

CRS Report for Congress

Adequate Yearly Progress (AYP): Implementation of the No Child Left Behind Act

Updated October 31, 2008

Wayne C. Riddle
Specialist in Education Policy
Domestic Social Policy Division



Prepared for Members and
Committees of Congress

Adequate Yearly Progress (AYP): Implementation of the No Child Left Behind Act

Summary

Title I, Part A of the Elementary and Secondary Education Act (ESEA), authorizes financial aid to local educational agencies (LEAs) for the education of disadvantaged children and youth at the preschool, elementary, and secondary levels. Over the last several years, the accountability provisions of this program have been increasingly focused on achievement and other outcomes for participating pupils and schools. Since 1994, and particularly under the No Child Left Behind Act of 2001 (NCLB), a key concept embodied in these requirements is that of “adequate yearly progress (AYP)” for schools, LEAs, and states. AYP is defined primarily on the basis of aggregate scores of various groups of pupils on state assessments of academic achievement. The primary purpose of AYP requirements is to serve as the basis for identifying schools and LEAs where performance is unsatisfactory, so that inadequacies may be addressed first through provision of increased support and, ultimately, a variety of “corrective actions.”

Under NCLB, the Title I-A requirements for state-developed standards of AYP were substantially expanded. AYP calculations must be disaggregated — determined separately and specifically for not only all pupils but also for several demographic groups of pupils within each school, LEA, and state. In addition, while AYP standards had to be applied previously only to pupils, schools, and LEAs participating in Title I-A, AYP standards under NCLB must be applied to *all* public schools, LEAs, and to states overall, if a state chooses to receive Title I-A grants. However, corrective actions for failing to meet AYP standards need be applied only to schools and LEAs participating in Title I-A. Another major break with the pre-NCLB period is that state AYP standards must incorporate concrete movement toward meeting an ultimate goal of all pupils reaching a proficient or advanced level of achievement by 2014.

The overall percentage of public schools identified as failing to make AYP for one or more years on the basis of test scores in 2006-2007 was approximately 28% of *all* public schools. The percentage of schools for individual states varied from 4% to 75%. Approximately 12% of Title I-A participating schools were in the “needs improvement” status (i.e., they had failed to meet AYP standards for 2 consecutive years or more) on the basis of AYP determinations for 2005-2006 and preceding school years.

The AYP provisions of NCLB are challenging and complex, and they have generated substantial interest and debate. Debates regarding NCLB provisions on AYP have focused on the provision for an ultimate goal, use of confidence intervals and data-averaging, population diversity effects, minimum pupil group size (n), separate focus on specific pupil groups, number of schools identified and state variations therein, the 95% participation rule, state variations in assessments and proficiency standards, and timing. The authorization for ESEA programs expired at the end of FY2008, and the 111th Congress is expected to consider whether to amend and extend the ESEA. This report will be updated regularly to reflect major legislative developments and available information.

Contents

Background: Title I Outcome Accountability and the AYP Concept	1
General Elements of AYP Provisions	2
Generic AYP Factors	2
AYP Provisions Under the IASA of 1994	5
Concerns About the AYP Provisions of the IASA	6
AYP Under NCLB Statute	8
ED Regulations and Guidance on Implementation of the AYP Provisions of NCLB	11
Recent Developments	12
Regulations Proposed in April 2008 on Title I-A Assessments and Accountability	12
Growth Models	13
Data on Schools and LEAs Identified as Failing to Meet AYP	26
Schools Failing to Meet AYP Standards for One or More Years	26
Schools Failing to Meet AYP Standards for Two Consecutive Years or More	27
LEAs Failing to Meet AYP Standards	27
Issues in State Implementation of NCLB Provisions	29
Introduction	29
Ultimate Goal	30
Confidence Intervals and Data-Averaging	32
Population Diversity Effects	33
Minimum Pupil Group Size (n)	33
Separate Focus on Specific Pupil Groups	34
Number of Schools Identified and State Variations Therein	36
95% Participation Rule	38
State Variations in Assessments and Proficiency Standards	38

List of Tables

Table 1. Categories of Pupils with Disabilities with Respect to Achievement Standards, Assessments, and AYP Determinations Under ESEA Title I-A	22
Table 2. Reported Percentage of Public Schools and Local Educational Agencies (LEAs) Failing to Make Adequate Yearly Progress (AYP) on the Basis of Spring 2007 Assessment Results	28

Adequate Yearly Progress (AYP): Implementation of the No Child Left Behind Act

Background: Title I Outcome Accountability and the AYP Concept

Title I, Part A of the Elementary and Secondary Education Act (ESEA), the largest federal K-12 education program, authorizes financial aid to local educational agencies (LEAs) for the education of disadvantaged children and youth at the preschool, elementary, and secondary levels.

Since the 1988 reauthorization of the ESEA (The Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988, or “School Improvement Act,” P.L. 100-297), the accountability provisions of this program have been increasingly focused on achievement and other outcomes for participating pupils and schools. Since the subsequent ESEA reauthorization in 1994 (the Improving America’s Schools Act of 1994, P.L. 103-382), and particularly under the No Child Left Behind Act of 2001 (NCLB, P.L. 107-110), a key concept embodied in these outcome accountability requirements is that of “adequate yearly progress (AYP)” for schools, LEAs, and (more recently) states overall. The primary purpose of AYP requirements is to serve as the basis for identifying schools and LEAs where performance is inadequate, so that these inadequacies may be addressed, first through provision of increased support and, ultimately, through a variety of “corrective actions.”¹

This report is intended to provide an overview of the AYP concept and several related issues, a description of the AYP provisions of the No Child Left Behind Act, and an analysis of the implementation of these provisions by the U.S. Department of Education (ED) and the states. The authorization for ESEA programs expired at the end of FY2008, and the 111th Congress is expected to consider whether to amend and extend the ESEA. This report will be updated regularly to reflect major legislative developments and available information.

