The Effects of State Decisions About NCLB Adequate Yearly Progress Targets

Andrew C. Porter, *Vanderbilt University*
Robert L. Linn, *University of Colorado at Boulder*, and
C. Scott Trimble, *Kentucky Department of Education (Retired)*

The No Child Left Behind Act allows states to vary (a) the trajectories they select to move from the baseline percent proficient or above in 2002 to the 100% proficient goal in 2014, (b) the minimum number of students required for reporting of disaggregated subgroup results, and (c) whether or not they will use confidence intervals when determining whether or not an annual measurable objective has been met. We use data from Kentucky for the years 2003 and 2004 to explore the consequences of different design decisions. The effect of design decisions on number and percentage of schools meeting adequate yearly progress (AYP) is large, with important implications for education practice.

Keywords: No Child Left Behind, accountability, state policy

It has been a little over 3 years since President Bush signed the No Child Left Behind Act (NCLB, 2001) into law in January 2002. Although some of the accountability provisions of the law have been the subject of considerable criticism, the law has also been widely praised for its focus on closing the achievement gap by improving achievement of students who have too often been ignored in the past (Chubb, Linn, Haycock, & Wiener, 2005).

Accountability is a prominent feature of NCLB. The NCLB accountability system is intended to contribute to improving education for all children by identifying schools that are in need of improvement so that corrective action can be taken. The system requires that all students achieve at the proficient level or higher (as defined by states) in mathematics and reading/English language arts by 2014. States must establish annual measurable objectives for mathematics and reading/English language arts in a way that leads to the goal of 100% proficiency by 2014.

Both schools and districts are to be held accountable and under similar rules. Here, we address school accountability, though the argument and general findings apply straightforwardly to districts as well. In order for a school to meet adequate yearly progress (AYP) requirements in a given year, students in the school must achieve at or above the state-established annual measurable objective that year in both mathematics and reading/English language arts. In keeping with the NCLB emphasis on closing the achievement gap, the academic achievement goals must be met not just for the total group of students considered as a whole but for each of the following specific subgroups: students who are economically disadvantaged, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency.

Title I schools in a state that fail to meet their AYP target for 2 consecutive years are identified as in need of improvement. They must develop a school improvement plan, spend 10% of their Title I funds on professional development for teachers, and offer school choice to all students in the building. Schools failing AYP for 3 years in a row must in addition provide supplemental services (e.g., tutoring) to students identified as being from low-income families. Providers of supplemental services must be state approved. Title I schools that fail to meet their AYP target for 4 consecutive years are identified for corrective action. While there is some latitude in the actions these schools must take, they must select and implement at least one of six corrective actions including, for example, replacing ineffective school staff or implementing a new curriculum. Schools that fail AYP for 5 years in a row must develop a school governance restructuring plan, and schools that fail for 6 years in a row must implement that plan (Cowan, 2003).

Andrew C. Porter is the Patricia and Rodes Hart Professor of Educational Leadership and Policy, Vanderbilt University, Learning Sciences Institute, Peabody #59, Nashville, TN 37203; Andy.porter@vanderbilt.edu. His areas of specialization are psychometrics, student assessment, and policy and research on teaching.

Robert L. Linn is Distinguished Professor Emeritus, University of Colorado at Boulder, P.O. Box 1815, Ouray, CO 81427; Robert.linn@colorado.edu. His areas of specialization are educational assessment and accountability.

C. Scott Trimble is an educational consultant (recently retired from the Kentucky Department of Education), 274 Highland Parkway, Frankfort, KY 40601; strimble@fewpb.net. His areas of specialization are large-scale assessment implementation/administration, inclusion of students with disabilities in large-scale assessments, and standards setting.

Although many of the AYP requirements are determined for states by NCLB and associated regulations and interpretations by the U.S. Department of Education, states have latitude in defining a number of critical features that determine whether or not a school meets AYP requirements. For example, states determine the proficient academic achievement standards for mathematics and reading/English language arts, and there is considerable state-to-state variation in the stringency of state definitions of proficient achievement (Linn, 2003; Linn, Baker, & Betebenner, 2002).

States must establish a starting point for measuring the percentage of students who meet or exceed the state's proficient level of academic achievement. Separate starting points are determined for reading/language arts and mathematics. The state must choose, at a minimum, the higher number from two separate calculations (or the state can set the starting point higher than either of those two numbers if it wishes). One number is simply the 2001/2002 percentage of students proficient for the lowest-achieving subgroup of students. The other is the percentage of students proficient in a school selected by rank ordering schools according to percentage of students proficient and then working from the lowest-ranked school up, accumulating 20% of the students in the state. The starting points for Iowa are 64% for math and 65% for English language arts, while the starting points for Missouri are 8.3% for math and 18.4% for English language arts (http://accountability.ccsso.org/index.asp).

