



## 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 towards these goals.

This report examines Wisconsin'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 Wisconsin's system is compared with other states. For this study, we selected 36 schools from around the nation, schools that vary by size, achievement, and diversity, among other factors, and determined whether each would make AYP under Wisconsin's system as well as under the systems of 27 other states. We used school data and proficiency cut score<sup>1</sup> estimates from academic year 2005–2006, but applied them against Wisconsin's AYP rules.

Here are some key findings:

- We estimate that **just 1 of 18 elementary schools and 11 of 18 middle schools** in our sample **failed to make AYP** in 2008 under Wisconsin's accountability system.
- **Looking across the 28 state accountability systems examined in the study, we find that Wisconsin has the greatest number of elementary schools making AYP in our sample.** In addition, at seven, Wisconsin

has the second highest number of middle schools making AYP in the sample (only Arizona has more) (See Figure 1).

- **The high number of schools making AYP in Wisconsin is likely due to the fact that Wisconsin's proficiency standards are extremely easy compared to other states, plus it uses a proficiency index, which means it gives “partial credit” to students performing below proficient.**
- The few schools in our sample that fail to make AYP in Wisconsin are meeting expected targets for their overall populations<sup>2</sup> but failing because of the performance of individual subgroups, particularly students

More schools make AYP in 2008 under **Wisconsin's** accountability system than in any other state in our sample. This is likely due to the fact that Wisconsin's proficiency standards (or cut scores) are relatively easy compared to other states (all of them are below the 30th percentile). Second, Wisconsin's minimum subgroup size for students with disabilities is 50, which is a bit larger than most other states (the size for their other subgroups is comparable to other states'). This means that Wisconsin schools must have more students with disabilities in order for that group to be held separately accountable. Third, Wisconsin's 99 percent confidence interval provides schools with greater leniency than the more commonly used 95 percent confidence interval. Last, unlike most states, Wisconsin measures its student performance with a proficiency index, which gives partial credit for students achieving “partial proficiency.” All of these factors work together so that 17 out of 18 elementary schools make AYP in Wisconsin, more than any other state in the study.

<sup>1</sup> A cut score is the minimum score on a student must receive on NWEA's Measures of Academic Progress (MAP) that is equivalent to performing proficient on the Wisconsin Knowledge and Concepts Examinations - Criterion Referenced Test (WKCE-CRT).

<sup>2</sup> 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 are simply not treated as their own subgroup.