¹ These corrective actions, as well as possible performance-based awards, are not discussed in detail in this report. For information on them, see CRS Report RL33371, *K-12 Education: Implementation Status of the No Child Left Behind Act of 2001 (P.L. 107-110)*, by Gail McCallion, et al., Section 4.

General Elements of AYP Provisions

ESEA Title I, Part A has included requirements for participating LEAs and states to administer assessments of academic achievement to participating pupils, and to evaluate LEA programs at least every two years, since the program was initiated in 1965. However, relatively little attention was paid to school- or LEA-wide outcome accountability until adoption of the School Improvement Act of 1988.² Under the School Improvement Act, requirements for states and LEAs to evaluate the performance of Title I-A schools and individual participating pupils were expanded. In addition, LEAs and states were for the first time required to develop and implement improvement plans for pupils and schools whose performance was not improving. However, in comparison to current Title I-A outcome accountability provisions, these requirements were broad and vague. States and LEAs were given little direction as to how they were to determine whether performance was satisfactory, or how performance was to be defined, with one partial exception.

The exception applied to schools conducting schoolwide programs under Title I-A. In schoolwide programs, Title I-A funds may be used to improve instruction for all pupils in the school, rather than being targeted on only the lowest-achieving individual pupils in the school (as under the other major Title I-A service model, targeted assistance schools). Under the 1988 version of the ESEA, schoolwide programs were limited to schools where 75% or more of the pupils were from low-income families (currently this threshold has been reduced to 40%). The School Improvement Act required schoolwide programs, in order to maintain their special authority, to demonstrate that the academic achievement of pupils in the school was higher than either of the following: (a) the average level of achievement for pupils participating in Title I-A in the LEA overall; or (b) the average level of achievement for disadvantaged pupils enrolled in that school during the three years preceding schoolwide program implementation.

The embodiment of outcome accountability in the specific concept of AYP began with the 1994 Improving America's Schools Act (IASA). Under the IASA, states participating in Title I-A were required to develop AYP standards as a basis for systematically determining whether schools and LEAs receiving Title I-A grants were performing at an acceptable level. Failure to meet the state AYP standards was to become the basis for directing technical assistance, and ultimately corrective actions, toward schools and LEAs where performance was consistently unacceptable.

Generic AYP Factors. Before proceeding to a description of the Title I-A AYP provisions under the IASA of 1994, we outline below the general types of major provisions frequently found in AYP provisions, actual or proposed.

Primary Basis: They are based primarily on aggregate measures of academic achievement by pupils. As long as Title I-A has contained AYP provisions, it has provided that these be based ultimately on state standards of curriculum content and

² For additional information on this legislation, see CRS Report 89-7, *Education for Disadvantaged Children: Major Themes in the 1988 Reauthorization of Chapter 1*, by Wayne Riddle (out of print, available from author [7-7382] upon request).

pupil performance, and assessments linked to these standards. More specifically, the Title I-A requirements have been focused on the percentage of pupils scoring at the “proficient” or higher level of achievement on state assessments, not a common national standard. However, when AYP provisions were first adopted in 1994, states were given an extended period of time to adopt and implement these standards and assessments, and for a lengthy period after the 1994 amendments, various “transitional” performance standards and assessments were used to measure academic achievement.³

Ultimate Goal: AYP standards may or may not incorporate an ultimate goal, which may be relatively specific and demanding (e.g., all pupils should reach the proficient or higher level of achievement, as defined by each state, in a specified number of years), or more ambiguous and less demanding (e.g., pupil achievement levels must increase in relation to either LEA or state averages or past performance). If there is a specific ultimate goal, there may also be requirements for specific, numerical, annual objectives either for pupils in the aggregate or for each of several pupil groups. The primary purpose of such a goal is to require that levels of achievement continuously increase over time in order to be considered satisfactory.

Subject Areas: With respect to subject areas, AYP standards might focus only on reading and math achievement, or they might include additional subject areas.

Additional Indicators: In addition to pupil scores on assessments, AYP standards often include one or more supplemental indicators, which may or may not be academic. Examples include high school graduation rates, attendance rates, or assessment scores in subjects other than those that are required.

Levels at Which Applied: States may be required to develop AYP standards for, and apply them to, schools, LEAs, or for states overall. Further, it may be required that AYP standards be applicable to *all* schools and LEAs, or only to those participating in ESEA Title I-A.

Disaggregation of Pupil Groups: AYP standards might be applied simply to all pupils in a school, LEA, or state, or they might also be applied separately and specifically to a variety of demographic groups of pupils — such as economically disadvantaged pupils, pupils with disabilities, pupils in different ethnic or racial groups, or limited English proficient pupils. In a program such as Title I-A, the purpose of which is to improve education for the disadvantaged, it may be especially important to consider selected disadvantaged pupil groups separately, to identify situations where overall pupil achievement may be satisfactory, but the performance of one or more disadvantaged pupil groups is not.

Basic Structure: The basic structure of AYP Models generally falls into one of three general categories. The three basic structural forms for AYP of schools or LEAs are the *group status*, *successive group improvement*, and *individual/cohort*

³ For additional information on the standard and assessment requirements under ESEA title I-A, see CRS Report RL31407, *Educational Testing: Implementation of ESEA Title I-A Requirements Under the No Child Left Behind Act*, by Wayne C. Riddle.

growth models. In the context of these terms, “group” (or “subgroup,” in the case of detailed demographic categories) refers to a collection of pupils that is identified by their grade level and usually other demographic characteristics (e.g., race, ethnicity, or educational disadvantage) as of a point in time. The actual pupils in a “group” may change substantially, or even completely, from one year to the next. In contrast, a “cohort” refers to a collection of pupils in which the same pupils are followed from year to year.

The key characteristic of the group status model is a required threshold level of achievement that is the same for all pupil groups, schools, and LEAs statewide in a given subject and grade level. Under this model, performance at a point in time is compared to a benchmark at that time, with no direct consideration of changes over a previous period and whatever the school’s or LEA’s “starting point.” For example, it might be required that 45% or more of the pupils in any of a state’s elementary schools score at the proficient or higher level of achievement in order for a school to make AYP. “Status” models emphasize the importance of meeting certain minimum levels of achievement for all pupil groups, schools, and LEAs, and arguably apply consistent expectations to all.