In addition to defining proficient academic achievement, states have been allowed considerable latitude with regard to three other design decisions that may have consequences for determining whether or not a school meets or fails to meet AYP requirements. First, states are allowed to vary in the trajectories they select to move from the baseline percent proficient or above in 2002 to the 100% proficient goal in 2014. Second, states are allowed to determine the minimum number of students required for reporting of disaggregated subgroup results. Third, states can choose whether or not they will use confidence intervals when determining whether or not an annual measurable objective has been met, as well as what confidence level to use and whether to use a one- or two-tail interval. The purpose of this study is to explore the consequences of the latter three design decisions for schools meeting or failing to meet AYP requirements. We use data from Kentucky for the years 2003 and 2004 to explore the consequences of different design decisions that a state may make. Before turning to the Kentucky results, however, we first elaborate on the three design decisions and present information about the prevalence of state choice with regard to these decisions. Chester (this issue) considers a number of additional design decisions within the context of Ohio implementation of NCLB.

State Uses of Design Decisions

Annual Measurable Target Trajectories

Shortly after NCLB was enacted, the U.S. Department of Education provided an illustration of how a state might define annual measurable objectives that would move from the 2002 starting point percent proficient to 100% proficient in 2014. Not surprisingly, the illustration used a straight-line trajectory. Thus, a state that had a 2002 starting point of, say, 40% proficient would increase the annual measurable objective by 5% each year to achieve the 100% goal 12 years later. The law, however, does not require annual increases in measurable objectives. Rather, it requires an increase after 2 years and increases at least every third year after that, and each increase must be of equal size (Public Law 107–110). Thus, states were allowed to set trajectories for their annual measurable objectives following a stair-step pattern, with the first increase in 2005 and subsequent increases every third year. States could use the straight-line approach modeled by the U.S. Department of Education or a stair-step approach with constant increases in the required years (which we refer to as straight with plateaus)—or they could use a front-loaded trajectory, with larger increases for the early steps between plateaus, or a back-loaded trajectory, with larger increases between plateaus in the out years closer to 2014.

Using the U.S. Department of Education's website (2005), we found that the trajectories were unspecified for 3 states. Of the remaining 47 states, only 4 chose to use the straight-line approach, no state chose the front-loaded approach, 19 states elected to use the straight approach with plateaus, and the remaining 24 states chose the back-loaded approach (Table 1). The straight approach with plateaus is illustrated in Figure 1 for Florida and the back-loaded approach is illustrated in Figure 2 for Michigan. Neither Florida nor Michigan elected to start increasing its AYP target until the required date of 2005. Michigan delayed the increases to the furthest extent possible, while Florida used equally spaced increases.

Minimum Number of Students in a Subgroup for Disaggregation

States were allowed to specify the minimum number of students in a subgroup that is required before disaggregated

Table 1. State Improvement Trajectories[a]

Back-Loaded	Straight With Plateaus	Straight	Front-Loaded	Unspecified
24	19[b]	4[c]	0	3

[a]Some states' increments of increase are only roughly equal across the trajectory, and most states' approximate increment levels vary based on grade level and/or subject area.

[b]All but one of the states whose trajectories were measured as straight with plateaus began increasing their annual measurement objectives in 2004/2005. The remaining state began its increase in 2003/2004. All but two of these trajectories have 3-year plateaus, meaning that there are 3 years between each AYP increase, the maximum period allowed for compliance. The remaining two states allow only 2 years between increases.

[c]Although officially implemented in 2002/2003, the requirement for state compliance holds that the first increment must take effect no later than the 2004/2005 school year. As such, trajectories were measured as straight for any state that began increasing on time. Of the four straight trajectory states, one began increasing in 2002/2003, one began in 2003/2004, and two began in 2004/2005.

