

Explainer

July 2007

States' Evidence: What It Means to Make 'Adequate Yearly Progress'
Under NCLB

By Erin Dillon and Andrew J. Rotherham

ACKNOWLEDGEMENTS

This publication was made possible by a grant from Carnegie Corporation of New York. The statements made and views expressed are solely the responsibility of the authors.

The authors thank Kathy Choi and Elizabeth Mehr for their research assistance on this project.

ABOUT THE AUTHORS

ERIN DILLON is a policy analyst at Education Sector. She can be reached at edillon@educationsector.org.

ANDREW J. ROTHERHAM is co-founder and co-director of Education Sector. He also serves on the Commonwealth of Virginia Board of Education. The board's duties include establishing policies related to "adequate yearly progress" for schools in Virginia. He can be reached at arotherham@educationsector.org.

ABOUT EDUCATION SECTOR

Education Sector is an independent education policy think tank devoted to developing innovative solutions to the nation's most pressing educational problems. We are nonprofit and nonpartisan, both a dependable source of sound thinking on policy and an honest broker of evidence in key education debates throughout the United States.

ABOUT THIS SERIES

Education Sector Explainers give lay readers insights into important aspects of education policymaking. They are not intended to be technical manuals.

© Copyright 2007 Education Sector. All rights reserved.

1201 Connecticut Ave., N.W., Suite 850, Washington, DC 20036
202.552.2840 • www.educationsector.org

When Congress passed the federal No Child Left Behind Act (NCLB) in 2001, the law’s passage initiated a new educational ritual that plays out each summer: States, in reporting on student achievement, announce the schools that did or did not make “adequate yearly progress,” or “AYP.” Under NCLB, states must hold schools accountable for improving student performance. Specifically, the law requires states to set performance targets that schools must meet. The goal is to ensure that all schools improve their performance over time and have almost all of their students score “proficient” on state standardized tests by 2014.

This accountability system is a linchpin of the law, which is the latest reauthorization of the Elementary and Secondary Education Act of 1965. And, under NCLB, the consequences for missing AYP are substantial. Schools that do not meet performance targets for multiple years are deemed to be “in need of improvement” and face an escalating series of interventions, including giving students the chance to transfer to other public schools or using school funds for extra tutoring. These interventions can culminate in a school being completely restructured or even closed and reopened under new governance. (See sidebar, Page 3.)

Also, NCLB’s requirements for AYP are especially noteworthy because in order to meet them, schools must improve the performance of *all* groups of students, including minority and low-income students and those with disabilities, not just the overall average for students in the school. This was not the case with previous reauthorizations of the 1965 law. For instance, the 1994 version required states to set performance targets, but allowed states tremendous discretion and did not include the type of enforcement measures that are found under NCLB.¹

Yet, because the consequences for missing AYP under NCLB can be substantial, discussions of AYP often focus on a yes or no question: Did a particular school or school district make AYP or not? But the question that provides the most insight into a school’s performance is not *whether* a school made AYP, but rather *how* a school did or did not make AYP.

In practice, there are several ways for schools and districts to make AYP. And making AYP looks different

from state to state since NCLB allows each state to determine the specifics of how it calculates AYP. States can decide, for example, to average test scores across grades and years. This flexibility gives schools some leeway in meeting NCLB’s requirements and makes the requirements less strict than they might appear at first. It also renders many of the common assumptions about AYP—that it requires schools to get every single student to proficiency by 2014, that it does not recognize year-to-year improvements in school test scores, and that all students must achieve at the same level—inaccurate. As NCLB comes up for reauthorization, much of the debate will be about AYP. Thus, understanding the “how” of AYP can help teachers, parents, the public, and the news media make sense of the debate and the central element of NCLB’s accountability system.

This Education Sector Explainer provides these audiences with an aid to understanding how NCLB’s accountability system works overall and in different states, without weighing in on the merits of the law’s 2014 goal. We discuss the basics of “making” AYP and the multiple routes schools can take to get there, and we include data showing what the requirements are in each state to meet AYP this year and for the past two years.

The Basics of AYP

No Child Left Behind is best understood as a set of requirements that states must use to construct an accountability system. In exchange for complying with NCLB’s requirements, states receive federal aid for

elementary and secondary education. NCLB gives states, as part of the accountability plan requirements, guidelines for calculating adequate yearly progress. Each state's AYP plan is codified in its "Consolidated State Accountability Workbook," and is filed with and approved by the U.S. Department of Education.² States may make changes to their accountability plans, but all changes must be approved by the Department of Education.

Before students sit down to take the assessments at the center of a state's accountability plan, states must define the academic standards for what students should know and be able to do at each grade level in reading/language arts and math. States then develop or purchase assessments to test students on these standards and set passing or "cut" scores, which ultimately will determine the students who are "proficient" under NCLB.³

After students have taken the state assessment, usually in the spring, and the scores are calculated, states then evaluate student performance in schools against NCLB's three requirements for making AYP. These goals include: 1) Reaching proficiency requirements both overall and within each student subgroup (categories of students that include major racial groups, low-income students, English language learners, and students with disabilities) on state math and reading tests; 2) Meeting participation requirements, which ensure that enough students take the state assessments so that schools cannot exclude students who are not likely to earn passing scores; and 3) Meeting performance requirements on an "additional indicator" of academic performance determined by the state.

While NCLB allows states to decide the specifics of their accountability systems (i.e., in developing academic standards and student assessments and setting cut scores), there are two important non-negotiable requirements. First, all states must have 100 percent proficiency on state assessments for almost all students (excluding those with more severe disabilities) by 2014 as their ultimate goal. This does not mean that almost all students must answer every question correctly on

the state assessment, but that almost all students must reach a minimum level of achievement, defined as the state's cut score for "proficient" on its tests. This is a vital distinction because some states set low cut scores on their tests, making it easier for students in those states to qualify as "proficient" under NCLB.⁴ Secondly, all states must disaggregate scores by student subgroup and hold schools accountable for student performance in each subgroup. (See sidebar on subgroups, Page 5.)

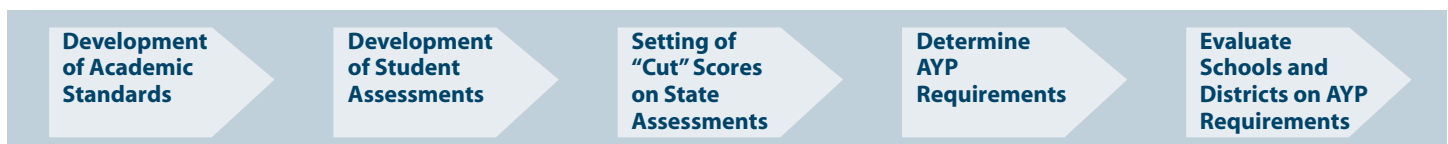
Proficiency Requirements

Annual proficiency benchmarks, called "annual measurable objectives," or "AMOs," are the primary measure of progress toward the 2014 goal for each state. For most states, these benchmarks are the minimum percentage of students in each NCLB subgroup that must score "proficient" on state assessments for a school or district to make AYP.