The key characteristic of the successive group improvement model is a focus on the rate of change in achievement in a subject area from one year to the next among groups of pupils in a grade level at a school or LEA (e.g., the percentage of this year’s 5th grade pupils in a school who are at a proficient or higher level in mathematics compared to the percentage of last year’s 5th grade pupils who were at a proficient or higher level of achievement).

Finally, the key characteristic of the individual/cohort growth model is a focus on the rate of change over time in the level of achievement among cohorts of the same pupils. Growth models are longitudinal, based upon the tracking of the same pupils as they progress through their K-12 education careers. While the progress of pupils is tracked individually, results are typically aggregated when used for accountability purposes. Aggregation may be by demographic group, by school or LEA, or other relevant characteristics. In general, growth models would give credit for meeting steps along the way to proficiency in ways that a status model typically does not.

Alternative or “Safe Harbor” Provisions: AYP systems often have alternative provisions under which schools or LEAs that fail to meet the usual requirements may still be deemed to have made AYP if they meet certain specified alternative conditions. For example, under a status model, it might be generally required that 45% or more of the pupils in any of a state’s elementary schools score at the proficient or higher level of achievement in order for the school to make AYP, but a school where aggregate achievement is below this level might still be deemed to have made AYP, through a “safe harbor” provision, if the percentage of pupils at the proficient or higher level in the school is higher than for the previous year by some specified degree. Such a concept may be seen as adding a successive group improvement model element to a status model of AYP.

Assessment Participation Rate: It might be required that a specified minimum percentage of a school's or LEA's pupils participate in assessments in order for the school or LEA to be deemed to have met AYP standards. The primary purposes of such a requirement are to assure that assessment results are broadly representative of the achievement level of the school's pupils, and to minimize the incentives for school staff to discourage test participation by pupils deemed likely to perform poorly on assessments.

Exclusion of Certain Pupils: Beyond general participation rate requirements (see above), states may be specifically required to include, or allowed to exclude, certain groups of pupils in determining whether schools or LEAs meet AYP requirements. For example, statutory provisions might allow the exclusion of pupils who have attended a school for less than one year in determining whether a school meets AYP standards.

Special Provisions for Pupils with Particular Educational Needs: Beyond requirements that all pupils be included in assessments, with accommodations where appropriate, there may be special provisions for limited English proficient (LEP) pupils or pupils with the most significant cognitive disabilities.

Averaging or Other Statistical Manipulation of Data: Finally, there are a variety of ways in which statistical manipulation of AYP-related data or calculations might be either authorized or prohibited. Major possibilities include averaging of test score data over periods of two or more years, rather than use of the latest data in all cases; or the use of "confidence intervals" in calculating whether the aggregate performance of a school's pupils is at the level specified by the state's AYP standards. These techniques, and the implications of their use, are discussed further below. In general, their use tends to improve the reliability and validity of AYP determinations, while often reducing the number of schools or LEAs identified as failing to meet AYP standards.

AYP Provisions Under the IASA of 1994

Under the IASA, states were to develop and implement AYP standards soon after enactment. However, states were given several years (generally until the 2000-2001 school year) to develop and implement curriculum content standards, pupil performance standards, and assessments linked to these for at least three grade levels in math and reading.⁴ Thus, during the period between adoption of the IASA in 1994 and of NCLB in early 2002, for most states the AYP provisions were based on "transitional" assessments and pupil performance standards that were widely varying in nature. AYP standards based on such "transitional" assessments were considered to be "transitional" themselves, with "final" AYP standards to be based on states' "final" assessments, when implemented. The subject areas required to be included in state AYP standards (as opposed to required assessments) were not explicitly specified in statute; ED policy guidance required states to include only math and

⁴ For more information on all aspects of the ESEA Title I-A assessment requirements, see CRS Report RL31407, *Educational Testing: Implementation of the ESEA Title I-A Requirements Under the No Child Left Behind Act*, by Wayne C. Riddle.

reading achievement in determining AYP. Further, the inclusion in AYP standards of measures other than academic achievement in math and reading on state assessments was optional.

With respect to the ultimate goal of the state AYP standards, the IASA provided broadly that there must be continuous and substantial progress toward a goal of having all pupils meet the proficient and advanced levels of achievement. However, no timeline was specified for reaching this goal, and most states did not incorporate it into their AYP plans in any concrete way.

The IASA's AYP standards were to be applied to schools and LEAs, but not to the states overall. Further, while states were encouraged to apply the AYP standards to *all* public schools and LEAs, states could choose to apply them only to schools and LEAs participating in Title I-A, and most did so limit their application.

The IASA provided that all relevant pupils⁵ were to be included in assessments and AYP determinations, although assessments were to include results for pupils who had attended a school for less than one year only in tabulating LEA-wide results (i.e., not for individual schools). LEP pupils were to be assessed in the language that would best reflect their knowledge of subjects other than English; and accommodations were to be provided to pupils with disabilities.

Importantly, while the IASA required state assessments to ultimately (by 2000-2001) provide test results that were disaggregated by pupil demographic groups, it did not require such disaggregation of data in AYP standards and calculations. The 1994 statute provided that state AYP standards must consider all pupils, "particularly" economically disadvantaged and LEP pupils, but did not specify that the AYP definition must be based on each of these pupil groups separately. Finally, the statute was silent with respect to data-averaging or other statistical techniques, as well as the basic structure of state AYP standards (i.e., whether a "group status," "successive group improvement," or "individual/cohort growth" model must be employed).

Concerns About the AYP Provisions of the IASA. Thus, the IASA's provisions for state AYP standards broke new ground conceptually, but were comparatively broad and ambiguous. Although states were required to adopt and implement at least "transitional" AYP standards, on the basis of "transitional" state assessment results, soon after enactment of the IASA, they were not required to adopt "final" AYP standards, in conjunction with final assessments and pupil performance standards, until the 2000-2001 school year. Further, states were not allowed to implement most corrective actions, such as reconstituting school staff, until they adopted final assessments, so these provisions were not implemented by most states until the IASA was replaced by NCLB.