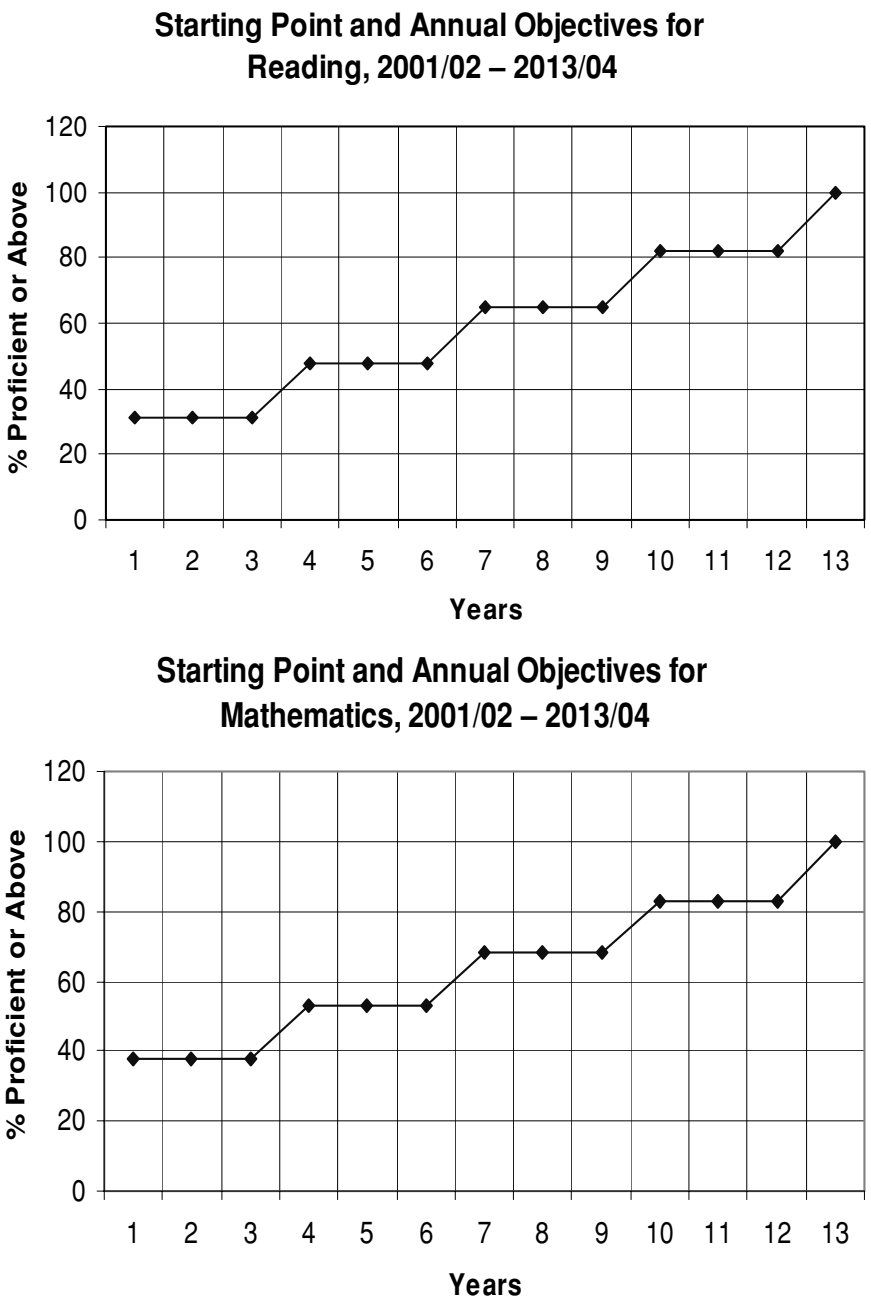

FIGURE 1. Example of straight with plateaus trajectory: Florida (duplicated as presented by the state with the vertical axis ending at 120%).

results are used to determine AYP for a school. The minimums selected by states and the number of states selecting each minimum for schools are listed in Table 2. As can be seen, 2 states did not specify a minimum. For the remaining 48 states, the minimum number of students (N) ranged from a low of 5 to a high of 100. The distribution is bimodal, with 14 states using a minimum of 40 and 13 using a minimum of 30. The smaller the minimum number, the larger the percentage of schools in the state that will be subject to accountability for disaggregated data and the greater the number of disaggregated groups per school. Thus, the smaller the minimum number of students, the higher the accountability standard.

States are not required to test in every grade in reading and mathematics for Grades 3 through 8 until the 2005/2006 school year. Most states begin the baseline year 2001/2002 testing at just two grade levels for Grades 3 through 8. As the number of grades tested increases, the number of students tested obviously increases and the percentage of schools that meet the minimum number of students required for reporting disaggregated data will also increase. Similarly, larger schools are more likely to meet the minimum number of students required. While data can be averaged over grade levels tested, they cannot be averaged over subjects (i.e., math and reading). Thus, much larger percentages of schools will be subjected to being held accountable for disaggregated data on identified subgroups after 2006.