States may also use "index scores" as their annual measurable objectives, even though this option was not a part of the original law. (See sidebar on index systems, Page 11.) An index system rewards schools for improving student performance, even if students don't cross the "proficiency" threshold. NCLB requires that students reach the "proficient" level on state assessments, but states have multiple performance levels, usually ranging from "below basic" to "proficient" to "advanced." In general, index systems assign a number of points to each of these achievement levels, with the points increasing as achievement increases until students reach proficiency. But states cannot award more points for increasing student performance at levels above proficiency, because the Department of Education does not allow states to make up for poor performance at lower achievement levels by increasing the achievement of students already achieving at higher levels.

The Department of Education also allows states to use a modified definition of "proficient" for some special education students. Regulations instituted in 2003 and 2005 allow states to now count as proficient for AYP

Figure 1. Establishing an Accountability System



those students who are tested on alternate assessments or under modified standards. States may count proficient scores from these students, but only up to 3 percent of all students tested (which is equivalent to approximately 30 percent of special education students). (See sidebar on special education provisions, Page 4.)

Participation Requirements

NCLB requires that at least 95 percent of students participate in the state assessments in order for a school or district to make AYP. If less than 95 percent of students participate, the school or district will miss AYP for that year. Because AYP is based on the *percent of students who pass the test*, this provision is designed to prevent schools from artificially inflating pass rates by not assessing students who are not likely to do well on state tests.

With the participation requirement, NCLB actually requires that almost (excluding severely disabled students) 100 percent of at least 95 percent of students be proficient on state assessments by 2014. In shorthand, this is referred to as 100 percent proficiency, but the 95 percent participation rule actually gives schools and districts added leeway in meeting annual performance goals because schools are not held accountable for the performance of up to 5 percent of students in the school.

Additional Indicator Requirements

Schools and school districts must also meet an “additional indicator” of academic achievement in order to make AYP. States can choose an indicator for elementary and middle schools, such as the school attendance rate, achievement on additional state assessments, or retention rates, but for secondary schools that have graduating classes, this additional indicator must be graduation rates. NCLB defines a high school graduate as a student who earns a regular diploma within the “standard number of years,” generally interpreted as four years. But states can request approval from the Department of Education to define the standard number of years differently for students with special needs. In those cases, a student will be counted as graduating on time if his or her Individual Education Plan (a modified education plan for students with special needs) states that the student needs additional years in high school. English language

What Happens When a School Misses AYP?

Year 1 of Missing AYP: There are no consequences for the first year a school misses AYP.

Year 2 of Missing AYP: If a school misses AYP for a second consecutive year, it is identified as “in need of improvement.” The school must develop a two-year improvement plan in consultation with parents, school staff, and the school district. The plan should address core academic subjects and any specific subjects the school is struggling with. Students enrolled in the school now have the option to transfer to another school within the school district that has not been identified as “in need of improvement.” Priority is given to the lowest achieving students from low-income families enrolled in the school.

Year 3 of Missing AYP: If a school misses AYP for another consecutive year, the school must continue to offer students the option to transfer to another school, and must offer tutoring and other “supplemental education services” to students.

Year 4 of Missing AYP: If a school misses AYP for a fourth consecutive year, the school is identified for “corrective action.” Corrective action involves more serious steps to improve the school’s academic performance. Steps can include replacing staff, introducing new curricula, bringing in outside consultants to help with school performance, extending the school day or year, or changing the management structure of the school.

Year 5 of Missing AYP: If, after a full year of corrective action, a school misses AYP for a fifth consecutive year, the school will be placed under “restructuring.” The school must prepare a plan for an alternative governance arrangement, which can include reopening the school as a charter school, contracting management to a private, outside management group, turning the school over to the state for reorganization, or any other changes to school governance that “make fundamental reforms.”

Year 6 of Missing AYP: If the school misses AYP for a sixth consecutive year, it must implement the restructuring plan developed in the prior year.

learner students can take up to five years to graduate high school.

For elementary and middle schools, the most commonly used additional indicator is the school attendance rate.⁵ Additional indicators can be used to identify schools as “needing improvement,” but cannot be used in the reverse, that is, to remove schools from the list of schools needing improvement or those that did not make AYP. For instance, if a high school meets the proficiency requirements for all subgroups of students, but misses the additional indicator (graduation rate) requirement, that school will not make AYP. Alternately, if a school meets the additional indicator requirement, but misses the proficiency requirement, that school will still not make AYP that year.

Additional indicators were included in AYP to add a non-test-based measure of school performance. But, in practice, additional indicators create more opportunities for schools to miss AYP targets and often add to the confusion over what it means to miss AYP. A school might miss AYP because of low attendance, but do well on the student performance measures. Consequently, while more measures and indicators may broaden the accountability system from being solely test-based, policymakers must weigh this against having some schools miss AYP for reasons other than academic performance.

Timeline to Proficiency

Before states can hold schools accountable for AYP progress under NCLB, states must establish a timeline to “100 percent proficiency.” The timeline must include both the annual proficiency benchmarks or annual measurable objectives (AMOs) and incremental increases in AMOs, called “intermediate goals.”

AMOs must be established for both reading/language arts and math, and must identify the “minimum percentage of students who are required to meet or exceed the proficient level on the academic assessments.”⁶

Intermediate goals—the increases in annual measurable objectives—must happen at least every three years, and must be equal in size. Within these requirements, states are free to establish their own trajectory of increases over time.

If a state changes its assessment or begins assessing new grades, it often needs to recalibrate its annual measurable objectives and intermediate goals. This requires establishing a new starting point and AMOs based on student performance on the new test or in the newly tested grades. The Department of Education must approve such changes.

Calculating the Starting Point

The first step in establishing the timeline is to calculate a starting point, or the first measurable goal. States can use one of two methods, both of which are based on student performance on previous state assessments. The starting point, which must be calculated separately for elementary, middle, and high schools, and for reading and math, is the higher of two numbers derived by the following methods:⁸

Special Education Provisions

The Department of Education has allowed states some flexibility in the definition of “proficient” for special education students. New regulations in 2003 and 2005 gave states more options in assessing this subgroup of students.

The 2003 provision allows states to include scores from alternate assessments of students with severe cognitive disabilities when calculating AYP. States are free to test as many students as needed with alternate assessments, but the number of proficient or advanced scores on alternate assessments that can be counted toward state or district AYP is capped at 1 percent of all students assessed in the state or district. This is equivalent to approximately 10 percent of special education students.

The 2005 provision allows states to count proficient scores from students tested under modified standards when calculating AYP. The number of proficient or advanced students tested under modified standards that can be counted toward AYP is capped at 2 percent of all students—approximately 20 percent of special education students—tested in the state or district. This regulation is intended for students who are not severely disabled, but cannot achieve grade-level standards in the typical time frame.

Both the 1 percent and 2 percent caps apply only at the district and state level. Because the size of the population of disabled students varies by school, schools can count a larger percentage of students tested under either alternate assessments or modified standards as proficient for calculating AYP.⁷ Schools and districts must still meet the participation requirement for special education students, whether they are tested with alternate assessments, modified standards, or regular state assessments.

- The percent of students proficient in the lowest performing subgroup of students in the state (e.g. the percent proficient of English language learners or low-income students); or
- The percent of students proficient in the school at the 20th percentile of total public school enrollment in the state, with the schools ranked from lowest to highest by percent proficient.