⁵ All pupils in states where AYP determinations were made for all public schools, or all pupils served by ESEA Title I-A in states where AYP determinations were made only for such schools and pupils.

A compilation was prepared by the Consortium for Policy Research in Education (CPRE) of the “transitional” AYP standards that states were applying in administering their Title I-A programs during the 1999-2000 school year.⁶ Overall, according to this compilation, the state AYP definitions for 1999-2000 were widely varied and frequently complex. General patterns in these AYP standards, outlined below, reflect state interpretation of the IASA’s statutory requirements.

- Most considered only achievement test scores, but some considered a variety of additional factors, most often dropout rates or attendance rates.
- Often, the state AYP standards set a threshold of some minimum percentage, or minimum rate of increase in the percentage, of pupils at the proficient or higher level of achievement on a composite of state tests. These thresholds were often based, at least in part, on performance of pupils in a school or LEA relative to statewide averages or to the school’s or the LEA’s performance in the previous year. Several states identified schools as failing to make AYP if they fail to meet “expected growth” in performance on the basis of factors such as initial achievement levels and statewide average achievement trends. These thresholds almost never incorporated a “ladder” of movement toward a goal of all pupils at the proficient level, or otherwise explicitly incorporated an ultimate goal to be met by some specific date.
- While some state AYP standards were based on achievement results for a single year, they were frequently based on two- or three-year rolling averages.
- The AYP standards generally referred only to all pupils in a school or LEA combined, without a specific focus on any pupil demographic groups. However, the AYP standards of some states included a focus on a single category of low-achieving pupils separately from all pupils, and a very few (e.g., Texas) included a specific focus on the performance of several pupil groups (African American, Hispanic, White, or Economically Disadvantaged). One state (New Mexico) compared school scores to predicted scores on the basis of such factors as pupil demographics.
- The state AYP standards under the IASA were sometimes substantially adjusted from year-to-year (often with consequent wide variations in the percentage of Title I-A schools identified as needing improvement). According to CPRE, two states (Iowa and New Hampshire) left AYP standards and determinations almost totally to individual LEAs in 1999-2000.

A report published by ED in 2004, on the basis of state AYP policies for the 2001-2002 school year, contains similar conclusions about state AYP policies in the period immediately preceding implementation of NCLB.⁷ There was tremendous

⁶ See [http://www.cpre.org/Publications/Publications_Accountability.htm].

⁷ U.S. Department of Education, Office of the Undersecretary, Policy and Program Studies (continued...)

variation among the states in the impact of their AYP policies under the IASA on the number and percentage of Title I-A schools and LEAs that were identified as failing to meet the AYP standards. In some states, a substantial majority of Title I-A schools were identified as failing to make AYP, while in others almost no schools were so identified. In July 2002, just before the initial implementation of the new AYP provisions of NCLB, ED released a compilation of the number of schools identified as failing to meet AYP standards for two or more consecutive years (and therefore identified as being in need of improvement) in 2001-2002 (for most states) or 2000-2001 (in states where 2001-2002 data were not available).⁸ The national total number of these schools was 8,652; the number in individual states ranged from zero in Arkansas and Wyoming to 1,513 in Michigan and 1,009 in California.⁹ While there are obvious differences in the size of these states, there were also wide variations in the percentage of all schools participating in Title I-A that failed to meet AYP for either one year or two or more consecutive years.

AYP Under NCLB Statute

NCLB provisions regarding AYP may be seen as an evolution of, and to a substantial degree as a reaction to perceived weaknesses in, the AYP requirements of the 1994 IASA. The latter were frequently criticized as being insufficiently specific, detailed, or challenging. Criticism often focused specifically on their failure to focus on specific disadvantaged pupil groups, failure to require continuous improvement toward an ultimate goal, and their required applicability only to schools and LEAs participating in Title I-A, not to all public schools or to states overall.

Under NCLB, the Title I-A requirements for state-developed standards of AYP were substantially expanded in scope and specificity. As under the IASA, AYP is defined primarily on the basis of aggregate scores of pupils on state assessments of academic achievement. However, under NCLB, state AYP standards must also include at least one additional academic indicator, which in the case of high schools must be the graduation rate. The additional indicators may not be employed in a way that would reduce the number of schools or LEAs identified as failing to meet AYP standards.¹⁰

⁷ (...continued)

Service, *Evaluation of Title I Accountability Systems and School Improvement Efforts (TASSIE): First-Year Findings*, 2004. Hereafter referred to as the TASSIE First-Year Report.

⁸ See the U.S. Department of Education, "Paige Releases Number of Schools in School Improvement in Each State," press release, July 1, 2002 at [<http://www.ed.gov/news/pressreleases/2002/07/07012002a.html>].

⁹ Another report published by ED in 2004 (the TASSIE First-Year Report — see footnote 7) stated that 8,078 public schools had been identified as failing to meet AYP standards for two or more consecutive years in the 2001-2002 school year.

¹⁰ As is discussed later in this report and in more detail in a separate report (RL33032, *Adequate Yearly Progress (AYP): Growth Models Under the No Child Left Behind Act*), a (continued...)

One of the most important differences between AYP standards under NCLB and previous requirements is that under NCLB, AYP calculations must be disaggregated; that is, they must be determined separately and specifically for not only all pupils but also for several demographic groups of pupils within each school, LEA, and state. Test scores for an individual pupil may be taken into consideration multiple times, depending on the number of designated groups of which they are a member (e.g., a pupil might be considered as part of the LEP and economically disadvantaged groups, as well as the “all pupils” group). The specified demographic groups are as follows:

- economically disadvantaged pupils,
- LEP pupils,
- pupils with disabilities, and
- pupils in major racial and ethnic groups, as well as all pupils.