Confidence Intervals

A number of states adopted the use of confidence intervals. If the confidence interval includes the annual measurable objective (AYP target), the school is counted as meeting the AYP requirement. For example, if 35% of the students in a given subgroup were proficient or above in a given subject when the objective for that subject and year was 40% or more proficient, the school would still meet its AYP requirement for that subject if the confidence interval had a width of 12 percentage points (i.e., $35 \pm 6\%$).

The use of confidence intervals by states in 2004 is summarized in Table 3. As can be seen, a substantial majority of states have elected to use some form of confidence interval. The states using confidence intervals differ, however, in the confidence level they use and in whether they use a one- or two-tailed interval.

The smaller the probability of a type I error (e.g., .01 versus .05), the larger the confidence interval. Similarly, a two-tailed confidence interval is larger than a one-tailed confidence interval. The larger the confidence interval, the more likely schools are to meet their AYP target. For example, a two-tailed confidence interval with a small probability of a type I error sets a relatively low standard for school accountability (i.e., everything else being equal, there will be a smaller percentage of schools identified as in need of improvement). While 11 states do not use a confidence interval—the highest standard—and 2 states use a confidence interval even narrower than a one-tailed 95% confidence interval, 16 states use a 99% confidence interval.

When the state design decision choices with regard to type of trajectory,

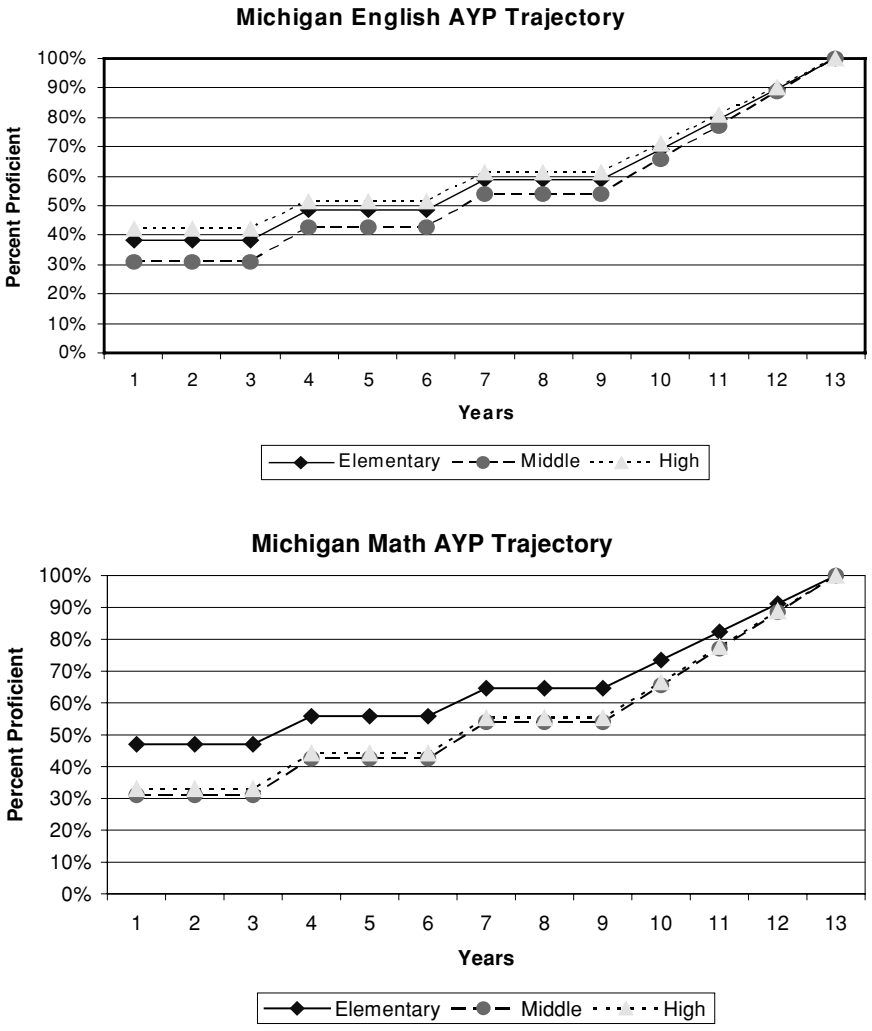

FIGURE 2. Example of back-loaded trajectory: Michigan.

minimum number of students per subgroup, and confidence interval are combined, states clearly vary substantially in their AYP approaches. The various AYP approaches with regard to these three design decisions are summarized in Table 4. Most cells in Table 4 contain three or fewer states and only one cell—back-loaded trajectory, minimum *N* of 40 or more, and a 99% confidence interval—has as many as five states. Among the states, these five chose to set school accountability to a relatively low standard. In contrast, two states use a straight-line trajectory, no confidence interval, and 40+ students for disaggregation. One of these two states, Illinois, set the first increase in percentage of students proficient for 2004 (before the deadline) and 40 students as the minimum for disaggregation, probably the highest standard for school accountability among all 50 states. (In the spring of 2005, however, Illinois received permission from the U.S. Department of Education to increase the minimum for disaggregation to 45 and use a 95% confidence interval.)