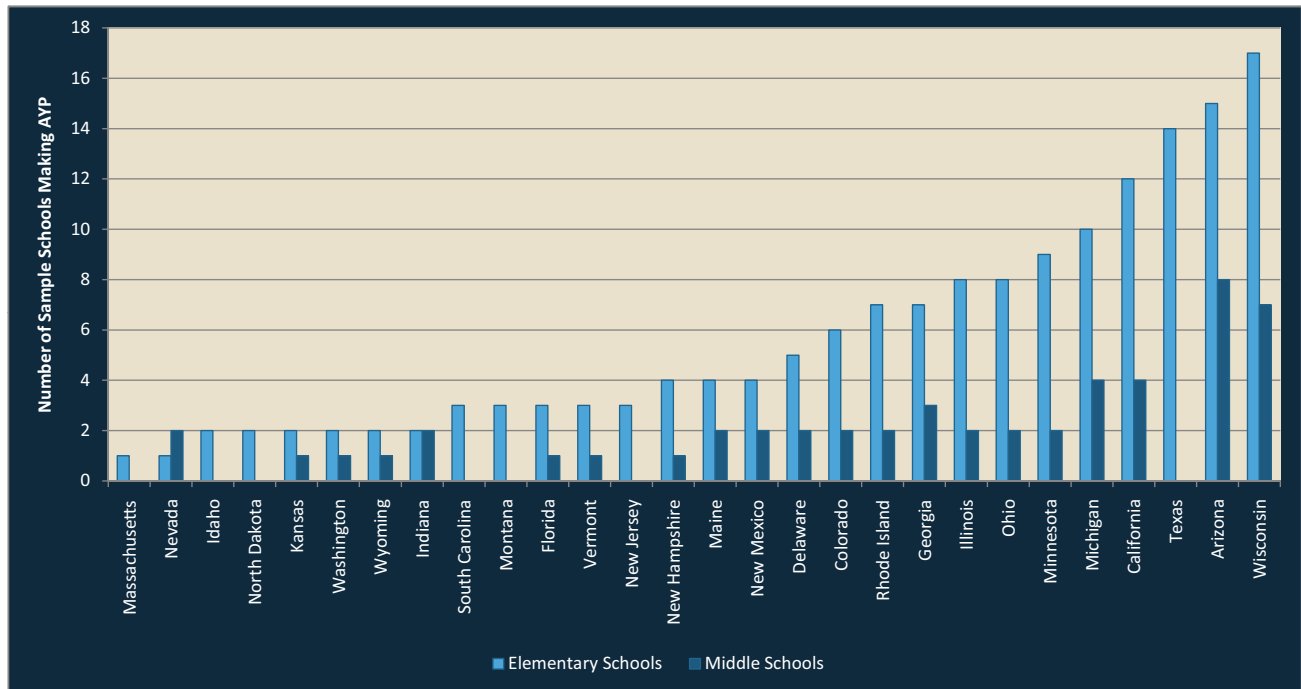


Figure 1. Number of sample schools making AYP by state

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.

with disabilities (SWDs)<sup>3</sup> and students with limited English proficiency (LEP).<sup>4</sup>

- Two sample schools that failed to make AYP in any other state made AYP in Wisconsin. Again, this is likely due to the fact that Wisconsin’s proficiency standards are relatively easy compared to other states, along with the fact that these two schools had fewer accountable subgroups.
- In Wisconsin, as is the case in most states, schools with fewer subgroups attain AYP more easily than schools with more subgroups, even when their average student performance is much lower. In other words, schools with greater diversity and size face greater challenges in making AYP.

- As in other states, middle schools have greater difficulty reaching AYP in Wisconsin than do elementary schools, primarily because their student populations are larger and therefore have more qualifying subgroups—not because their student achievement is any lower than in the elementary schools.
- A strong predictor of a school making AYP under Wisconsin’s system is whether it has enough SWDs to qualify as a separate subgroup. Almost all schools with qualifying subgroups in this category failed to meet their AYP targets, particularly at the middle school level.<sup>5</sup> **Ironically, Wisconsin has one of the largest minimum  $n$  sizes for SWDs in our sample; still, when enough SWDs exist to comprise a subgroup, they do not perform well.**

<sup>3</sup> SWDs are defined as those students following individualized education plans.

<sup>4</sup> Note that we use “LEP students” and “English language learners” interchangeably to refer to students in the same subgroup.

<sup>5</sup> It should be noted that our subgroup findings for Limited English Proficient (LEP) and students with disabilities (SWDs) may be slightly more negative than would be seen under real world conditions. This is mostly due to the differences in testing practices between how LEP students and SWDs are treated in the Wisconsin Knowledge and Concepts Examinations - Criterion Referenced Test (WKCE-CRT) 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 the current study, however, no valid MAP scores were omitted from consideration.

## Introduction

*The Proficiency Illusion* (Cronin et al. 2007a) linked student performance on Wisconsin'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) of 2001. That study revealed profound differences in states' proficiency standards (i.e., how difficult it is to achieve proficiency on the state test), and even across grades within a single state.

Our study expands on *The Proficiency Illusion* by examining other key factors of state NCLB accountability plans and how they interact with state proficiency standards to determine whether the schools in our sample made adequate yearly progress (AYP) in 2008. Specifically, we estimated how a single set of schools, drawn from around the country, would fare under the differing rules for determining AYP in 28 states (the original 25 in *The Proficiency Illusion* plus 3 others for which we now have cut score estimates). In other words, if we could somehow move these entire schools—with their same mix of characteristics—from state to state, how would they fare in terms of making AYP? Will schools with high-performing students consistently make AYP? Will schools with low-performing students consistently fail to make AYP? If AYP determinations for schools are not consistent across states, what leads to the inconsistencies?

NCLB requires every state, as a condition of receiving Title I funding, to implement an accountability system that aims to get 100% of its students to the proficient level on the state test by academic year 2013–2014. In the intervening years, states set annual measurable objectives (AMOs). This is the percentage of students in each school, and in each subgroup within the school (such as low income<sup>6</sup> 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 10 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 ten 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.

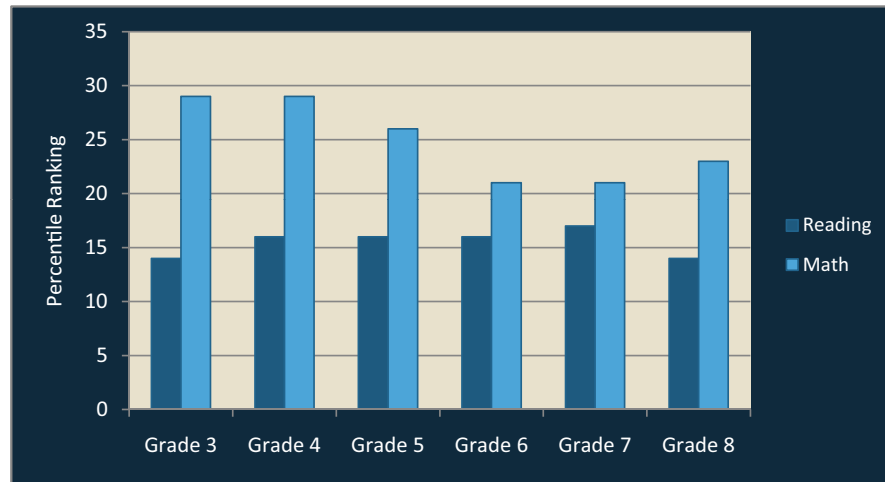
All of these AYP rules vary by state, which means that a school that makes AYP in Arizona or Ohio, for example, might not make it under South Carolina's or Idaho's rules (U.S. Department of Education 2008).