The second option was added during the drafting of NCLB because the proficiency benchmarks under the first calculation—the percent proficient of the lowest performing subgroup of students—was in some cases too low to serve as a credible starting point. A starting point in the single digits, which it would have been for many states under the first option, presents a minimal challenge to most schools, undermining the goal of making NCLB an ambitious education reform measure. For example, when Hawaii calculated its starting points, the calculation of the lowest subgroup would have established a percent proficient starting point of 6 percent in reading/language

arts and only 2 percent in math. Under the second method, commonly referred to as the “20th percentile rule,” Hawaii established starting points of 30 percent for reading and 10 percent for math.⁹

Using the second method, states rank all schools in order from lowest performing (lowest percent of students scoring proficient on the state test) to highest performing (highest percent of students scoring proficient). States then count enrollments in the schools, starting at the

bottom, until they reach 20 percent of total school enrollment in the state. The percent proficient of the school at this point—the 20th percentile of enrollment—becomes the starting point for all schools in the state.

Choosing Objectives, Goals

Once the state has established its starting point, the next step is to determine the annual measurable objectives and intermediate goals that will establish the state’s timeline

NCLB’s Subgroups

NCLB’s goal is to improve academic achievement among *all* students. Thus, states are required to disaggregate student test scores by subgroup. Showing test scores for each group of students prevents states from masking low student performance among some groups with high performance of students overall.

For example, in Connecticut, 75 percent of students in grades 3–8 scored proficient on the state reading assessment, meaning that the state met the performance target for 2005–06 in overall student performance. But, this data disaggregated by subgroup tells a different story. Most groups did not reach the performance target, and some, such as students with disabilities and English language learners scored far below the AMO (see figure below).

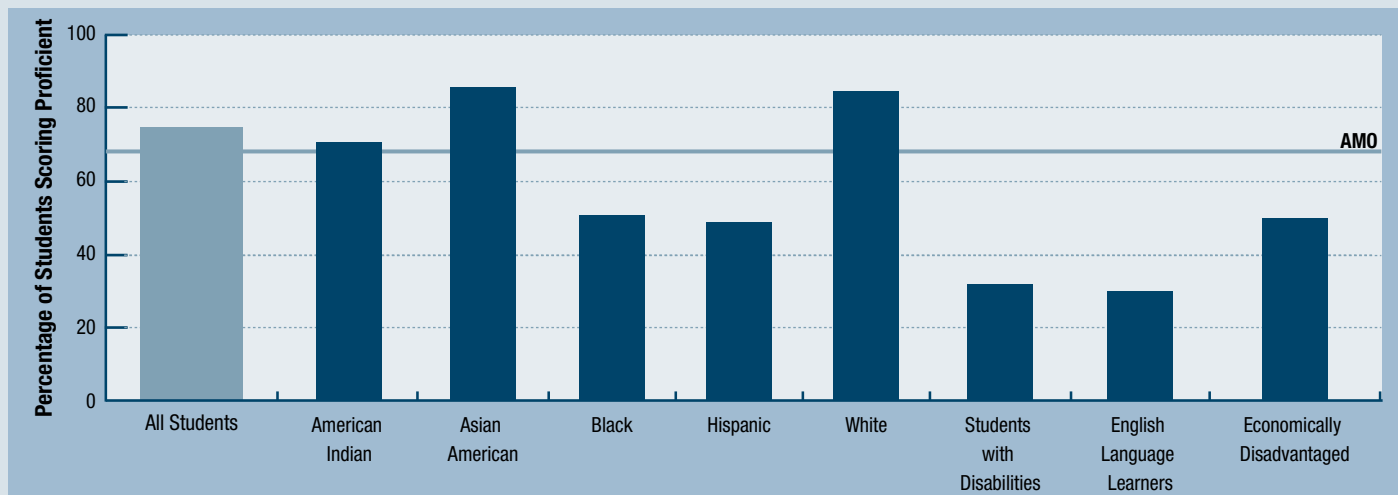
The original NCLB legislation defines four kinds of subgroups:

- Economically disadvantaged: This is defined as students who receive free or reduced price lunch.
- Major racial and ethnic subgroups: These subgroups include African American students, white students, Asian/Pacific Islander students, Native American students, and Latino students.

- Students with disabilities: The Individuals with Disabilities Education Act (IDEA) provides definitions for identifying students with disabilities. Neither IDEA nor NCLB define severe cognitive disabilities, which is the category used for administering alternate assessments. States can establish their own definition of severe cognitive disabilities.
- Students with Limited English Proficiency, or English language learners: The Department of Education allows states to decide whether to include assessment results in AYP calculations for students who are enrolled in their first year of school in the United States. States may also include students in this subgroup for up to two years after they have attained English proficiency.

In 2005, the Department of Education added a subgroup for students who were displaced by Hurricanes Katrina and Rita. States that enrolled a large number of displaced students could apply for a waiver for the 2005–06 school year that would allow them to create a displaced students subgroup. Schools in those states would then not be held accountable for the performance of students in that subgroup for the 2005–06 school year.¹⁷

Connecticut State Reading Results, Grades 3–8, 2005–06



Source: Connecticut State Department of Education, *Connecticut No Child Left Behind (NCLB) State Report: 2005–06 School Year*.

of proficiency benchmarks that schools and districts must meet in order to make AYP.

NCLB requires that increases in AMOs occur at least every three years and be equal in size. This prevents states from establishing a timeline with small increases initially and big jumps as 2014 nears. But states can make increases in benchmarks more frequent, and many states do, expecting increases every three years in the beginning and then increases every year in the last few years. Michigan, for instance, uses this type of trajectory (see Figure 2). Chester E. Finn, an education analyst and former Department of Education official, compares this option to the lending industry’s “balloon mortgages,” where borrowers make small payments in the beginning and put off substantial payments until the final years of a loan.¹⁰

Not all states use the “balloon mortgage” approach. Some have opted for a stair-stepped approach across all 12 years, and others have chosen to increase AMOs each year to 100 percent proficiency. Arkansas’ AYP timeline is one example of this continual increase approach (see Figure 3).

Because states can establish their own starting points and intermediate goals, AMOs vary widely between states. The result is that making AYP can mean a very different thing from one state to the next. As states progress to the 100 percent goal, annual benchmarks will begin to

