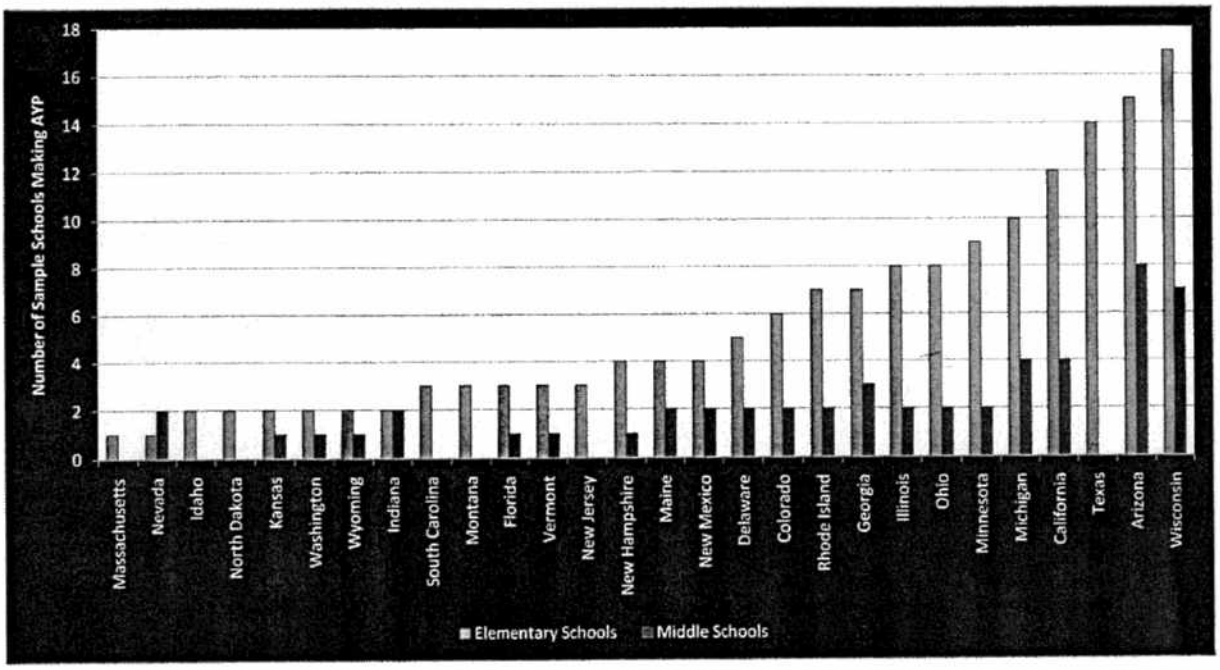

Figure 1. Number of sample schools making AYP by state

Note: Middle schools were not included for Texas and New Jersey; absence of a middle school bar in those states means "not applicable" as opposed to zero. States like Idaho and North Dakota, however, have zero passing middle schools.

- Schools with fewer subgroups attained AYP more easily in Michigan 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. This is the case in other states as well.
- Middle schools had greater difficulty reaching AYP in Michigan than did 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 a school making AYP under Michigan's system is whether it has enough SWDs to qualify as a separate subgroup. More than half of the schools with enough qualifying SWDs failed to meet their AYP targets.[4]

Introduction

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

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

dards 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. 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,[6] among other subgroups.

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

Proficiency cut score estimates for the Michigan Educational Assessment Program (MEAP) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Michigan's definitions of proficiency ranked below the average compared with the standards set by the other 25 states in that study. These cut scores were used to estimate whether students would have scored as proficient or better on the Michigan test, given their performance on MAP. Student test data and subgroup designations were then used to determine how these 18 elementary

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

[6] Note that we use "LEP students" and "English language learners" interchangeably to refer to students in the same subgroup.

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

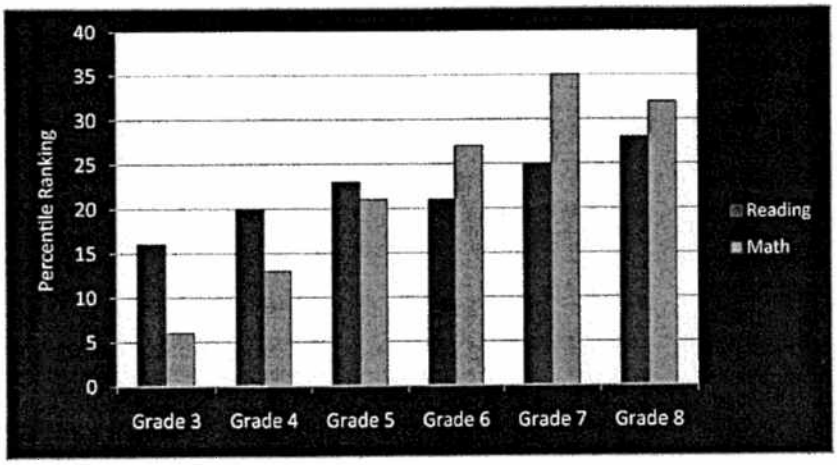

Figure 2. Michigan reading and math cut score estimates, expressed as percentile ranks (2006)