## 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, such as students with limited English proficiency (LEP), 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-

<sup>6</sup> Low-income students are those who receive a free or reduced-price lunch.



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

Note: This figure illustrates the difficulty of Wisconsin’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 Wisconsin’s math cut scores are below the 30th percentile and all its reading cut scores are below the 18th percentile.

price lunches). Appendix 1 contains a complete discussion of the methodology for this project along with the characteristics of the school sample.<sup>7</sup>

Proficiency cut score estimates for the Wisconsin Knowledge and Concepts Examinations - Criterion Referenced Test (WKCE-CRT) are taken from *The Proficiency Illusion* (as shown in Figure 2), which found that Wisconsin’s definitions of proficiency were generally below average compared with the standards set by the other 25 states in that study (especially in reading). These cut scores were used to estimate whether students would have scored as proficient or better on the Wisconsin test, given their performance on MAP. Student test data and subgroup designations are then used to determine how these 18 elementary and 18 middle schools would have fared under Wisconsin 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 Wisconsin’s 2008 AYP rules.

Table 1 shows the pertinent Wisconsin AYP rules that were applied to elementary and middle schools in this

study. Wisconsin’s minimum subgroup size for most subgroups is 40, which is comparable to most other states we examined.<sup>8</sup> However, for students with disabilities (SWDs) the minimum is 50, which is a bit larger than most other states.

Furthermore, although most states examined in the study apply confidence intervals (or margins of statistical error) to their measurements of student proficiency rates, Wisconsin’s 99% confidence interval gives schools greater leniency than the more commonly used 95% confidence interval used by most other states. So, for instance, although schools are supposed to get 74% of their grade 3–8 students (as well as 74% of their grade 3–8 students in each subgroup) to the proficient level on the state reading test, applying the confidence interval means that the real target can actually be lower, particularly with smaller groups.<sup>9</sup>

Unlike most states, Wisconsin measures its student performance with a proficiency index, which gives partial credit for students achieving “partial proficiency.” In the short term, the index makes it easier for Wisconsin

<sup>7</sup> We gave all schools in our sample pseudonyms in this report.

<sup>8</sup> Keep in mind, however, that school size and *n* size are related (e.g., small *n* sizes make sense for small schools).

<sup>9</sup> 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.** Wisconsin AYP rules for 2008

Subgroup minimum <i>n</i>	Race/ethnicity: 40	
	SWDs: 50	
	Low-income students: 40	
	LEP students: 40	
CI	Applied to proficiency rate calculations?	
	Yes; 99% CI used	
AMOs	Baseline proficiency levels as of 2002 (index)	2008 targets (index)
READING/LANGUAGE ARTS		
Grade 3	61	74
Grade 4	61	74
Grade 5	61	74
Grade 6	61	74
Grade 7	61	74
Grade 8	61	74
MATH		
Grade 3	37	58
Grade 4	37	58
Grade 5	37	58
Grade 6	37	58
Grade 7	37	58
Grade 8	37	58