Consequences of Design Decisions

We used data for the state of Kentucky to explore the consequences of various design decisions for schools meeting or failing to meet AYP requirements. Kentucky design decisions are to hold schools accountable on disaggregated subgroups when there are 60 or more students of that subgroup in the school. They use a two-tailed 99% confidence interval and a back-loaded trajectory. The first increase in AYP target is 2005 (the latest allowed by NCLB), flat for 2 years, and a straight line from 2007 to 2014. Thus among all states, Kentucky design decisions are neither among the most demanding nor among the least demanding.

For 2002, 2003, and 2004—years for which we have data—Kentucky tested reading in Grades 4, 7, and 10 and mathematics in Grades 5, 8, and 11. In 2003,

Table 2. Minimum Number of Students Tested in a School for Subgroup Accountability[a,b]

Minimum Number	Number of States Using Minimum
100	1
60	1
52	1
50	3
45	2
42	1
40	14
34	1
30	13
25	2
20	4
11	1
10	3
5	1
None	2

[a]Seven states qualify their minimum numbers based on the percentage of the total membership of a school that a subgroup constitutes. For instance, California uses 100 as its minimum number except in schools where the subgroup comprises 15% or more of the school population. In this case, California uses 50 as its minimum number for holding schools accountable for that particular subgroup. In addition, Florida uses a minimum number of 10 for small schools.

[b]Eight states have separate, larger minimum numbers for limited English proficiency and/or students with disabilities subgroups.

Table 3. Use of Confidence Intervals

Type I Error	Number of States
Lower Than 95%	2
95	14 (3 One-Tailed)
98	1 (One-Tailed)
99	16 (2 One-Tailed)
Unspecified	6
Not Used	11

Table 4. State AYP Approach

	Back-Loaded				Straight With Plateaus				Straight				Unspecified			
	40+	25–39	<25	None	40+	25–39	<25	None	40+	25–39	<25	None	40+	25–39	<25	None
95%	3	2	1		3	2	1				1					1
99%	5	1	3		2	4		1								
Unspecified[a]	2	1	1		1		2		1				1			
None	1	4			2	1			2					1		

[a]The six states that reported that they did employ a confidence interval but did not report the size of that confidence interval, the two states that used one or one-half of a standard deviation for their confidence intervals, and the one state that applied a 98% confidence interval were all coded as "Unspecified."

the data set contained 1,160 schools, and for 2004, 1,150 schools. In the 2002 baseline year, 60% of Kentucky students were proficient or above in reading at the fourth grade and 56% were proficient at the seventh grade. In mathematics, 36% were proficient or above at the fifth grade, and 26% were proficient at the eighth grade. By 2004, these percents proficient or above had increased from 4 to 12 percentage points.

Table 5 lists the number of schools in Kentucky that would have failed various numbers of AYP categories (e.g., disaggregated groups and overall) under three scenarios: (1) using a 99% confidence interval and the trajectories with plateaus as actually employed by Kentucky, (2) using the regular trajectory that Kentucky has chosen but without confidence intervals, and (3) using a straight-line trajectory without confidence intervals. As required by NCLB, in our analyses a school fails to meet AYP if it fails to meet either the reading or the math targets. While NCLB also requires that schools test 95% or more of their students, fewer than 20 schools in Kentucky failed this criterion; our analyses ignore the 95% tested rule. Further, NCLB states that schools that otherwise failed AYP may be judged to meet AYP if they meet the safe harbor criterion. Safe harbor specifies that a school that otherwise fails AYP for a group of students will still meet AYP for that group if (1) the percentage of students not proficient decreased by 10% of the percentage not proficient from the previous year and (2) the subgroup made progress toward another academic indicator. Our analyses ignore the safe harbor criterion as well.

Since the first increase in AYP target is 2005 for Kentucky, the 99% confidence interval/no increase column in Table 5 reproduces Kentucky's actual results. All results in Table 5 are based on a minimum number of students in a subgroup of 60 to report disaggregated results, the standard used in Kentucky. The upper and lower panels of Table 5 provide results for 2003 and 2004, respectively. Using a 99% confidence interval, 1,042 schools (90%) met AYP requirements in 2003. That number increased to 1,081 (94%) in 2004 due to the fact that the annual measurable objective remained constant across those 2 years and there was some improvement in student achievement from 2003 to 2004.