Note: This figure illustrates the difficulty of Michigan'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 Michigan's cut scores are at or below the 35th percentile.

and 18 middle schools would have fared under Michigan 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 Michigan's 2008 AYP rules.

Table 1 shows the pertinent Michigan AYP rules that we applied to elementary and middle schools in this study. Michigan employs a "sliding" minimum subgroup size of 30 or 1% of the school population, whichever is larger, up to a maximum of 200 students.[8] Thirty is a smaller number than is used in most states, which helps ensure that smaller subgroups will still be accountable. Most states, however, employ a fixed number rather than a sliding one, increasing the likelihood that larger schools will be accountable for more subgroups than small schools.

Unlike most states, which apply a confidence interval to measurements of group proficiency rates, Michigan applies standard errors to individual student scores. Technically, this is a more appropriate strategy than using confidence intervals—that is, if the motivation is to correct for test measurement error. However, rather than treating the measurement error correctly (a student's "true" score could be higher OR lower), Michigan merely *adds* the standard error to the student's score, making it easier for students to achieve proficiency on the state test (thus the technical advantage of using standard errors over confidence intervals is lost). Ironically enough, all of the states in the study that use confidence intervals follow essentially this same practice, by treating the margin of error as if it only went in one direction—the one favoring school outcomes. Strictly speaking, such practices cannot be justified purely by a desire to correct for measurement error, because measurement error is seldom unidirectional.

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.

[8] In Michigan, the minimum subgroup size is generally 1% of the total school population. Overall, this means that the subgroup size grows with the school size. However, there's also a clause that specifies the minimum subgroup size can't be less than 30 or more than 200. For example, a school with a total population of 3900 would have a minimum subgroup size of 39 (i.e., 1%), but a school with only 900 students would have a minimum subgroup size of 30, since 1% of 900 (i.e., 9) is below the minimum. Similarly, a hypothetical school of 25,000 would have a minimum subgroup size of 200, since 1% of 25,000 (i.e., 250) is greater than the maximum value.

Table 1. Michigan AYP rules for 2008

Subgroup minimum *n*	Race/ethnicity: 1% of school population, but can't be less than 30 or more than 200	
	SWDs: 1% of school population, but can't be less than 30 or more than 200	
	Low-income students: 1% of school population, but can't be less than 30 or more than 200	
	LEP students: 1% of school population, but can't be less than 30 or more than 200	
CI	**Applied to proficiency rate calculations?**	
	Not used, but 2 standard errors added to individual test scores	
AMOs	**Baseline proficiency levels as of 2002 (%)**	**2008 targets (%)**
READING/LANGUAGE ARTS		
Grade 3	38	59
Grade 4	38	59
Grade 5	38	59
Grade 6	31	54
Grade 7	31	54
Grade 8	31	54
MATH		
Grade 3	47	65
Grade 4	47	65
Grade 5	47	65
Grade 6	31	54
Grade 7	31	54
Grade 8	31	54

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

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.

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

Figure 3 illustrates the AYP performance of the sample elementary schools under Michigan's 2008 AYP rules. **Ten elementary schools made AYP and eight failed to make it.** The triangles in the figure 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