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

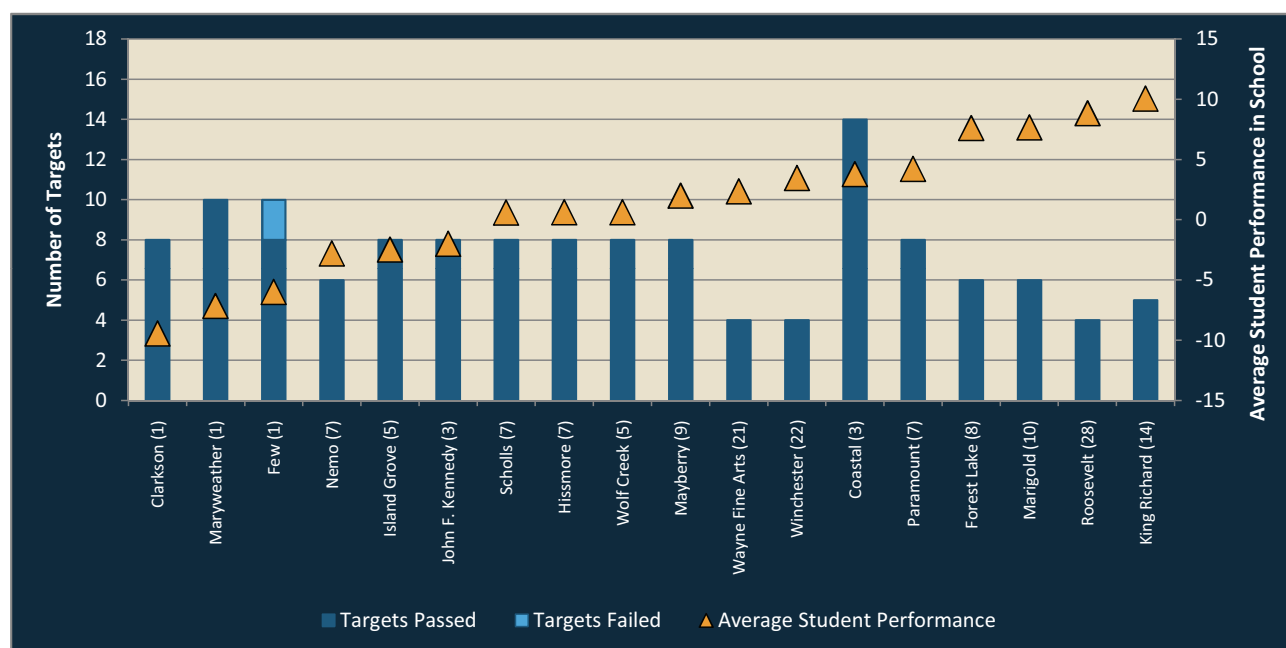
schools to meet their targets, although the effect of the index diminishes as the targets approach 100% proficiency requirement dictated under NCLB for 2014.<sup>10</sup>

**Note that we were unable to examine the effect 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 is 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

<sup>10</sup> In six of the states studied (Massachusetts, Minnesota, Rhode Island, Vermont and New Hampshire, as well as Wisconsin), an index is used that gives full credit to students who achieve proficient (or better) and partial credit to students performing at lower levels. Consequently, the resultant score in states using this “hybrid” model is always higher than the actual proficiency percentage (giving students partial credit for achieving lower proficiency levels is obviously better than no credit, at least for the schools’ ratings). The index provides a fair amount of help when annual targets are below 50%; however, once targets rise above 75%, the index has far less impact.



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

Note: This figure indicates how each elementary school within the sample fared under Wisconsin's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet 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. Few Elementary, for example, met eight of its ten 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), which 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 AYP in the study.

schools. In addition, federal law requires 95% of each school's students—and 95% of students in each school's subgroups—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 annual AMOs. Again, Appendix 1 supplies further methodological detail.

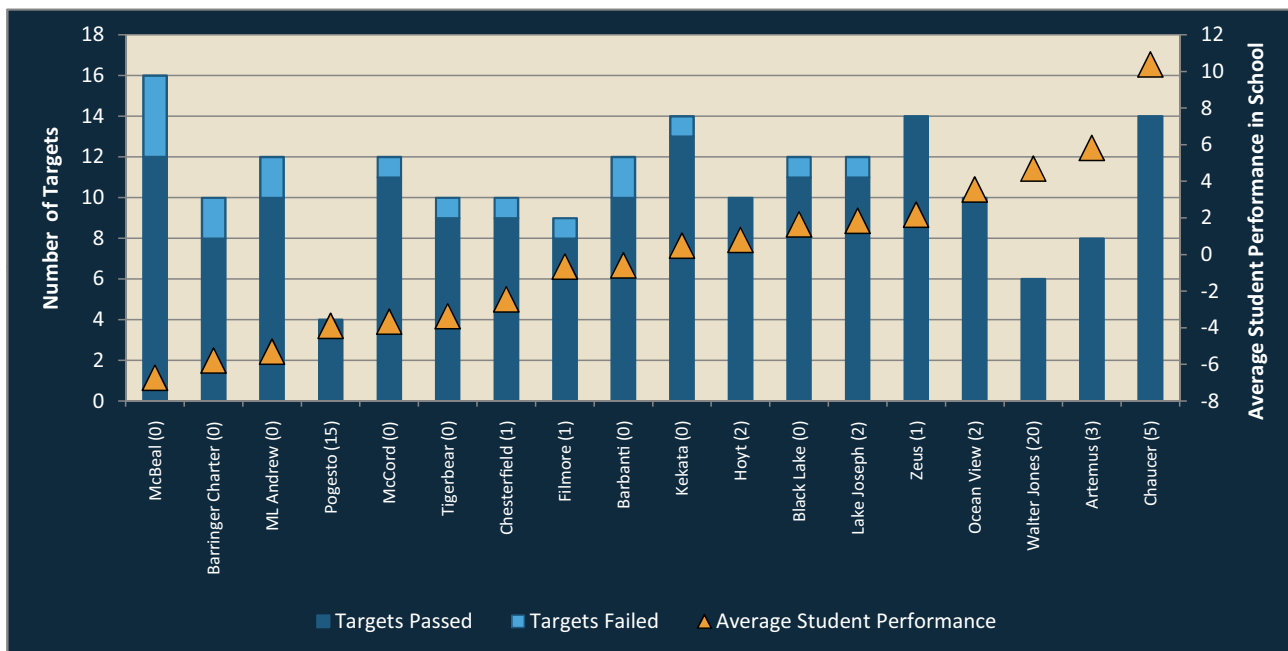
### How Did the Sample Schools Fare under Wisconsin's AYP Rules?