However, as is discussed further below, there are three major constraints on the consideration of these pupil groups in AYP calculations. First, pupil groups need not be considered in cases where their number is so relatively small that achievement results would not be statistically significant or the identity of individual pupils might be divulged.¹¹ As is discussed further below, the selection of the minimum number (n) of pupils in a group for the group to be considered in AYP determinations has been left largely to state discretion. State policies regarding “n” have varied widely, with important implications for the number of pupil groups actually considered in making AYP determinations for many schools and LEAs, and the number of schools or LEAs potentially identified as failing to make AYP. Second, it has been left to the states to define the “major racial and ethnic groups” on the basis of which AYP must be calculated. And third, as under the IASA, pupils who have not attended the same school for a full year need not be considered in determining AYP for the school, although they are still to be included in LEA and state AYP determinations.

In contrast to the previous statute, under which AYP standards had to be applied only to pupils, schools, and LEAs participating in Title I-A, AYP standards under NCLB must be applied to *all* public schools, LEAs, and for the first time to states overall, if a state chooses to receive Title I-A grants. However, corrective actions for failing to meet AYP standards need only be applied to schools and LEAs participating in Title I-A.

Another major break with the past is that state AYP standards must incorporate concrete movement toward meeting an ultimate goal of all pupils reaching a proficient or advanced level of achievement by the end of the 2013-2014 school year. The steps — that is, required levels of achievement — toward meeting this goal, known as Annual Measurable Objectives (AMOs), must increase in “equal increments” over time. The first increase in the thresholds must occur after no more

¹⁰ (...continued)
growth model pilot project has been initiated by ED.

¹¹ In addition, program regulations (*Federal Register*, December 2, 2002) do not require graduation rates and other additional academic indicators to be disaggregated in determining whether schools or LEAs meet AYP standards.

than two years, and remaining increases at least once every three years. As is discussed further below, several states have accommodated this requirement in ways that require much more rapid progress in the later years of the period leading up to 2013-2014 than in the earlier period.

The primary basic structure for AYP under NCLB is specified in the authorizing statute as a group status model. A “uniform bar” approach is employed: states are to set a threshold percentage of pupils at proficient or advanced levels each year that is applicable to all pupil subgroups of sufficient size to be considered in AYP determinations. The threshold levels of achievement are to be set separately for reading and math, and may be set separately for each level of K-12 education (elementary, middle, and high schools). The minimum¹² starting point for the “uniform bar” in the initial period is to be the greater of (a) the percentage of pupils at the proficient or advanced level of achievement for the lowest-achieving pupil group in the base year,¹³ or (b) the percentage of pupils at the proficient or advanced level of achievement for the lowest-performing quintile (fifth)¹⁴ of schools statewide in the base year.¹⁵ The “uniform bar” must generally be increased at least once every three years, although in the initial period it must be increased after no more than two years.

In determining whether scores for a group of pupils are at the required level, the averaging of scores over two to three years is allowed. In addition, NCLB includes a safe harbor provision, under which a school that does not meet the standard AYP requirements may still be deemed to meet AYP if it experiences a 10% (not a 10 percentage point) reduction in the gap between 100% and the base year for the specific pupil groups that fail to meet the “uniform bar,” and those pupil groups make progress on at least one other academic indicator included in the state’s AYP standards. As noted earlier, this alternative provision adds successive group improvement as a secondary AYP model under NCLB. In addition, as is discussed below, under a pilot project, nine states have been approved to use a third model of AYP — a “growth model” — for AYP determinations.

Finally, NCLB AYP provisions include an assessment participation rate requirement. In order for a school to meet AYP standards, at least 95% of all pupils, as well as at least 95% of each of the demographic groups of pupils considered for

¹² States may, of course, establish starting points above the required minimum level.

¹³ The “base year” is the 2001-2002 school year.

¹⁴ This is determined by ranking all public schools (of the relevant grade level) statewide according to their percentage of pupils at the proficient or higher level of achievement (on the basis of all pupils in each school), and setting the threshold at the point where one-fifth of the schools (weighted by enrollment) have been counted, starting with the schools at the lowest level of achievement.

¹⁵ Under program regulations [4 C.F.R. § 200.16(c)(2)], the starting point may vary by grade span (e.g., elementary, middle, etc.) and subject.

AYP determinations for the school or LEA, must participate in the assessments that serve as the primary basis for AYP determinations.¹⁶

ED Regulations and Guidance on Implementation of the AYP Provisions of NCLB

States began determining AYP for schools, LEAs, and the states overall on the basis of NCLB provisions beginning with the 2002-2003 school year. The deadline for states to submit to ED their AYP standards based on NCLB provisions was January 31, 2003, and according to ED, all states met this deadline. On June 10, 2003, ED announced that accountability plans had been approved for all states. However, many of the approved plans required states to take additional actions following submission of their plan.¹⁷

In the period preceding ED's review of state accountability plans under NCLB, the Department published two relevant documents. Regulations, published in the *Federal Register* on December 2, 2002, mirrored the detailed provisions in the authorizing statute. The second document, a policy letter published by the Secretary of Education on July 24, 2002,¹⁸ emphasized flexibility, stating that "The purpose of the statute, for both assessments and accountability, is to build on high quality accountability systems that States already have in place, not to require every state to start from scratch." The letter went on to list 10 criteria that it said would be applied by ED in the process of reviewing state AYP standards. These criteria included most, but not all, of the specifications regarding AYP from the authorizing statute and regulations (e.g., applicability to all public schools and their pupils, and specific focus on individual pupil groups). In response to concerns that large numbers of schools might be identified as failing to make AYP (as is discussed further below), ED officials emphasized the importance of taking action to identify and move to improve underperforming schools, no matter how numerous. They also emphasized the possibilities for flexibility and variation in taking corrective actions with respect to schools that fail to meet AYP, depending on the extent to which they fail to meet those standards.

Aspects of state AYP plans that apparently received special attention in ED's reviews included (1) the pace at which proficiency levels are expected to improve (e.g., equal increments of improvement over the entire period, or much more rapid improvement expected in later years than at the beginning); (2) whether schools or LEAs must fail to meet AYP with respect to the same pupil group(s), grade level(s), or subject areas to be identified as needing improvement, or whether two consecutive years of failure to meet AYP with respect to *any* of these categories should lead to

¹⁶ If the number of pupils in a specified demographic group is too small to meet the minimum group size requirements for consideration in AYP determinations, then the participation rate requirement does not apply.