Without confidence intervals, considerably fewer schools would have met AYP requirements with either the regular Kentucky trajectory with plateaus (no increase in AYP target) or the straight-line trajectory. The straight-line trajectory that has the first increase in AYP targets in 2003 leads to more schools failing to meet AYP requirements than using the regular Kentucky trajectory with plateaus. For example, in 2003 using the actual Kentucky design decisions, 90% of the schools met AYP. Dropping the use of confidence intervals would have led to 61% of schools meeting AYP, and dropping the confidence interval and using a straight-line projection would have led to only 45% of the schools meeting AYP. The results for 2004 are parallel in every respect but somewhat larger percentages. Perhaps not surprisingly, the bulk of the schools that fail AYP fail on just one category.

Table 5. Number and Percent of KY Schools That Would Have Failed Various Numbers of AYP Categories as a Function of Trajectory Used and the Use of Confidence Intervals (Disaggregation Minimum *N* = 60)[a]

Number of Categories Failed	99% Confidence Intervals/ No Increase	No Confidence Interval/ No Increase	No Confidence Interval/ Straight-Line Trajectory
	2003 Results		
0	1,042 (90%)	708 (61%)	521 (45%)
1	61 (5%)	272 (23%)	315 (27%)
2	24 (2%)	87 (8%)	142 (12%)
3	24 (2%)	74 (6%)	154 (13%)
4	7 (.8%)	16 (1%)	25 (2%)
5	2 (.2%)	3 (1%)	3 (1%)
	2004 Results		
0	1,081 (94%)	826 (72%)	679 (59%)
1	38 (39%)	202 (18%)	248 (22%)
2	17 (1%)	43 (4%)	83 (7%)
3	10 (1%)	67 (5%)	121 (11%)
4	4 (1%)	12 (1%)	18 (18%)
5	0 (0%)	0 (0%)	1 (0%)

[a]Kentucky uses a 99% confidence interval, minimum number of students of 60, and straight with plateaus trajectory with the first increase in 2005.

Table 6. The Effects of Varying Minimum Number of Students for Disaggregation on Number and Percent of Schools Meeting AYP

	2003	2004
Disaggregation ($N = 30$)		
Confidence Interval	970 (84%)	1,025 (89%)
Regular KY Trajectory	527 (45%)	656 (57%)
Straight-Line Trajectory	364 (31%)	510 (44%)
Disaggregation ($N = 40$)		
Confidence Interval	1,003 (86%)	1,058 (92%)
Regular KY Trajectory	599 (52%)	725 (68%)
Straight-Line Trajectory	417 (36%)	577 (50%)
Disaggregation ($N = 50$)		
Confidence Interval	1,027 (89%)	1,074 (93%)
Regular KY Trajectory	659 (57%)	786 (68%)
Straight-Line Trajectory	476 (41%)	639 (56%)
Disaggregation ($N = 60$)		
Confidence Interval	1,042 (90%)	1,081 (94%)
Regular KY Trajectory	708 (61%)	826 (72%)
Straight-Line Trajectory	521 (45%)	679 (59%)

Table 6 displays the number of schools that would have met all AYP requirements in 2003 or 2004 under the three approaches reported in Table 5 when combined with different minimum *N*s for disaggregated subgroup reporting. As would be expected, the number of schools that meet AYP requirements increases as the minimum *N* increases for all three scenarios. The most lenient approach uses the trajectory with plateaus as adopted by Kentucky with 99% confidence intervals and a minimum *N* of 60 per subgroup, while the most stringent is the straight-line trajectory without confidence intervals and a minimum *N* of 30. The number of schools meeting AYP requirements under these two extremes of the design conditions ranged from 364 (31%) to 1,042 (90%) for the 2003 data, and from 510 (44%) to 1,081 (94%) for the 2004 data.

We investigated the effects of school level and student body composition. In particular, we contrasted schools with 25% or more black students to schools with fewer than 25% black students to see what percentage of the schools failed AYP for each of the three conditions investigated in Table 5 (99% confidence intervals/no increase; no confidence intervals/no increase; no confidence interval/straight-line trajectory). The distribution of schools on the variable percent of black students was highly positively skewed with 15.3% of the schools having more than 25%. We selected 25% or more black as the cut because it represents a substantial fraction of the student body and a higher cut would have resulted in too few schools in the category 25% or more. We learned that schools with 25% or more black students were much more likely to fail AYP regardless of the design decisions. In fact, the difference between the two groups of schools was 25–30 points regardless of the three conditions in Table 5.