Figure 3 illustrates the AYP performance of the sample elementary schools under Wisconsin's 2008 AYP rules.

**Seventeen elementary schools made AYP, while only one (Few Elementary) 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.

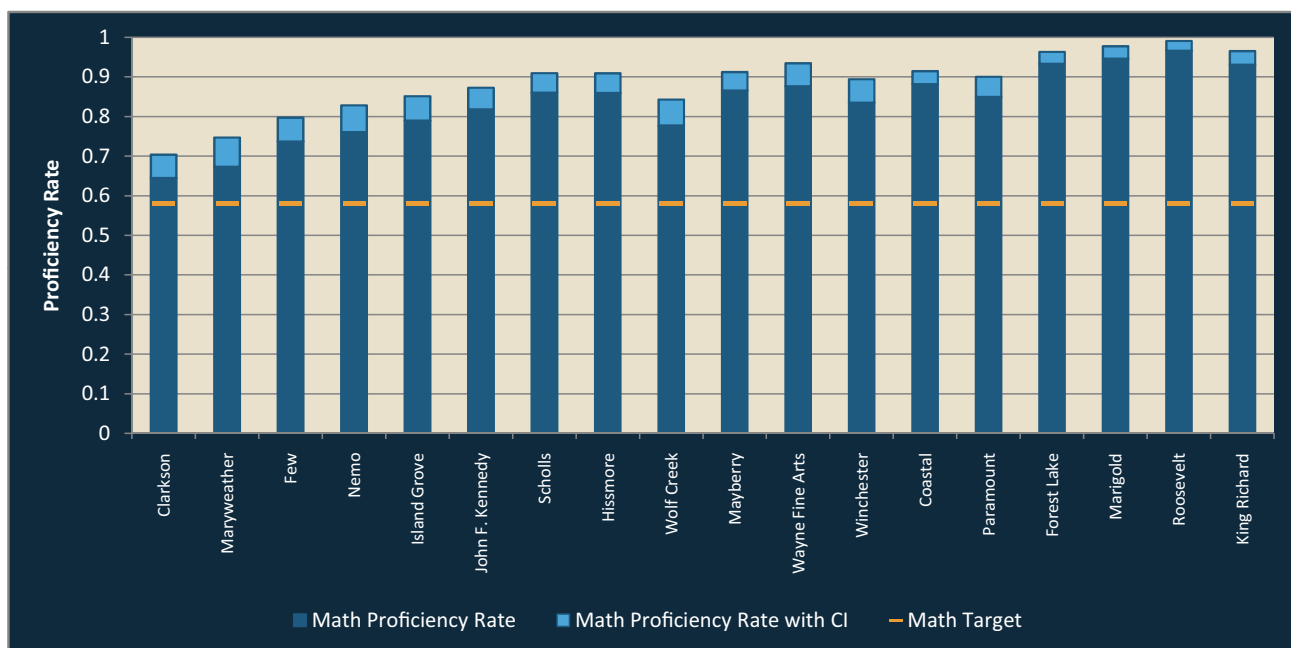
Figure 4 illustrates the AYP performance of the sample middle schools under the 2008 Wisconsin AYP rules. **Out of 18 middle schools in our sample, 7 made AYP**—one low-performance school (Pogesto), which has relatively few qualifying subgroups and six higher performing schools (Hoyt, Zeus, Ocean View, Walter Jones, Artemus, and Chaucer).