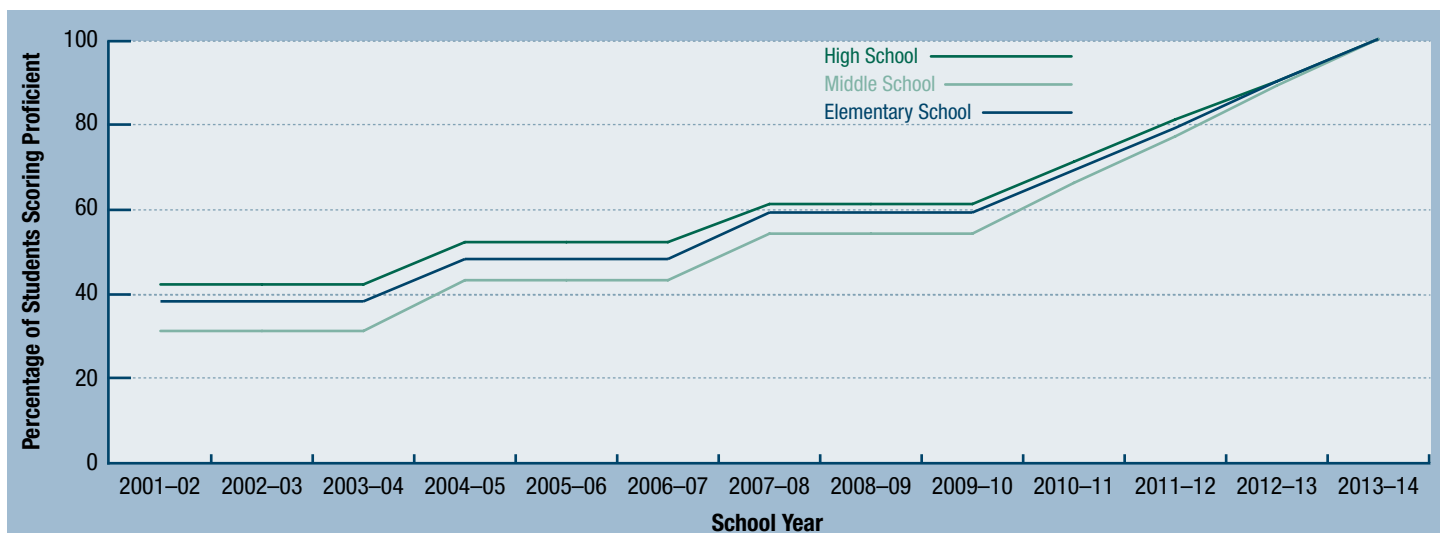
converge across states. But because states are currently only halfway to the 2014 goal, proficiency benchmarks vary among states. As Table 1 (on Page 8) shows, the median percent of students that need to be proficient for a school to make AYP ranges from 50 percent to 67 percent. But there is also wide variation between states, with the lowest AMOs ranging from 20 percent to 24 percent, and the highest ranging from 75 percent to 90 percent. (See Appendices for the exact annual measurable objectives for each state in the 2004–05, 2005–06, and 2006–07 school years.)

Overall, today’s annual goals are actually quite modest. Only five states have *any* AMOs in the 80 percent to 100 percent range for the 2006–07 school year, at any grade level or subject. In other words, in only five states do at least eight in 10 students have to pass the state test for the school to make AYP right now. Of these, only three—Colorado, South Dakota, and Tennessee—have AMOs in that range for multiple grade levels or subject areas. (See Appendices.)

The AYP Decision

Once the annual benchmarks are established, states must evaluate schools and districts against those benchmarks and determine whether a school or district made AYP. AYP decisions can be divided into two primary categories—“regular” AYP and “safe harbor” AYP. Regular AYP is the

Figure 2. Michigan’s Reading/Language Arts AYP Trajectory



Source: Michigan Department of Education, *Michigan Consolidated State Application Accountability Workbook*, updated June 2005.

first cut for schools and districts, and those that do not make regular AYP are then evaluated for safe harbor AYP. For both regular and safe harbor AYP, schools must meet the 95 percent participation requirement and the state’s additional indicator requirement.

Within these two ways of making AYP, multiple factors are considered before a decision on AYP is made. And NCLB allows each state to decide key elements of how AYP will be calculated in that state. In particular, states are able to decide how many students must be in a subgroup for it to count for accountability purposes, whether to average scores over multiple years or across multiple grade-levels, and whether to use statistical techniques like “confidence intervals” when making AYP determinations. How states decide to employ these options can either increase the reliability of an accountability system or diminish the performance requirements.¹¹

‘Regular’ AYP

The first test of whether a school or district makes AYP for a given year is whether or not the percent of students scoring proficient—overall and in each student subgroup—for both math and reading meets or exceeds that year’s annual measurable objective. This decision—whether a school meets the AMO for each subgroup—is not as simple as a clear yes or no. Again, states use multiple calculations to determine if a school has met that year’s AMO. Under NCLB’s provision that

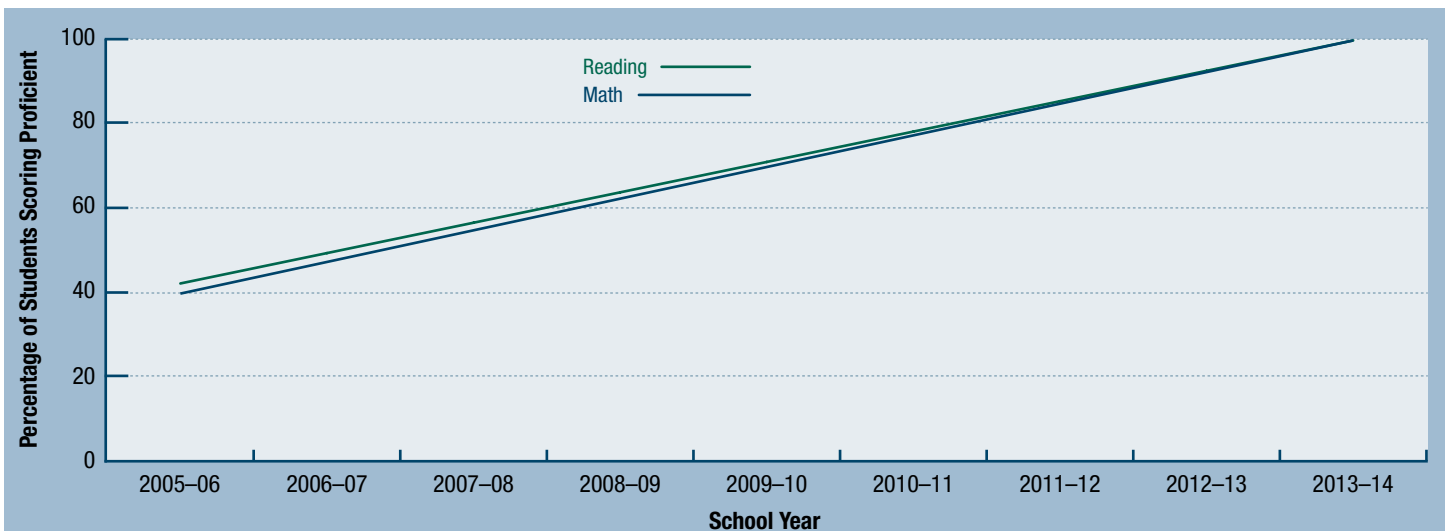
AYP should be “statistically valid and reliable,” states have successfully petitioned the Department of Education to add provisions to protect schools from being incorrectly identified as not making AYP. But the inclusion of some of these considerations can work against NCLB’s primary goal of holding schools accountable for students in each subgroup, and of ensuring that schools reach 100 percent proficiency.