¹⁷ The plans have been posted online by ED at [<http://www.ed.gov/admins/lead/account/stateplans03/index.html>].

¹⁸ See [<http://www.ed.gov/news/pressreleases/2002/07/07242002.html>].

identification;¹⁹ (3) the length of time over which pupils should be identified as being LEP; (4) the minimum size of pupil groups in a school in order for the group to be considered in AYP determinations or for reporting of scores; (5) whether to allow schools credit for raising pupil scores from below basic to basic (as well as from basic or below to proficient or above) in making AYP determinations; and (6) whether to allow use of statistical techniques such as “confidence intervals” (i.e., whether scores are below the required level to a statistically significant extent) in AYP determinations.

Recent Developments

Regulations Proposed in April 2008 on Title I-A Assessments and Accountability. Several new final regulations affecting the Title I-A assessment, AYP, and accountability policies were published in the “Federal Register” on October 29, 2008 (pages 64435-64513). Most of the proposed regulations deal with policy areas other than AYP. Many of the regulations clarify previous regulations or codify as regulations policies that have previously been established through less formal mechanisms (such as policy guidance or peer reviewer guidance). The proposed regulations related to AYP are briefly described below.

Group Size-Related Provisions in State AYP Policies. States must provide a more extensive rationale than previously required for their selection of minimum group sizes, use of confidence intervals, and related aspects of their AYP policies. Although no specific limits are placed on these parameters, states must explain in their Accountability Workbooks how their policies provide statistically reliable information while minimizing the exclusion of designated pupil groups in AYP determinations, especially at the school level. States must also report on the number of pupils in designated groups that are excluded from separate consideration in AYP determinations due to minimum group size policies. In addition, the regulations codify provisions for the National Technical Advisory Council that was established in August 2008 to advise the Secretary on a variety of technical aspects of state standards, assessments, AYP, and accountability policies. Each state is required to submit its Accountability Workbook, modified in accordance with the proposed regulations, to ED for a new round of technical assistance and peer review. Workbooks must be submitted in time to implement any needed changes before making AYP determinations based on assessment results for the 2009-2010 school year.

Assessments and Accountability Policies in General. The proposed regulations clarify that assessments required under Title I-A may include multiple formats as well as multiple academic assessments within each subject area (reading, mathematics, and science). This does not include the concept of “multiple measures,” as this term has been used by many to refer to proposals to expand NCLB

¹⁹ ED has approved state accountability plans under which schools or LEAs would be identified as failing to meet AYP only if they failed to meet the required level of performance in the same subject for two or more consecutive years, but has not approved proposals under which a school would be identified only if it failed to meet AYP in the same subject *and* pupil group for two or more consecutive years.

through inclusion of a variety of indicators *other than* standards-based assessments in reading, mathematics, and science. Also, states are required to include results from the most recent National Assessment of Educational Progress (NAEP) assessments on their state and LEA performance report cards. Further, ED policies regarding provisions for states to request waivers allowing them to use growth models of AYP are codified in the October 2008 regulations (previously they were published only in policy guidance and peer reviewer guidance documents.)

Graduation Rates. Numerous changes have been made to previous policies regarding graduation rates used as the “additional indicator” in AYP determinations for high schools. Previously, states were allowed a substantial degree of flexibility in their method for calculating graduation rates and were not required to disaggregate the rates by pupil group (except for reporting purposes). Also, although states were required to determine a level of, or rate of improvement in, graduation rates that would be adequate for AYP purposes, they were not required to set an ultimate goal toward which these rates should be progressing.

Under the October 2008 regulations, states must adopt a uniform method for calculating graduation rates. This method must be used for school, LEA, and state report cards showing results of assessments administered during the 2010-2011 school year, and for purposes of determining AYP based on assessments administered during the 2011-2012 school year (states unable to meet these deadlines may request an extension). This method has been endorsed by the National Governors Association. The graduation rate is defined as the number of students who graduate from high school in four years²⁰ divided by the number of students in the cohort for the students’ class, adjusted for student transfers among schools. States may also propose using a supplementary extended-year graduation rate, in addition to the four-year rate, in order to accommodate selected groups of students (such as certain students with disabilities) who may need more than four years to graduate. These graduation rates must be disaggregated by subgroup.

States must set an ultimate goal for graduation rates that they expect all high schools to meet. No federal standard is established, but the state goal, as well as annual targets toward meeting that goal, must be approved by ED as part of the state’s accountability policy.

Growth Models. In November 2005, the Secretary of Education announced a growth model pilot program under which initially up to 10 states would be allowed to use growth models to make AYP determinations for the 2005-2006 or subsequent school years.²¹ In December 2007, the Secretary lifted the cap on the number of states that could participate in the growth model pilot, and regulations published in October 2008²² incorporate this expanded policy. The models proposed by the states must meet at least the following criteria (in addition to a variety of criteria applicable

²⁰ This includes students who graduate following a summer program after their fourth year.

²¹ See [<http://www.ed.gov/news/pressreleases/2005/11/11182005.html>].

²² See the “Federal Register” for October 29, 2008 (pages 64435-64513).

to all state AYP policies — that is, measure achievement separately in reading/language arts and mathematics):

- they must incorporate an ultimate goal of all pupils reaching a proficient or higher level of achievement by the end of the 2013-2014 school year;
- achievement gaps among pupil groups must decline in order for schools or LEAs to meet AYP standards;
- annual achievement goals for pupils must not be set on the basis of pupil background or school characteristics;
- annual achievement goals must be based on performance standards, not past or “typical” performance growth rates;
- the assessment system must produce comparable results from grade-to-grade and year-to-year; and
- the progress of individual students must be tracked within a state data system.