Figures 5 and 6 indicate the degree to which schools' math proficiency rates are aided by the confidence interval for elementary and middle schools, respectively. On these figures, the darker portions of the bars show the actual proficiency rates at each school, and the lighter



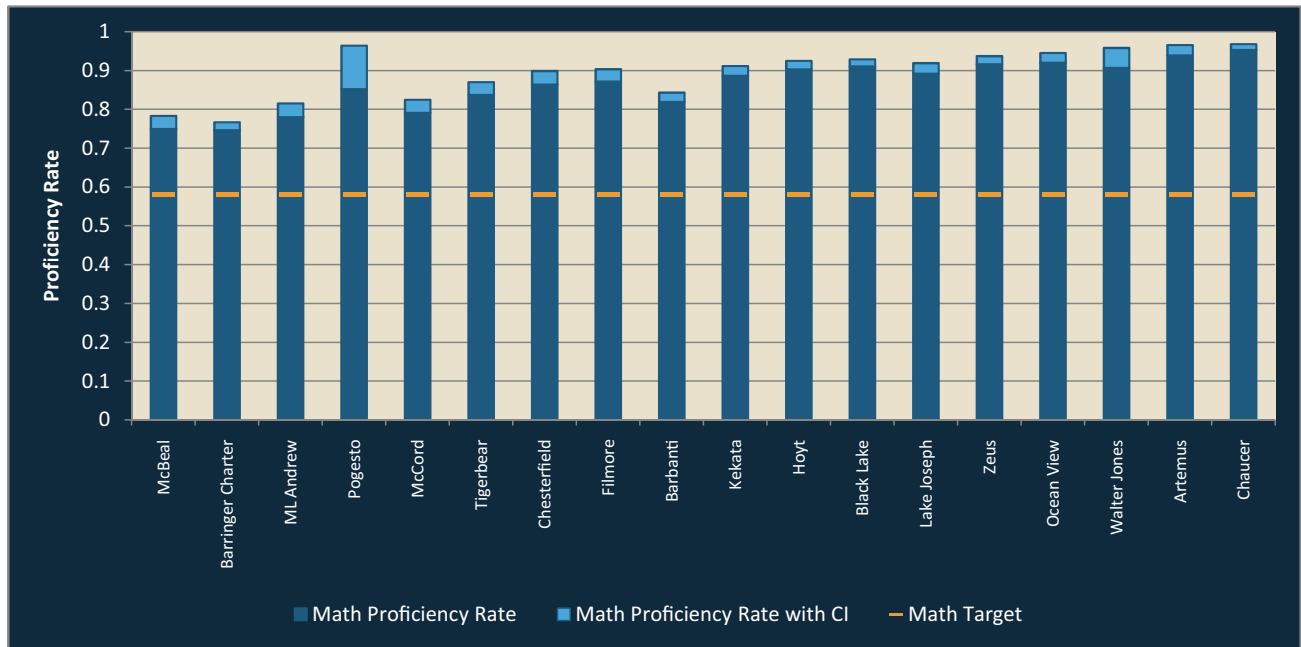
**Figure 4.** AYP performance of the middle school sample under the Wisconsin 2008 AYP rules

Note: This figure indicates how each middle school within the sample would have fared under Wisconsin's AYP rules (as described in Table 1). The bars show the number of targets that each school has to meet 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 does not make AYP, so any light blue means that the school failed. Black Lake, 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), which is measured by 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 under the Wisconsin 2008 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 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.



**Figure 6.** Impact of the confidence interval on middle school math proficiency rates under the Wisconsin 2008 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 none of the sample middle 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.

portions of the bars show the degree to which these proficiency rates were increased by applying the confidence interval. The orange lines show the annual target needed to meet AYP. These figures show that **none of the sample elementary or middle schools was assisted by the confidence intervals, because the math targets in Wisconsin are so low, relative to the schools' overall performance.** The picture is much the same for reading proficiency rates at the elementary and middle school levels (not shown). No school is assisted by the confidence interval because the reading targets are so low. **In short, applying the confidence interval, even though it is a lenient one, has no effect on whether or not sample schools meet their overall reading and math targets.**<sup>11</sup>

### Where do schools fail?

Figures 3 and 4 illustrate how many subgroup targets each sample school is held accountable, and whether or not each school made AYP. However, these figures do

not indicate which subgroups failed or passed in which school. Tables 2 and 3 list information on individual subgroup performance 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 (AA), Asian/Pacific Islander (Asian), Hispanic/Latino (Hispanic), American Indian/Alaska Native (AI/AN), and white. Tables 2 and 3 also show whether a school met AYP under the 2008 Wisconsin rules, and the total number of states within the study in which that school met AYP.

<sup>11</sup> In the current analyses, confidence intervals were applied to both the overall school population and to all eligible subgroups in our sample schools. Thus, the ultimate impact of the confidence interval may be larger than the impact depicted in Figures 5 and 6. However, we chose not to show how the confidence interval impacted subgroup performance because it would have added greatly to the report's length and complexity.