Below are the most common factors states include in AYP calculations:

Minimum Sample Size or “n” Size: Under NCLB, states are responsible for establishing the minimum number of students that can be used in calculating AYP for a school or district. This is commonly referred to as the “n” size, referring to the use of “n” to denote population size in statistics. If a school has a small number of students in a particular subgroup or grade, the performance of just one or two of those students can have a big impact on the average performance of the entire group. To avoid having one or two students skew the results of an entire group and potentially affect the accountability status of a school, states establish a minimum group size that needs to be tested in order for a school to be held accountable for that group’s performance. The most common size is 40 students.¹²

But a large minimum sample size means schools won’t be held accountable for some groups of students as originally

Figure 3. Arkansas’ Elementary Grades AYP Trajectory



Source: Arkansas Department of Education, *Arkansas Consolidated State Application Accountability Workbook*, updated February 15, 2007.

intended in NCLB. For example, California has a minimum sample size of up to 100 students, meaning that if a school has 98 students who are English language learners, the scores for those students are not evaluated as a separate subgroup. Instead, the scores are grouped with all student scores and may be overshadowed by the achievement of another group of students. The Commission on No Child Left Behind, a bipartisan group charged with providing recommendations to improve NCLB, suggests minimum n-sizes of no larger than 20 students in order to ensure accountability for all students.¹³

States must also establish a minimum group size for reporting purposes. This is meant to protect students' privacy when schools are publicly reporting student test-score data. The minimum group size for reporting purposes is generally smaller, around 10 students, than the minimum sample size for accountability purposes.

Confidence Intervals: In addition to minimum sample sizes, states have received approval from the Department of Education to use confidence intervals for calculating AYP. Confidence intervals create a “buffer zone”—a percentage above and below a school’s test score—to ensure schools and districts are not unfairly penalized for random fluctuations in student test scores. This plus or minus gives schools and districts more leeway in the percent of students that need to be proficient in order to meet AYP for a given year. In 2006, nearly every state used confidence intervals in determining AYP.¹⁴

The most common confidence intervals are 95 percent or 99 percent, which means they are set at a level to ensure that 95 out of 100 times or 99 out of 100 times the number falls within the buffer zone. The higher the confidence interval the larger the buffer zone must be, and fewer students will need to be proficient to meet the AMO. The size of the school, grade, or subgroup being measured has a large impact on how much leeway a school has to make AYP. The smaller the group size, the smaller the percent of students that need to be proficient for the group to meet the performance targets.

Confidence intervals can serve the important function of ensuring that schools are not identified as missing AYP because of random year-to-year test-score fluctuations. On the other hand, large confidence intervals can distort a school’s performance, qualifying it as making AYP when it may not be anywhere near the proficiency

Table 1. Target Percentage of Students Scoring Proficient on State Assessments, 2006–07

	Median	Minimum (State)	Maximum (State)
Elementary Math	56%	21% (ME)	84% (CO)
Elementary Reading	67%	24% (CA)	83% (CO)
Middle School Math	50%	20% (NM)	79% (TN)
Middle School Reading	60%	24% (CA)	83% (TN)
High School Math	54%	20% (ME)	75% (TN)
High School Reading	64%	22% (CA)	90% (TN)

Note: Eight states with index scores as AMOs were excluded from analysis. States included in the analysis (n = 43) report AMOs as the minimum percent of students scoring proficient on the state assessments.

benchmark. Washington, for example, uses a large confidence interval of 99 percent, meaning that if a grade contains 37 students and the performance target is 64.2 percent proficiency, only 45 percent of students actually need to score proficient for the school to make AYP. The Commission on No Child Left Behind recommends limiting confidence intervals to 95 percent, which provides schools with a smaller “buffer zone” in which to meet their performance targets.

Rolling averages: NCLB allows states to use rolling averages to determine if a school or district makes AYP. States may average a school or district’s current years’ test scores with the previous one or two years and use the average to determine AYP. States also may combine this technique with a “high score” measure so that the AYP determination is based on the higher of the multiyear average or the current year’s scores. States also can average scores across grades in a given school. Averaging scores across years and grades further decreases the chance that random variation in scores from year to year, or grade to grade, could impact a school or district’s AYP status. Nearly every state allows schools or districts to average scores across multiple years and/or grades when calculating AYP.

Growth Models: The Department of Education has fully or conditionally approved growth model pilot programs for nine states as of July 2007.¹⁵ These states use assessments and data systems that allow them to track the growth of individual students over time and hold schools and districts that might otherwise fall short of

AMOs under the regular AYP approach accountable for substantial progress toward the 2014 goal. This is different from measures like safe harbor that look at year-to-year changes in achievement for a particular grade or group at a school, but do not measure achievement gains for individual students. The goal of growth models is to give schools credit for making noteworthy progress without undermining the policy goal of having students on grade-level by 2014.

Full Academic Year: Schools are only held accountable for students who have been enrolled for a full academic year. States must define “full academic year” in their state accountability plans, and the definition varies from state to state. Recently, states have started to define “full year” students as those who are enrolled in a school from the test date of one year to the test date of the next year. In other words, students need to be enrolled in the same school over a summer in order for their test scores to count toward the school’s accountability status. These provisions are intended to ensure that schools are not held accountable for the performance of students they did not have the opportunity to teach.

This factor can have large consequences for schools with mobile student populations, which are often the same schools struggling with low student performance. Under these rules, many students at these schools can be excluded from NCLB accountability requirements.

‘Safe Harbor’ AYP

NCLB legislation includes a “safe harbor” provision, a measure allowing schools to make AYP by improving their student performance, even if they do not meet the proficiency benchmarks. While this is a type of “growth model” in that it measures change in student performance from one year to the next, it only measures changes in achievement at a particular grade level or school; it does not follow achievement changes for individual students or cohorts of students as the new pilot growth models do.

Safe harbor AYP requires that a school or district reduce the percentage of students who are *below* proficient by 10 percent from the previous year. For example, if a school has 30 percent of students scoring proficient on the reading assessment, and the annual measurable objective is 40 percent proficiency, that school would not make regular AYP. If, however, only 22 percent of

students scored proficient the previous year, meaning that 78 percent were below proficient, the school would have reduced the percentage of students not proficient by just over 10 percent (from 78 percent to 70 percent) and made AYP that year through safe harbor. Schools and districts can use safe harbor calculations for specific subgroups of students or for students overall. But in order to use safe harbor a school also must meet the performance target for its additional indicator requirement, for instance, attendance rates, graduation rates, or achievement in another subject.

The safe harbor provision is intended to reward schools for improving student performance, even if they are not yet meeting the annual measurable objectives. But, because safe harbor requires a 10 percent reduction from the previous year in the percent of non-proficient students, a school could potentially make AYP each year through safe harbor and never reach 100 percent proficiency by 2014. This is a little-understood provision in the law and another reason the 2014 goal is somewhat less absolute than implied in the political and public debates about NCLB.

As with regular AYP, there are a number of factors that states consider when calculating safe harbor AYP.

Safe Harbor Confidence Intervals: Approximately half of the states have added confidence intervals to the safe harbor provision. This provides schools and districts with even more leniency in achieving the 10 percent reduction required to make safe harbor AYP. To mitigate the problem of too much leniency in the safe harbor provision, the Department of Education has limited states to using lower confidence intervals, typically 75 percent, which provide less leeway than higher confidence intervals of 95 percent or 99 percent.

Secondary Safe Harbor: Louisiana, Pennsylvania, and South Carolina use index scores as a secondary safe harbor measure. Index scores credit schools for increases in student achievement all along the achievement scale, and not just for the students who get over the proficiency bar. In Pennsylvania, schools can make AYP either through regular AYP, through the safe harbor provision, or by meeting Pennsylvania Performance Index targets.