In addition, applicant states must have their annual assessments for each of grades 3-8 approved by ED, and these assessments must have been in place for at least one year previous to implementation of the growth models.

In January 2006, ED published peer review guidance for growth model pilot applications.²³ In general, this guidance elaborates upon the requirements described above, with special emphasis on the following: (a) pupil growth targets may not consider their “race/ethnicity, socioeconomic status, school AYP status, or any other non-academic” factor; (b) growth targets are to be established on the basis of achievement standards, not typical growth patterns or past achievement; and (c) the state must have a longitudinal pupil data system, capable of tracking individual pupils as they move among schools and LEAs within the state.

The requirements for growth models of AYP under this pilot are relatively restrictive. The models must be consistent with the ultimate goal of all pupils at a proficient or higher level by 2013-2014, a major goal of the statutory AYP provisions of NCLB. More significantly, they must incorporate comparable annual assessments, at least for each of grades 3-8 plus at least one senior high school year, and those assessments must be approved by ED and in place for at least one year before implementation of the growth model. Further, all performance expectations must be individualized, and the state must have an infrastructure of a statewide, longitudinal database for individual pupils. Proposed models would have to be structured around expectations and performance of individual pupils, not demographic groups of pupils in a school or LEA, although individual results would have to be aggregated for the demographic groups designated in NCLB.

Two states, North Carolina and Tennessee, were initially approved to use proposed growth models in making AYP determinations on the basis of assessments administered in the 2005-2006 school year. Nine additional states — Arkansas, Delaware, Florida, Iowa, Ohio, Alaska, Arizona, Michigan, and Missouri — have

²³ See [<http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf>].

been approved to participate in the pilot program subsequently, contingent in the case of Missouri on adoption of a uniform minimum group size for all pupil groups. The growth models for these states are briefly described below.

The *North Carolina* policy does not actually provide for a separate AYP model, but rather the addition of a projection component to the current group status model. If the achievement level of a non-proficient pupil is on a trajectory toward proficiency within four years, then the pupil is added to the proficient group. All other provisions of the current group status and successive group improvement models would continue to apply. Thus, the ultimate goal becomes: by the end of the 2013-2014 school year, all pupils will be either at a proficient or higher level, or on a four-year trajectory toward proficiency (without use of confidence intervals). The trajectory calculations will be made for pupils in the 3rd through 8th grades. SEA staff estimate that 4% of the schools in North Carolina that failed to meet AYP standards on the basis of 2004-2005 assessment results would have met AYP standards if this growth model had been in place.

Under the *Tennessee* policy, schools and LEAs will have two options for meeting AYP: meeting either the AYP standards under the group status or successive group improvement models of current law, or meeting AYP standards according to a “projection model.” Under the projection model, pupils are deemed to be at a proficient or higher level of achievement if their test scores are projected to be at a proficient or higher level three years into the future, on the basis of past achievement levels for individual pupils. It should be noted that under this model, pupils who currently score at a proficient level, but who would be projected to score below a proficient level in three years, would not be counted as proficient. Further, the Tennessee growth/projection model implicitly assumes that pupils attend schools performing at a state average level. If, in actuality, they attend low-performing schools, their future achievement level may be overestimated.

Tennessee’s projection model will not be applied to high schools. SEA staff estimate that 13% of the schools in Tennessee that failed to meet AYP standards based on 2004-2005 assessment results would have met AYP standards if this model had been in place.

Under the *Delaware* growth model, AYP will be calculated each year on the basis of both the statutory provisions and using the state’s growth model. A school will meet AYP standards if it qualifies using either method. Individual pupil performance will be tracked from one year to the next. Specified numbers of points (up to 300) will be awarded on the basis of changes (if any) in pupils’ performance level. Points will be awarded for partial movement toward proficiency, but the points awarded for movement to advanced levels beyond proficiency will be the same as for movement to proficiency. (Maintaining a level of proficient or higher awards 300 points as well.) The average growth scores for schools and LEAs to meet AYP standards increase steadily until 2013-2014, by which time all pupils would be expected to achieve at a proficient or higher level.²⁴

²⁴ Delaware’s proposal included the use of confidence intervals at an unspecified level in (continued...)

Under the *Arkansas* policy, AYP will be calculated each year on the basis of both statutory provisions and using the state's growth model. A school will meet AYP standards if it qualifies using either method. Under the growth model, pupils in grades 4-8 will be deemed to be proficient if they are on a growth path toward proficiency by the end of 8th grade. Pupils already proficient must be on a path to continue to be proficient through grade 8 (i.e., growth path criteria will be applied to all pupils, proficient and non-proficient). Individual annual proficiency thresholds and growth increments are designed to enable non-proficient students to reach proficiency by grade 8, and proficient students to continue to be proficient. Mobile pupils will be associated with the school they attended at the time of assessment administration in the previous year.

Under the *Florida* model, AYP will be determined separately for each pupil subgroup in each school or LEA (i.e., not for schools or LEAs as a whole) using the statutory models (status and safe harbor) plus a growth model. The school or LEA will meet AYP standards if each pupil subgroup makes AYP using one of the three models.

Florida's growth model will be essentially the same as the current status model, except that proficient pupils will include both those currently scoring at a proficient or higher level plus those who are on an individual path toward proficiency within three years. The combined percentage of pupils rated proficient will be compared to the standard AMO. The model will be applied to AYP determinations for grades 3-10 (with some modifications for pupils in grade 3). In its application, the Florida SEA estimated that for 2006-2007, 938 of the state's public schools would meet AYP standards with the growth model applied, compared to 743 schools without (out of a total of 3,200 schools).

Under the *Iowa* model, pupil tests score ranges below proficient have been divided into 3 categories: Hi Marginal, Lo Marginal, and Weak. A student who rises from one of these levels to a higher level, and has not previously attained the higher level, will be deemed to have met "Adequate Yearly Growth" (AYG). AYG is considered to be more than a typical year's growth over a one-year period. For schools and LEAs that have not met AYP though application of the standard status and safe harbor models, students making AYG will be added to those scoring proficient or above, and this combined total will be used in determining whether the school or LEA makes AYP for the year. Students scoring below the proficient level must continue to move to a higher sub-proficient level each year in order to be included in the combined proficient + AYG student count. This implies that students beginning at the Weak level must reach proficiency within three years, those beginning at Lo Marginal must become proficient within two years, and those beginning at Hi Marginal must reach proficiency within one year. By 2014, the growth model would no longer be used, and all pupils will be expected to achieve at a proficient or higher level.