Similarly, we investigated the percentage of students on free and reduced price lunch, grouping schools according to having 66% or more such students, schools with 33–65% such students, and schools with fewer than 33%. The distribution of schools on the variable percentage of students on free and reduced price lunch was approximately normal with schools distributed across the full range from 0 to 100%. Again, we found schools with high percentages of students on free and reduced price lunch much more likely to fail AYP. For example, 16% of schools with 66% or more students on free and reduced price lunch failed AYP using a 99% confidence interval and no increase, while 59% failed when there was no confidence interval and no increase, and a full 73% of the schools failed when there was no confidence interval and a straight-line trajectory. In contrast, only 5% of the schools with fewer than 33% students on free and reduced price lunch failed AYP for a 99% confidence interval and no increase, 17% for no confidence interval and no increase, and 30% for no confidence interval and a straight-line trajectory. Thus, the effect of SES composition of the student body was more pronounced for the more demanding design decisions.

We investigated the percentage of schools failing AYP, contrasting elementary schools, middle schools, and high schools. We found that the percentage of elementary schools failing AYP was substantially less than either middle schools or high schools. For example, for the 99% confidence interval and no increase, only 5% of elementary schools failed AYP, while 21% of middle schools and 17% of high schools failed. The difference between elementary school and middle and high schools in percentage of schools failing AYP remained relatively unchanged when moving to the condition of no confidence interval/no increase and no confidence interval/straight-line trajectory. As before, however, the percentage of schools failing AYP increased when no confidence interval was used and again when there was a straight-line trajectory. Seventy-six percent of middle schools and 73% of high schools failed AYP under the most demanding design condition of no confidence interval/straight-line trajectory.

Using the Kentucky decision—99% confidence interval, disaggregated data for schools with 60 or more students, and no increase in AYP targets until 2005—we investigated the stability between 2003 and 2004 of school failure to meet AYP for the three groups most likely in Kentucky to result in an AYP failure: students receiving free and reduced price lunch, those in special education, and black students. The results of our consistency analyses are reported in Table 7. Of the 1,090 schools in Kentucky for which data could be disaggregated for these three groups in both years, 1,014, or 93%, met their AYP target for all three categories both years. Of the 35 schools that failed to meet their AYP target for free and reduced price lunch students in 2003, 10 also failed to meet the AYP target for students with free and reduced price lunch in 2004. Again, because there was no increase in AYP targets between 2003 and 2004 and some improvement in student achievement, the tendency was for schools that had failed their AYP target in 2003 to meet their AYP target in 2004.

Table 7. Cross-Tabulation of Categories of Failure to Meet AYP in 2003 and 2004[a]

	2003								
	Met	**1**	**2**	**3**	**1 & 2**	**1 & 3**	**2 & 3**	**1, 2, 3**	**Total**
2004									
Met	1,014	25	1	2	0	6	0	0	1,048
1	9	10	0	1	1	0	0	0	21
2	4	0	1	0	0	0	0	0	5
3	3	0	0	1	0	2	0	0	6
1 & 2	1	0	0	0	0	0	0	0	1
1 & 3	0	0	0	2	0	7	0	0	9
2 & 3	0	0	0	0	0	0	0	0	0
1, 2, 3	0	0	0	0	0	0	0	0	0
Total	1,031	35	2	6	1	15	0	0	1,090

[a]1 = free and reduced price lunch; 2 = special education; and 3 = black students.

For example, of the 59 schools that failed to meet their AYP target in 2003 in one or more of the three categories reported on in Table 7, 34 met all three AYP targets in 2004. Conversely, of the 42 schools that did not meet one or more of the three AYP targets in 2004, only 17 had met those targets in 2003. Clearly, there is some year-to-year instability in whether or not a school meets AYP. For these data with the vast majority of schools meeting AYP, however, the stability was good. As one might hope, the tendency was for schools failing AYP in 2003 to meet AYP in 2004.

Conclusions

In implementing NCLB school accountability requirements, states are left with discretion to make a number of important decisions. For example, states are free to design their own assessments of student achievement, though they must provide evidence that those assessments are aligned to their challenging academic content standards. States are also left free to make decisions about defining performance standards, including setting the cut for determining how much achievement is necessary for a student to be judged proficient. Here, we investigated the effects on numbers and percentages of schools meeting AYP of states' decisions about three design parameters in determining AYP. While our analyses are for all schools and all schools are labeled as meeting or failing to meet AYP, NCLB requires that consequences apply only to Title I schools.