Additional Indicator and Participation Rate Safe Harbor: States also have implemented safe harbor

provisions for the additional indicator requirement of AYP and the participation requirement. In these states, schools can meet their additional indicator requirement by reducing the percent of non-graduates, absences, or students not participating in state assessments by 10 percent of the previous year.

Understanding AYP

There is no doubt: AYP is complicated. The variation among states in how AYP is calculated, along with the multiple ways an individual school could potentially make AYP, leads to considerable confusion over what, exactly, it means to make AYP. As AYP is currently set up, there is no single answer to that question. Instead, it is necessary to understand how each state calculates AYP, and how a particular school made—or missed—those targets.

Adding to the confusion, the rhetoric about AYP in public and political debates often makes NCLB's accountability requirements appear stricter than they are in practice. States have considerable flexibility in deciding how to

calculate AYP and how to hold schools accountable for student performance. Safe harbor, the special education provisions, confidence intervals, and minimum group sizes, for example, all give schools leeway in reaching the 100 percent proficiency mark.

Because of this flexibility, the number of schools labeled “in need of improvement” under NCLB has been smaller than originally predicted, especially by the law's critics.¹⁶ But a small number of schools identified for improvement also may indicate that the accountability system lacks rigor—especially against a national backdrop of substantial achievement gaps and high dropout rates.¹⁷

Ultimately, no accountability system is perfect. The goal for policymakers is to maximize reliability within the constraints of existing data and assessment systems and minimize problems and unintended consequences. Understanding how NCLB's accountability system works today and what we can learn from how it is implemented at the school, district, and state level is essential to understanding both the debate about the law, and what it means for an individual school to make or miss AYP.

Index Systems

Most states set their annual measurable objectives or AMOs as the minimum percent of students that need to be proficient for a school to make AYP. But some states use an index system to determine if schools made AYP.

Instead of having one proficiency benchmark for students, states have turned to index systems in order to allow schools to demonstrate growth below the proficient level. An index system credits schools and districts for raising student achievement, even for students who did not reach the proficiency threshold. But schools and districts are not credited for increases in student achievement above proficiency. This prevents schools from using the performance of high-achieving students to compensate for students who are below proficient. There are several methods for calculating and using index scores.

The “partial credit” method: Minnesota and Wisconsin use a “partial credit” method for calculating index scores. In these two states, students scoring at the lowest level on the state assessment receive zero points, students at the next level receive a half-point, and students at the proficient level or above receive full points. The number of students at each level is counted and multiplied by the number of points for that level. The resulting sum is the school or district’s index score.

In New York, a school’s index score is calculated by multiplying the percentage of students scoring at the “basic” level by one, and adding that to the percentage of students scoring at “proficient” or above multiplied by two. In New York’s system, the maximum score a school could achieve, meaning that all students score proficient, is 200.

Weighted average method: The annual measurable objectives in Alabama, Mississippi, Iowa, Michigan, Washington, and Wyoming are expressed as percent proficient, not an index score. But schools and districts in these states are not simply compared with the AMO benchmark. Instead, the annual measurable objective is used to calculate a weighted average index score across grades. This allows the state to reward schools for growth and to combine scores across different grade-level combinations. If the index score is above zero, the school has made AYP. If it is below zero and outside any confidence interval requirements, the group has not made AYP. Below is an example of the weighted average formula:

The percent proficient cut-off (AMO) for elementary reading in a state is 40 percent. For the subgroup of low-income students in this school, the following percent scored proficient in each grade:

Grade 1: 45% proficient, n = 50

Grade 2: 38% proficient, n = 45

Grade 3: 35% proficient, n = 48

The annual measurable objective is then subtracted from the percent proficient in each grade:

Grade 1: $45 - 40 = +5\%$

Grade 2: $38 - 40 = -2\%$

Grade 3: $35 - 40 = -5\%$

A weighted average is calculated by multiplying the percent by which each grade is above or below the annual measurable object by its proportion of the total enrollment.

Grade 1: $5\% \times .35 = 1.75$

Grade 2: $-2\% \times .31 = -.62$

Grade 3: $-5\% \times .34 = -1.70$

These numbers are summed across grades and the resulting number, if it is above zero, indicates that the school made AYP. If it is below zero, the school did not make AYP, barring any confidence interval considerations.

$1.75 + -.62 + -1.70 = -.57$

This school did not make AYP, but only barely, which means that it might make AYP with confidence interval calculations.

Average points per student method: Massachusetts uses its index score to combine the achievement of students on its state assessment, the MCAS, with student achievement on the MCAS-Alt, the state assessment for students with disabilities. The number of students scoring at each performance level is multiplied by the number of points that level receives. The points are then summed and divided by the total number of students tested, with the result being the average number of points per student. Below is an example of this system:

Performance Level	Points per Level	Number of Students Tested	Total Points
Proficient & Advanced	100	200	20,000
Needs Improvement–High	75	180	13,500
Needs Improvement–Low	50	150	7,500
Warning/Failing–High	25	75	1,875
Warning/Failing–Low	0	60	0
Total		665	42,875

For this school, the Composite Performance Index score is 64.5.

New Hampshire, Vermont, and Rhode Island follow a similar system to calculate index scores. South Carolina, Louisiana, and Pennsylvania also use a similar formula, but use the results as a supplementary safe harbor calculation for schools. In addition to setting annual measurable objectives of percent proficient, these states have established index score annual measurable objectives. If a school misses the percent proficient requirement, but meets the AMO for index scores, it is considered to have made AYP.

Oklahoma’s method: Oklahoma calculates the index score for schools and districts based on a school or district’s performance on seven indicators. The first indicator is student performance on the Oklahoma state assessment in both reading and math. Other indicators included in calculating Oklahoma’s index scores are “school completion” measures, including attendance, drop out, and school completion rates; and “academic excellence” measures, which include ACT scores, college remediation rates, and Advanced Placement credit.