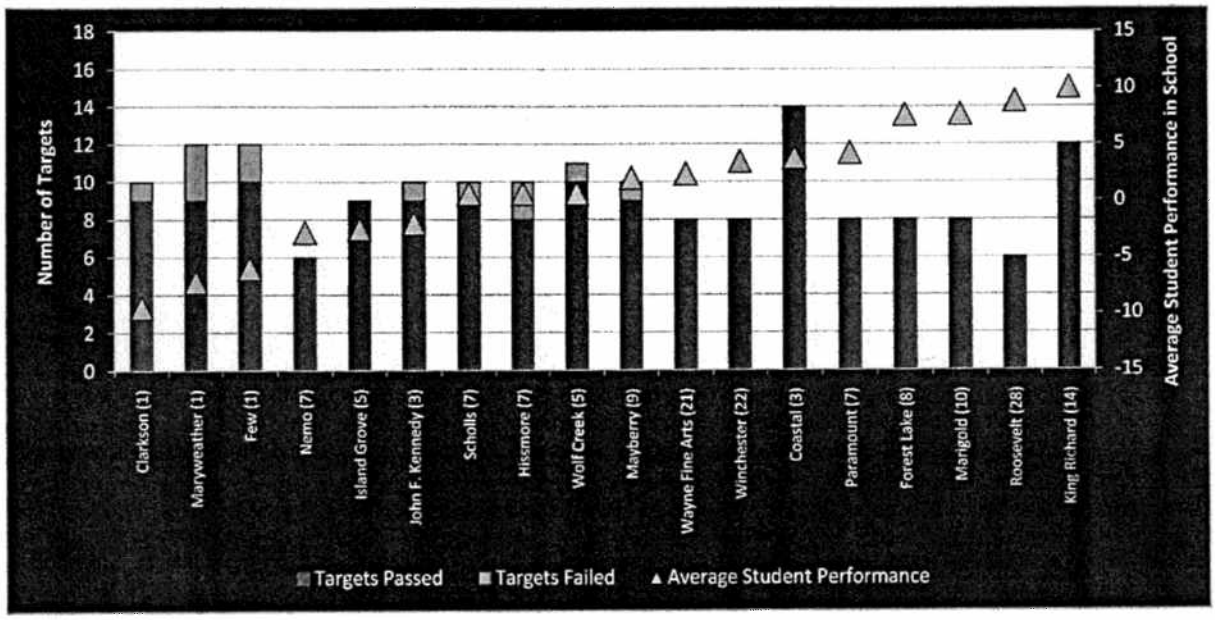

Figure 3. AYP performance of the elementary school sample under the Michigan 2008 AYP rules

Note: This figure indicates how each of the elementary schools within the sample fared under Michigan'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 9 of its 10 targets, but because it didn't meet them all, it didn't make AYP. Schools are ordered from lowest to highest average student performance (shown by the orange triangles). 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.

schools making AYP are in the right half of the figure, meaning that the highest performing students were found at these schools.

Of the schools with lower performing students, the only ones that made AYP are those with relatively few qualifying subgroups—and thus the fewest targets to meet. For example, Nemo and Island Grove made AYP but have only six and nine targets each, respectively. Each had to make AYP for its overall student population in reading and math (two targets), for its low-income population (two targets), and for its white population (two more targets). Island Grove also had to make AYP for its LEP population in reading (one target) and for its Hispanic population (two targets).

Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Michigan AYP rules. Of **18 in our sample, only 4 made AYP**—one low-performance school (Pogesto), one middle-performance school (Hoyt), and two high-performance schools (Walter Jones and Chaucer). All but Chaucer (the highest performing school in the sample) have relatively few qualifying subgroups.