One state decision concerns annual measurable target trajectories. All states must bring 100% of their students to proficient by the year 2014. States must increase their AYP target no later than 2005 and must increase it again every 3 years. There are a variety of approaches that states can and have used to set trajectories. States can, for example, postpone increases in their AYP targets (back loading) or alternatively have the bulk of their increases in the early years (front loading). Other alternatives are a straight-line projection from 2002 to 2014 or straight with plateaus. Perhaps not surprisingly, the two most popular approaches taken by states are back loaded (24 states) and straight with plateaus (19 states).

Another design decision concerns the minimum number of students necessary in a subgroup before school accountability applies at the disaggregated subgroup level. Here again, there is considerable variance among states, with two states applying the disaggregated criteria regardless of group size and one state requiring 100 students or more.

A third design decision is whether or not to use confidence intervals, and if so, the nature of the confidence interval. Again, there is variability among states, with 11 not using confidence intervals and 16 using 99% confidence intervals. Thus there is considerable variability among states in the design decisions they have made and, as a result, in how stringent their standards are for having a school meet AYP.

Using Kentucky data for the years 2002, 2003, and 2004, we conducted simulations to see size of effect of design decisions on numbers and percentages of schools judged as meeting AYP. The effects were substantial. For example, Kentucky design decisions are to use a disaggregation group size of 60, a two-tailed confidence interval of 99%, and no increase in AYP target until 2005. As a result, 90% of Kentucky schools met AYP in 2003, and 94% met AYP in 2004. Had the state decided to use 30 as the minimum number of students required for disaggregated subgroup accountability, these percentages would have dropped to 84% and 89%, respectively. Alternatively, had the state decided to drop its use of a confidence interval, the percent of schools meeting AYP would have dropped to 61% and 72%, respectively, for 2003 and 2004. Dropping both the confidence interval and moving to a straight-line trajectory that has an AYP target increase in both 2003 and 2004 would drop the percentage of schools meeting AYP to 45% and 59%, respectively. Combining all three design decisions to form the most stringent target for meeting AYP, only 31% of the schools in 2003 and 44% in 2004 would have met AYP.

The effects of the design decisions on number and percentage of schools meeting AYP can be quite large. As AYP targets increase over time, as they must, the design decisions may make the difference between having all schools fail AYP and having only significant numbers of schools fail AYP. In future years, the decision about minimum group size will become less important for states that are currently testing in only a few selected grades, because all states will be testing in Grades 3 through 8 and in at least one high school grade, thus substantially increasing the number of students tested per school. In contrast, decisions about trajectory will become increasingly important as the majority of states elected to postpone increases in AYP targets. In the summer of 2005, 47 states had requests into the U.S. Department of Education to make changes in their design decisions, with the clear majority aimed at making it easier for schools (and districts) to meet AYP targets.

Having increasingly large numbers of schools failing to meet AYP has implications for state resources. Title I schools are identified for improvement if they fail AYP for two consecutive years. Such schools must provide their students with the option to transfer to another public school in the district that has not been identified for improvement. The school must develop a school plan for

improvement covering a 2-year period that draws on research-based practices for improving teaching and learning in core academic subjects. The district must provide technical assistance as the school develops and implements the school plan. Schools identified as in need of improvement who continue to fail AYP must be provided even more technical assistance, including supplemental educational services, and may ultimately be restructured. Presumably, these remedial strategies are affordable and manageable by states and districts only if some reasonably small fraction of the total number of schools in the state is identified. The three design decisions investigated here may prove pivotal in the years to come in terms of a state's ability to provide required service to all schools identified and schools' ability to continue to exist.

Note

Based on a paper presented at the Annual Meeting of the National Council on Measurement in Education, as part of a symposium entitled Challenges, Contributions, and Consequences of State Accountability Systems, Montreal, Canada, April 12, 2005. Thanks to Tim Zeidner for his able assistance in analyzing the data.

References

Chubb, J., Linn, R., Haycock, K., & Wiener, R. (2005). Do we need to repair the monument? Debating the future of No Child Left Behind. *Education Next*, *5*(2), 9–19.

Cowan, K. T. (2003). *The new title I: The changing landscape of accountability*. Washington, DC: Thompson Publishing Group.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, *32*(7), 3–13.

Linn, R., Baker, E., & Betebenner, D. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, *31*(6), 3–16.

No Child Left Behind Act of 2001 (NCLB). (2001). Pub. Law No. 107.110.

United States Department of Education (2005). *Approved state accountability plans*. Retrieved from http://www.ed.gov/admins/lead/account/stateplans03/index.html.