Endnotes

- ¹ The 1994 version of the Elementary and Secondary Education Act, the “Improving America’s Schools Act,” required states to set performance targets, but these provisions allowed states tremendous discretion and state and federal enforcement was uneven. As a result, when the law came up for renewal in 1999 a top goal of civil rights and children’s advocacy organizations, for example the Citizens Commission on Civil Rights and The Education Trust, was the enactment of a more prescriptive accountability system that held schools more strictly accountable for the performance of all students, not just overall averages. Their efforts led directly to the AYP requirements in the current iteration of the Elementary and Secondary Education Act, the No Child Left Behind Act. See, for instance, *Title I in Midstream: The Fight to Improve Schools for Poor Kids* (Washington, D.C.: Citizens Commission on Civil Rights, 1998); *Promising Results, Continuing Challenges: The Final Report of the National Assessment of Title I* (Washington, D.C.: United States Department of Education, 1999); Patrick J. McGuinn, *No Child Left Behind and the Transformation of Federal Education Policy, 1965-2005* (Lawrence, KS: University of Kansas Press, 2006).
- ² State accountability plans are available online through the U.S. Department of Education at: <http://www.ed.gov/admins/lead/account/stateplans03/index.html>.
- ³ For more about how states set cut scores see Andrew J. Rotherham, *Making the Cut: How States Set Passing Scores on Standardized Tests* (Washington, D.C.: Education Sector, July 2006).
- ⁴ Ibid.
- ⁵ Institute for Education Sciences, *National Assessment of Title I: Interim Report, Volume 1* (Washington, D.C.: U.S. Department of Education, February 2006).
- ⁶ *No Child Left Behind Act of 2001*, Public Law 107-110, 107th Cong., (January 8, 2002), Subpart 1, Sect. 1111, State Plans.
- ⁷ The U.S. Department of Education’s final regulations on the 2003 special education changes are available online at: <http://www.ed.gov/legislation/FedRegister/finrule/2003-4/120903a.html>. Final regulations for the 2005 changes are available online at: <http://www.ed.gov/legislation/FedRegister/finrule/2007-2/040907a.html>.
- ⁸ *The ABCs of ‘AYP’* (Washington, D.C.: The Education Trust, Updated September 2004).
- ⁹ Accountability Resource Center Hawaii, *Reading and Mathematics AYP Starting Points, Intermediate Goals, and Annual Measurable Objectives* (Honolulu: State of Hawaii Department of Education, February 2003).
- ¹⁰ Chester E. Finn Jr., “Adequate Yearly Progress or Balloon Mortgage?” *The Education Gadfly*, January 30, 2003.
- ¹¹ See Kevin Carey, *Hot Air: How States Inflate Their Educational Progress Under NCLB* (Washington, D.C.: Education Sector, 2006).
- ¹² *No Child Left Behind at Five: A Review of Changes to State Accountability Plans* (Washington, D.C.: Center on Education Policy, January 2007).
- ¹³ Commission on No Child Left Behind, *Beyond NCLB: Fulfilling the Promise to Our Nation’s Children* (Washington, D.C.: The Aspen Institute, 2007).
- ¹⁴ *No Child Left Behind at Five: A Review of Changes to State Accountability Plans*.
- ¹⁵ North Carolina, Tennessee, Arkansas, Delaware, Florida, Iowa, Arizona, and Alaska have received full approval. Ohio has received conditional approval.
- ¹⁶ Lynn Olson, “Data Show Schools Making Progress on Federal Goals,” *Education Week*, September 8, 2004.
- ¹⁷ Kevin Carey, *Hot Air: How States Inflate Their Educational Progress Under NCLB*.
- ¹⁸ Ellen Forte and William J. Erpenbach, *Statewide Educational Accountability Under the No Child Left Behind Act—A Report on 2006 Amendments to State Plans* (Washington, D.C.: Council of Chief State School Officers, November 2006).

Appendix I. State Annual Measurable Objectives by Subject Area for Elementary School, School Years 2004–05 Through 2006–07

School Year	2004–05		2005–06		2006–07	
	Math	Reading	Math	Reading	Math	Reading
Alabama	61	68	67	73	67	73
Alaska	57.61	71.48	57.61	71.48	57.61	71.48
Arizona	*43.3(G3)	*53.3(G3)	54	45	54	45
Arkansas	*46.14	*48.84	40	42.4	47.5	49.6
California	26.5	24.4	26.5	24.4	26.5	24.4
Colorado	81.9	82.69	83.64	82.69	83.64	82.69
Connecticut	74	68	74	68	74	68
D.C.	*48.67	*41.92	40.28	47.37	40.28	47.37
Delaware	41	62	41	62	50	68
Florida	44	37	50	44	56	51
Georgia	58.3	66.7	58.3	66.7	58.3	66.7
Hawaii	28	44	28	44	28	44
Idaho	60	72	60	72	70	76
Illinois	47.5	47.5	47.5	47.5	55	55
Indiana	57.1	58.8	64.3	65.7	64.3	65.7
Iowa	68.3	70	68.3	70	68.3	70
Kansas	60.1	63.4	60.1	63.4	66.8	69.5
Kentucky	32.14	53.86	32.14	53.86	32.14	53.86
Louisiana	41.8	47.4	41.8	47.4	41.8	47.4
Maine	21	41	21	41	21	41
Maryland	53.6	57.8	58.8	62.5	63.9	67.2
Massachusetts†	64.75	78.05	68.7	80.5	72.6	82.95
Michigan	56	48	56	48	56	48
Minnesota†	69.6(G3)*	66.5(G3)*	69.64	69.48	69.64	69.48
Mississippi	62	75	62	75	62	75
Missouri	17.5	26.6	26.6	34.7	35.8	42.9
Montana	40	55	51	74	51	74
Nebraska	74	72	74	72	74	72
Nevada	*45.4	*39.6	*45.4	*39.6	43.3	39.6
New Hampshire†	**	**	76	82	76	82
New Jersey	62	75	62	75	62	75
New Mexico	24.13	40.85	28	45	33	49
New York†	*142	*131	86	122	86	122
North Carolina	*81	76.7	65.8	76.7	65.8	76.7
North Dakota	59.3	73.8	63.8	76.7	68.3	79.6
Ohio	46.6	71.2(G3)	60.6(G3)	71.2(G3)	60.6(G3)	71.2(G3)
Oklahoma†	790	768	790	768	932	914
Oregon	49	50	49	50	49	50
Pennsylvania	45	54	45	54	45	54
Rhode Island†	68.1	80.1	68.1	80.1	68.1	80.1
South Carolina	36.7	38.2	36.7	38.2	36.7	38.2
South Dakota	54	78	65	78	65	82
Tennessee	79	83	79	83	79	83
Texas	42	53	42	53	50	60
Utah	64	71	64	71	71	77
Vermont†	*361	*414	390	403	390	403
Virginia	63	65	67	69	71	73
Washington	47.3	64.2	47.3	64.2	47.3	64.2
West Virginia	67	72	67	72	72.5	76.7
Wisconsin†	47.5	67.5	47.5	67.5	47.5	67.5
Wyoming	36.5	42	36.5	42	36.5	42

†Indicates states whose annual measurable objectives are reported as index score targets. All other AMOs are reported as the percent of students scoring proficient.

States establish annual measurable objectives either across all grades, by grade span, or by individual grade. Except where noted in parentheses, elementary refers to the grade span K–5, or individual grade 4; middle refers to the grade span 6–8, or individual grade 8; high school refers to the grade span 9–12, or individual grades 10 or 11.

*These annual measurable objectives are based on an old assessment system. Subsequent AMOs reflect revised performance targets. Generally, AMOs were revised because of a change in the state assessment or in the grades assessed.

**New Hampshire determined AYP status for elementary and middle grades by attendance rate only for school year 2004–05 because of a transition to a grades 3–8 testing system. High School AYP status will be determined solely on graduation rates for school year 2006–07 because of a transition from spring to fall testing.