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

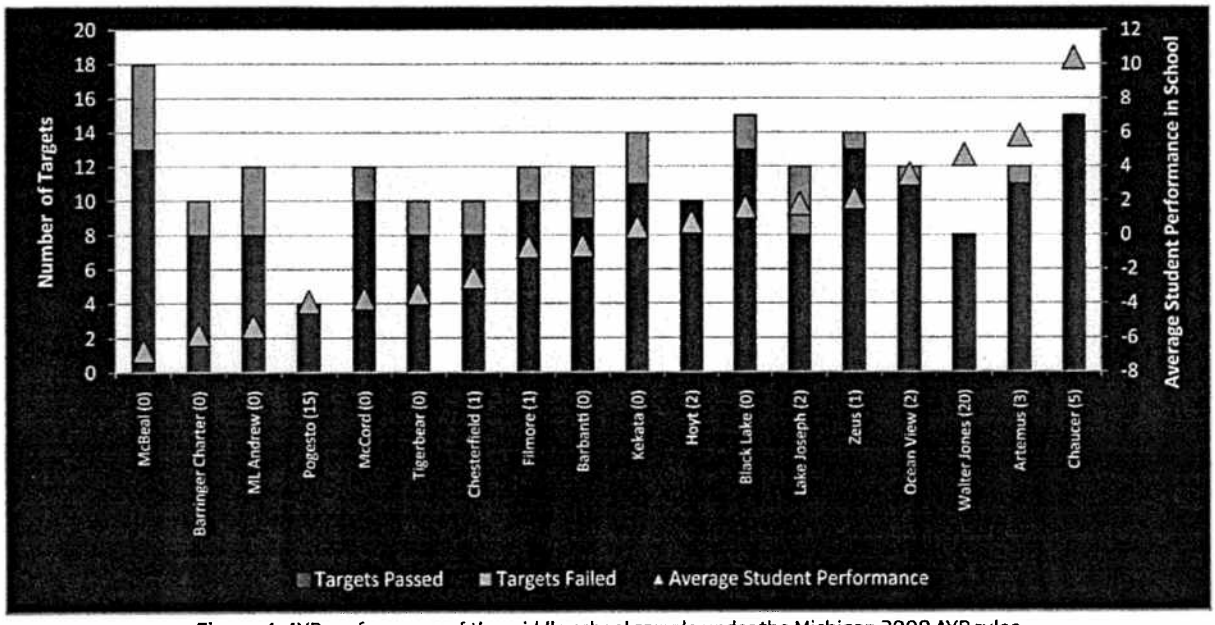

Figure 4. AYP performance of the middle school sample under the Michigan 2008 AYP rules

Note: This figure shows how each of the middle schools within the sample fared under Michigan'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. Artemus, for example, met 11 of its 12 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.

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, 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 Michigan 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:

- All elementary and middle schools met reading and math targets for their overall populations (again, most likely because of Michigan's relatively easy proficiency standards compared to other states).
- Six of the 8 failing elementary schools (Clarkson, JFK, Scholls, Hissmore, Wolf Creek, Alice Mayberry) and 6 of the 14 failing middle schools (Barringer, Tigerbear, Chesterfield, Filmore, Black Lake, and Artemus) missed AYP only for the SWD subgroup.
- Two middle schools (Zeus and Ocean View) fail only because of their LEP subgroups.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. We can see that elementary students did well on Michigan's math test and middle school students performed better in reading than math. This may be because Michigan's proficiency scores are easier in math than in reading at the elementary grades and easier in reading than in math at the middle grades (see Figure 2). Second, the performance of SWDs is proving challenging for schools under Michigan's system, particularly in middle schools, where this subgroup tends to have enough students to meet the state's minimum *n* size. Finally, we see that low-income and minority subgroups performed relatively well under Michigan's accountability system.

Table 2. Elementary school subgroup performance of sample schools under the 2008 Michigan 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	88.2%	74.1%	Y	Y	Y	N	Y	Y	Y	Y					Y	Y					10	9	90%	N	1
Maryweather	88.1%	74.4%	Y	Y	N	N	Y	N	Y	Y					Y	Y			Y	Y	12	9	75%	N	1
Few	90.4%	77.7%	Y	Y	Y	N	Y	N	Y	Y					Y	Y			Y	Y	12	10	83%	N	1
Nemo	91.6%	89.8%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	7
Island Grove	93.7%	87.2%	Y	Y				Y	Y	Y					Y	Y			Y	Y	9	9	100%	Y	4
JFK	96.3%	86.2%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	3
Scholls	96.6%	88.1%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	7
Hissmore	94.3%	90.1%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	7
Wolf Creek	92.7%	88.6%	Y	Y	Y	N		Y	Y	Y					Y	Y			Y	Y	11	10	91%	N	5
Alice Mayberry	97.2%	92.4%	Y	Y	Y	N			Y	Y	Y	Y							Y	Y	10	9	90%	N	9
Wayne Fine Arts	97.7%	97.7%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	21
Winchester	96.7%	94.3%	Y	Y	Y	Y									Y	Y			Y	Y	8	8	100%	Y	22
Coastal	94.5%	88.5%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	14	100%	Y	3
Paramount	92.9%	89.9%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	7
Forest Lake	98.9%	95.2%	Y	Y	Y	Y			Y	Y									Y	Y	8	8	100%	Y	8
Marigold	99.3%	96.0%	Y	Y	Y	Y			Y	Y									Y	Y	8	8	100%	Y	10
Roosevelt	99.7%	98.6%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	28
King Richard	97.6%	97.3%	Y	Y	Y	Y	Y	Y	Y	Y					Y	Y			Y	Y	12	12	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.

Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Michigan's NCLB accountability system is, in some 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 Michigan, too.

But Michigan is also home to a few anomalies. First, consider Island Grove Elementary (see Figure 3). It failed to make AYP in 24 of the 28 states in our sample, yet made AYP in Michigan. In examining Table 2, we can see that Island Grove didn't meet the minimum numbers for the SWD subgroup, which created difficulty for so many other schools within the sample. With fewer accountable subgroups, and with relatively easy proficiency standards (Figure 2), Island Grove made AYP, even when other schools with higher average performance didn't.