**Table 2.** Elementary subgroup performance of sample schools under the 2008 Wisconsin 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	64.5%	79.6%	Y	Y			Y	Y	Y	Y					Y	Y					8	8	100%	Y	1	
Maryweather	67.4%	79.0%	Y	Y			Y	Y	Y	Y					Y	Y			Y	Y	10	10	100%	Y	1	
Few	73.7%	80.6%	Y	Y	Y	N	Y	N	Y	Y					Y	Y					10	8	80%	N	1	
Nemo	76.0%	91.2%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	7	
Island Grove	79.0%	89.3%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	4	
JFK	81.8%	88.6%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	3	
Scholls	86.0%	91.4%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	7	
Hissmore	85.9%	92.2%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	7	
Wolf Creek	77.8%	90.0%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	5	
Alice Mayberry	86.6%	93.4%	Y	Y					Y	Y	Y	Y							Y	Y	8	8	100%	Y	9	
Wayne Fine Arts	87.6%	97.7%	Y	Y															Y	Y	4	4	100%	Y	21	
Winchester	83.5%	95.0%	Y	Y															Y	Y	4	4	100%	Y	22	
Coastal	88.1%	90.8%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y			Y	Y	14	14	100%	Y	3	
Paramount	84.9%	90.7%	Y	Y					Y	Y					Y	Y			Y	Y	8	8	100%	Y	7	
Forest Lake	93.3%	96.6%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	8	
Marigold	94.6%	96.2%	Y	Y					Y	Y									Y	Y	6	6	100%	Y	10	
Roosevelt	96.6%	99.2%	Y	Y															Y	Y	4	4	100%	Y	28	
King Richard	93.1%	97.3%	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.

The school-by-school findings in Tables 2 and 3 show that:

- All schools met both their math and reading targets for their overall student populations.
- Nine of the 11 failing middle schools only missed targets for the students with disabilities subgroup.
- One middle school (Kekata) failed to make AYP only because of its LEP subgroup.
- Unlike any other state in the study, *all* of the low-income, African American, Hispanic, Asian, American

Indian, and white subgroups met both their reading and math targets.

Tables 4 and 5 summarize subgroup performance for elementary and middle schools, respectively. First, there are very few qualifying SWD and LEP subgroups at the elementary level (keep in mind that the minimum *n* size for SWDs is rather large at 50). But when there are large enough numbers of these students to comprise subgroups at the middle school level, they tend to struggle. In fact, one of the two elementary schools and most of the middle schools in the study that have qualifying SWD subgroups fail to make AYP. Students with LEP

**Table 3.** Middle school subgroup performance of sample schools under the 2008 Wisconsin 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	74.9%	80.2%	Y	Y	N	N	N	N	Y	Y	Y	Y			Y	Y	Y	Y	Y	Y	16	12	75%	N	0	
Barringer Charter	74.7%	88.5%	Y	Y	N	N			Y	Y	Y	Y			Y	Y					10	8	80%	N	0	
ML Andrew	78.0%	89.1%	Y	Y	N	N			Y	Y	Y	Y			Y	Y				Y	Y	12	10	83%	N	0
Pogesto	85.2%	92.6%	Y	Y																Y	Y	4	4	100%	Y	15
McCord Charter	79.1%	89.5%	Y	Y	N	Y			Y	Y	Y	Y			Y	Y				Y	Y	12	11	92%	N	0
Tigerbear	83.7%	86.4%	Y	Y	Y	N			Y	Y	Y	Y								Y	Y	10	9	90%	N	0
Chesterfield	86.4%	89.9%	Y	Y	Y	N			Y	Y	Y	Y								Y	Y	10	9	90%	N	1
Filmore	87.2%	92.3%	Y	Y		N			Y	Y					Y	Y				Y	Y	9	8	89%	N	1
Barbanti	82.0%	87.7%	Y	Y	N	N	Y	Y	Y	Y					Y	Y				Y	Y	12	10	83%	N	0
Kekata	88.7%	90.1%	Y	Y	Y	Y	Y	N	Y	Y	Y	Y			Y	Y				Y	Y	14	13	93%	N	0
Hoyt	90.3%	91.5%	Y	Y	Y	Y			Y	Y	Y	Y								Y	Y	10	10	100%	Y	2
Black Lake	91.1%	91.6%	Y	Y	Y	N			Y	Y	Y	Y			Y	Y				Y	Y	12	11	92%	N	0
Lake Joseph	89.2%	92.8%	Y	Y	Y	N	Y	Y	Y	Y					Y	Y				Y	Y	12	11	92%	N	2
Zeus	91.7%	91.9%	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y			Y	Y				Y	Y	14	14	100%	Y	1
Ocean View	92.0%	96.0%	Y	Y			Y	Y	Y	Y					Y	Y				Y	Y	10	10	100%	Y	2
Walter Jones	90.7%	94.3%	Y	Y					Y	Y										Y	Y	6	6	100%	Y	20
Artemus	93.9%	93.9%	Y	Y					Y	Y					Y	Y				Y	Y	8	8	100%	Y	3
Chaucer	95.3%	97.2%	Y	Y	Y	Y	Y	Y	Y	Y				Y	Y	Y	Y			Y	Y	14	14	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.