Appendix II. State Annual Measurable Objectives by Subject Area for Middle School, School Years 2004–05 Through 2006–07

School Year	2004–05		2005–06		2006–07	
	Math	Reading	Math	Reading	Math	Reading
Alabama	48	43	48	51	55	51
Alaska	57.61	71.48	57.61	71.48	57.61	71.48
Arizona	22.5	42.5	22.5	42.5	22.5	42.5
Arkansas	*36.48	*38.59	29.1	35.2	37.96	43.3
California	26.5	24.4	26.5	24.4	26.5	24.4
Colorado	69.63	80.21	69.63	80.21	69.63	80.21
Connecticut	74	68	74	68	74	68
D.C.	*33.17	*28.08	40.55	43.58	40.55	43.58
Delaware	41	62	41	62	50	68
Florida	44	37	50	44	56	51
Georgia	58.3	66.7	58.3	66.7	58.3	66.7
Hawaii	28	44	28	44	28	44
Idaho	60	72	60	72	70	76
Illinois	47.5	47.5	47.5	47.5	55	55
Indiana	57.1	58.8	64.3	65.7	64.3	65.7
Iowa	65	66.7	65	66.7	65	66.7
Kansas	60.1	63.4	60.1	63.4	66.8	69.5
Kentucky	26.93	52.4	26.93	52.4	26.93	52.4
Louisiana	41.8	47.4	41.8	47.4	41.8	47.4
Maine	22	42	22	42	22	42
Maryland	35.8	56.7	42.9	61.5	50	66.3
Massachusetts†	64.75	78.05	68.7	80.5	72.6	82.95
Michigan	43	43	43	43	43	43
Minnesota†	*76.1(G7)	*77.8(G7)	58.39	64.04	58.39	64.04
Mississippi	42	48	42	48	42	48
Missouri	17.5	26.6	26.6	34.7	35.8	42.9
Montana	40	55	51	74	51	74
Nebraska	69	71	69	71	69	71
Nevada	*43.3	*47.5	*43.3	*47.5	43.3	39.6
New Hampshire†	**	**	76	82	76	82
New Jersey	49	66	49	66	49	66
New Mexico	10.58	34.14	15	38	20	42
New York†	*93	*116	86	122	86	122
North Carolina	*81	76.7	65.8	76.7	65.8	76.7
North Dakota	50	71.1	55.5	74.3	61.1	77.5
Ohio	47.3(G6)	46.7(G6)	47.5	73.8	47.5	73.8
Oklahoma†	790	768	790	768	932	914
Oregon	49	50	49	50	49	50
Pennsylvania	45	54	45	54	45	54
Rhode Island†	55.1	73.3	55.1	73.3	55.1	73.3
South Carolina	36.7	38.2	36.7	38.2	36.7	38.2
South Dakota	54	78	65	78	65	82
Tennessee	79	83	79	83	79	83
Texas	42	53	42	53	50	60
Utah	64	71	64	71	71	77
Vermont†	*341	*382	390	403	390	403
Virginia	63	65	67	69	71	73
Washington	38	47.6	38	47.6	38	47.6
West Virginia	64	75	64	75	70	79.17
Wisconsin†	47.5	67.5	47.5	67.5	47.5	67.5
Wyoming	37.75	45.42	37.75	45.42	37.75	45.42

†Indicates states whose annual measurable objectives are reported as index score targets. All other AMOs are reported as the percent of students scoring proficient.

States establish annual measurable objectives either across all grades, by grade span, or by individual grade. Except where noted in parentheses, elementary refers to the grade span K–5, or individual grade 4; middle refers to the grade span 6–8, or individual grade 8; high school refers to the grade span 9–12, or individual grades 10 or 11.

*These annual measurable objectives are based on an old assessment system. Subsequent AMOs reflect revised performance targets. Generally, AMOs were revised because of a change in the state assessment or in the grades assessed.

**New Hampshire determined AYP status for elementary and middle grades by attendance rate only for school year 2004–05 because of a transition to a grades 3–8 testing system. High School AYP status will be determined solely on graduation rates for school year 2006–07 because of a transition from spring to fall testing.

Appendix III. State Annual Measurable Objectives by Subject Area for High School, School Years 2004–05 Through 2006–07

State	2004–05		2005–06		2006–07	
	Math	Reading	Math	Reading	Math	Reading
Alabama	68	81	73	84	73	84
Alaska	57.61	71.48	57.61	71.48	57.61	71.48
Arizona	25	35.8	25	35.8	25	35.8
Arkansas	*32.81	*39.63	29.2	35.5	38.05	43.56
California	20.9	22.3	20.9	22.3	20.9	22.3
Colorado	60.25	84.74	60.25	84.74	60.25	84.74
Connecticut	69	72	69	72	69	72
D.C.	*33.17	*28.08	40.55	43.58	40.55	43.58
Delaware	41	62	41	62	50	68
Florida	44	37	50	44	56	51
Georgia	62.3	81.6	68.6	84.7	68.6	84.7
Hawaii	28	44	28	44	28	44
Idaho	60	72	60	72	70	76
Illinois	47.5	47.5	47.5	47.5	55	55
Indiana	57.1	58.8	64.3	65.7	64.3	65.7
Iowa	74.2	74.2	74.2	74.2	74.2	74.2
Kansas	46.8	58	46.8	58	55.7	65
Kentucky	29.79	29.35	29.79	29.35	29.79	29.35
Louisiana	41.8	47.4	41.8	47.4	41.8	47.4
Maine	20	50	20	50	20	50
Maryland	40.7	53.3	47.3	58.5	53.9	63.7
Massachusetts†	64.75	78.05	68.7	80.5	72.6	82.95
Michigan	44	52	44	52	44	52
Minnesota†	*76.7	*82.3	28.13	64.77	28.13	64.77
Mississippi	28	37	28	37	28	37
Missouri	17.5	26.6	26.6	34.7	35.8	42.9
Montana	40	55	51	74	51	74
Nebraska	72	75	72	75	72	75
Nevada	52.3	77.9	52.3	77.9	52.3	77.9
New Hampshire†	64	77	64	77	**	**
New Jersey	64	79	64	79	64	79
New Mexico	18.29	37.3	22	41	27	45
New York†	139	148	146	154	152	159
North Carolina	70.8	35.4	70.8	35.4	70.8	35.4
North Dakota	43.1	57.2	49.4	61.9	55.7	66.7
Ohio	60	71.8	60	71.8	60	71.8
Oklahoma†	790	768	790	768	932	914
Oregon	49	50	49	50	49	50
Pennsylvania	45	54	45	54	45	54
Rhode Island†	54	68.8	54	68.8	54	68.8
South Carolina	50	52.3	50	52.3	50	52.3
South Dakota	67	66	54	66	54	72
Tennessee	75	90	75	90	75	90
Texas	42	53	42	53	50	60
Utah	47	70	47	70	59	76
Vermont†	326	384	326	384	326	384
Virginia	63	65	67	69	71	73
Washington	43.6	61.5	43.6	61.5	43.6	61.5
West Virginia	59	71	59	71	65.83	75.83
Wisconsin†	47.5	67.5	47.5	67.5	47.5	67.5
Wyoming	46.5	57	46.5	57	46.5	57

†Indicates states whose annual measurable objectives are reported as index score targets. All other AMOs are reported as the percent of students scoring proficient.

States establish annual measurable objectives either across all grades, by grade span, or by individual grade. Except where noted in parentheses, elementary refers to the grade span K–5, or individual grade 4; middle refers to the grade span 6–8, or individual grade 8; high school refers to the grade span 9–12, or individual grades 10 or 11.

*These annual measurable objectives are based on an old assessment system. Subsequent AMOs reflect revised performance targets. Generally, AMOs were revised because of a change in the state assessment or in the grades assessed.

**New Hampshire determined AYP status for elementary and middle grades by attendance rate only for school year 2004–05 because of a transition to a grades 3–8 testing system. High School AYP status will be determined solely on graduation rates for school year 2006–07 because of a transition from spring to fall testing.