Second, look at Pogesto Middle School (see Figure 4). Even with its relatively low average performance, it made

Table 3. Middle school subgroup performance of sample schools under the 2008 Michigan 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	68.8%	73.9%	Y	Y	N	N	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	18	13	72%	N	0
Barringer Charter	83.3%	83.9%	Y	Y	N	N			Y	Y	Y	Y			Y	Y					10	8	80%	N	0
ML Andrew	70.6%	82.1%	Y	Y	N	N			N	Y	N	Y			Y	Y			Y	Y	12	8	67%	N	0
Pogesto	70.4%	85.2%	Y	Y															Y	Y	4	4	100%	Y	15
McCord Charter	73.0%	84.7%	Y	Y	N	Y			Y	Y	N	Y			Y	Y			Y	Y	12	10	83%	N	0
Tigerbear	77.8%	80.7%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	0
Chesterfield	82.8%	84.6%	Y	Y	N	N			Y	Y	Y	Y							Y	Y	10	8	80%	N	1
Filmore	82.5%	89.4%	Y	Y	N	N	Y	Y	Y	Y					Y	Y			Y	Y	12	10	83%	N	1
Barbanti	75.7%	82.9%	Y	Y	N	N	N	Y	Y	Y					Y	Y			Y	Y	12	9	75%	N	0
Kekata	84.3%	84.2%	Y	Y	N	Y	N	N	Y	Y	Y	Y			Y	Y			Y	Y	14	11	79%	N	0
Hoyt	87.0%	88.6%	Y	Y	Y	Y			Y	Y	Y	Y							Y	Y	10	10	100%	Y	2
Black Lake	87.7%	87.9%	Y	Y	N	N	Y		Y	Y	Y	Y	Y	Y	Y	Y			Y	Y	15	13	87%	N	0
Lake Joseph	85.2%	89.7%	Y	Y	N	N	N	N	Y	Y					Y	Y			Y	Y	12	8	67%	N	2
Zeus	88.4%	88.6%	Y	Y	Y	Y	N	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	13	93%	N	1
Ocean View	89.6%	93.7%	Y	Y	Y	Y	N	Y	Y	Y					Y	Y			Y	Y	12	11	92%	N	2
Walter Jones	93.0%	92.6%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	20
Artemus	91.5%	90.7%	Y	Y	Y	N			Y	Y			Y	Y	Y	Y			Y	Y	12	11	92%	N	3
Chaucer	93.4%	95.9%	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y			Y	Y	15	15	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.

AYP in Michigan, but failed to do so in 13 of 28 states. Like Island Grove, its AYP success in Michigan is likely attributable to the relatively small number of targets (four) it has to meet (shown in Table 3), along with Michigan's relatively easy proficiency standards, compared to other states.

This is consistent with the patterns shown in Table 6, which compares schools that did and didn't made 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 academically disadvantaged (e.g., low-income) 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 Michigan's AYP rules (and AMOs) for 2008. Among this sample, 10 elementary schools and 4 middle schools—

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

Table 5. Summary of subgroup performance of sample middle schools under the 2008 Michigan 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	11	10
Students with limited English proficiency	9	6	3
Low-income students	17	1	0
African-American students	11	3	0
Asian/Pacific Islander students	4	0	0
Hispanic students	14	0	0
American Indian/Alaska Native students	1	0	0
White students	17	0	0

14 out of a sample of 36—would have made AYP in Michigan. Looking across the 28 state accountability systems examined in the study, this puts Michigan at the high end of the sample distribution in terms of the number of schools making AYP (see Figure 1). In addition, several sample schools made AYP in Michigan that failed to make AYP in most other states, most likely because Michigan's proficiency standards are relatively easy compared to other states and its schools generally have fewer accountable subgroups.

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, Michigan's NCLB accountability system is working exactly as Congress intended: identify-

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	10	8	4	14
Average student body size	260	361	586	937
Average % low income	28	69	37	47
Average % nonwhite	29	56	30	48
Average performance†	4.28	-2.59	2.99	-0.93
Average % growth‡	124	104	118	92
Average number of targets to meet	9	11	9	13

† 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.

ing 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. Each of the sample schools made AYP in Michigan for its student populations as a whole. 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 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] doesn't the failure of many of these students to meet Michigan'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 students, but when sizable numbers of schools (particularly at the middle school level) are unable to meet the goal, perhaps that indicates that the goal is unrealistic. These will be critical considerations for Congress as it takes up NCLB re-authorization 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

[9] See footnote 4.

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.