are also struggling somewhat to meet the state's targets; three schools with a large enough LEP population to qualify as a separate subgroup fail to meet targets for these students.

### Characteristics of Schools that Did and Didn't Make AYP

A close look at Figures 3 and 4 indicates that Wisconsin's NCLB accountability system is, at least somewhat, behaving like those in other states. For example, Wayne Fine Arts and Walter Jones made AYP in many of the

states—21 and 20, respectively. And these schools made AYP in Wisconsin, too.

But Wisconsin is also home to a few anomalies. First, consider Clarkson and Maryweather elementary schools (see Table 2). They each failed to make AYP in 27 of the 28 states in our sample, yet made AYP in Wisconsin. In examining Table 2, we can see that Clarkson and Maryweather didn't meet the minimum numbers for the SWD subgroup, which create difficulty for so many other schools in the study. Without fewer accountable subgroups and easy proficiency standards (see Figure 2),

**Table 4.** Summary of subgroup performance of sample elementary schools under the 2008 Wisconsin AYP rules

SUBGROUP	Number of schools with qualifying subgroups	Number of schools where subgroup failed to meet math target	Number of schools where subgroup failed to meet reading target
Students with disabilities	2	0	1
Students with limited English proficiency	4	0	1
Low-income students	15	0	0
African-American students	5	0	0
Asian/Pacific Islander students	0	0	0
Hispanic students	7	0	0
American Indian/Alaska Native students	0	0	0
White students	16	0	0

**Table 5.** Summary of subgroup performance of sample middle schools under the 2008 Wisconsin 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	14	5	9
Students with limited English proficiency	7	1	2
Low-income students	17	0	0
African-American students	10	0	0
Asian/Pacific Islander students	1	0	0
Hispanic students	13	0	0
American Indian/Alaska Native students	1	0	0
White students	17	0	0

these schools made AYP even when schools with higher average performance failed.

Second, look at Pogesto Middle School (Figure 4). Even with its relatively low average performance it made AYP in Wisconsin, but failed to do so in 13 of 28 states. Like Clarkson and Maryweather, its AYP success in Wisconsin is most likely attributable to its relatively small number of targets (four, as shown in Table 3) along with the easy pro-

ficiency standards in Wisconsin compared to 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, lower percentages of low income stu-

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

	Elementary Schools		Middle Schools	
	Made AYP	Failed to make AYP	Made AYP	Failed to make AYP
Number of schools in sample	17	1	7	11
Average student body size	290	550	663	984
Average % low income	44	90	33	53
Average % nonwhite	38	89	26	55
Average performance†	1.65	-5.99	3.36	-2.23
Average % growth‡	114	135	105	94
Average number of targets to meet	7	10	9	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.

dents, and fewer subgroups (and thus fewer 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 fare under Wisconsin's AYP rules and AMOs for 2008. We found that 17 elementary schools and 7 middle schools—24 in all, from a total of 36—would have made AYP in Wisconsin. Looking across the 28 state accountability systems examined in the study, we see that Wisconsin has the greatest number of elementary schools making AYP in our sample. In addition, at 7, Wisconsin has the second highest number of middle schools making AYP in the sample (only Arizona has more). This is likely due to the easy proficiency standards in Wisconsin (the state's reading cut scores are below the 18th percentile), as well as its proficiency "index" which awards partial credit to students performing below proficient.

Because the overriding goal of the federal 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, Wisconsin'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. All of the sample schools met the Wisconsin math and reading targets for their student populations as a whole. In the pre-NCLB era, such schools might have been considered to be effective or at least not in need of improvement, even though sizable numbers of their pupils weren't meeting state standards. Disaggregating data by race, income, and so on has made those students visible. That is surely a positive step.

Yet NCLB's design flaws are also readily apparent. Does it make sense that the size of a school's enrollment has so much influence over making AYP? Does it make sense that having fewer subgroups enhances the likelihood of making AYP? In the case of Wisconsin, is it "fair" that students receive partial credit for performing below proficient? Or that many subgroups meet their targets not because of improved performance but largely due to low cut scores? 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 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.