²⁴ (...continued)

implementing the growth model; however, ED approved use of the model without confidence intervals.

Confidence intervals will continue to be applied to determine whether the combined proficient + AYP student count meets the required threshold to make AYP. This growth model will be applied statewide to test scores for grades 3-8 and 11, and to grades 9 and 10 as well in the LEAs that administer the Iowa Tests in those grades. The Iowa growth model does not currently include students with the most significant cognitive disabilities, who take the Iowa Alternate Assessment.

Ohio has adopted a variation of the “projection” or “on track to proficiency” approach that is common to the models for all of the other participating states except Delaware and Iowa. After application of the standard status and safe harbor models, if any pupil group fails to meet AYP, then a determination will be made if a sufficient proportion of pupils in the group is on track toward meeting the required proficiency threshold as of a “target grade.” In the case of elementary and middle schools, the target grade will be either the grade level following the highest grade offered by the school (i.e., for a K-5 school, the 6th grade), or 4 grades beyond the pupil’s current grade, whichever comes first. In the case of a high school, pupils would have to be on track toward proficiency by the 11th grade.

Pupils currently scoring at a proficient level but who are projected to be below the proficient level by the target grade will not be considered to be proficient in *Ohio*’s projection model. Student achievement trajectories will be projected on an individual basis. Projections will be based on past test results (in all subjects, but with greater weight applied to past test results in the same subject) for each pupil.

Under *Alaska*’s growth model, pupils will be included in the proficient group if their achievement level trajectory is on a growth path toward proficiency within 3 additional years for pupils in grades 4-9, or within 2 additional years for pupils in grade 10. (Alaska currently has no standards-based assessments for grades beyond 10.) Pupils in the third grade (the earliest grade at which state assessments are administered) will be measured on the basis of status only, not growth. The growth model will not apply to pupils with disabilities who take alternate assessments. While Alaska had proposed that confidence intervals be applied, at a relatively low level (68%), under the growth model, the state agreed to drop this in the approved version. In its application, Alaska estimated that approximately 13% of pupils currently not proficient are on track toward proficiency, under the terms of the state’s growth model.

In *Arizona*, the growth model will be applicable to pupils in grades 4-8 only. Pupils will be included in the proficient group if their achievement level trajectory is on a growth path toward proficiency within three years or by 8th grade, whichever comes first. Pupils in the third grade (the earliest grade at which state assessments are administered) will be measured on the basis of status only, not growth. Unlike some other states participating in the growth model pilot, pupils with disabilities who take the state’s alternate assessment (AIMS-A) will be included in the Arizona growth model. Such pupils with disabilities who move up one performance level (i.e., from “falls far below” to “approaches” or from “approaches” to “meets” the proficiency standard) will be deemed to have met their growth target.

In *Missouri*, schools and LEAs will first be evaluated under the status model of AYP. If the school or LEA does not make AYP under that model, the growth model

will be applied. If the school or LEA still does not make AYP after application of the growth model, then a Safe Harbor calculation will be applied. If the school or LEA does not meet any of these 3 criteria, then it fails to make AYP.

In the growth calculation, it will be determined whether students currently scoring below a proficient level are on track to be proficient within either four years or by 8th grade, whichever occurs first. If so, they will be added to the number of students currently scoring at a proficient or higher level. Students in grades 3 and 8 will be evaluated on the basis of the status model and Safe Harbor only (grade 3 scores will be used as the baseline for growth trajectory calculations). No confidence intervals will be applied to growth model calculations. Only the current status and Safe Harbor models will be used for AYP determinations for grades 9-12. Students with disabilities, including those taking the state's alternate assessment for students with the most severe cognitive disabilities, will be included in the growth model, applying trajectories and achievement levels associated with either the regular or alternate assessments.

In *Michigan*, the approved growth model provides a third option for deeming student achievement to be proficient for purposes of AYP determinations. Currently, Michigan students are deemed to be proficient if their achievement test scores are at a proficient or advanced level, or if the scores of individual students are within 2 standard errors of measurement (in effect, a 95% confidence interval) of the test score cut point for proficiency.²⁵ The latter students are considered to be “provisionally proficient” and are treated the same as students scoring proficient or above in AYP determinations. The growth model adds a third category of students “on trajectory” toward proficiency.

To determine whether students are on trajectory toward proficiency, each of the four proficiency levels (not proficient/below basic, partially proficient/basic, proficient, and advanced) is divided into 3 sub-levels (low, middle, high). Similar, but slightly different, procedures are applied to Michigan's alternate assessment for students with mild cognitive impairment. The growth model does not cover high school students or students with disabilities taking alternate assessments who have moderate or severe cognitive impairment. If a student's performance improves over the previous year by a number of sub-levels such that, if the improvement continued at the same rate in the future, they would reach proficiency within three years, they are counted as being on trajectory toward proficiency. Confidence intervals will not be applied to the growth model determinations.

Thus, the number of students deemed proficient will be the total of students scoring proficient or above, plus students on trajectory to proficiency, plus students provisionally proficient. If this number of students divided by total students tested meets or exceeds the Annual Measurable Objective, then AYP is met with respect to the subject and student group in question. Since many students may meet both the trajectory toward proficiency and the provisionally proficient criteria, it will first be

²⁵ Most states use confidence intervals in their AYP determinations. However, in most cases, the confidence intervals are applied to group average percentages of students scoring proficient or above, not to individual student scores.

determined whether students are on trajectory, then whether any remaining non-proficient students meet the provisionally proficient criterion. It is estimated that use of the growth model will add only minimally (0.7-1.3%) to the number of students already deemed to be proficient or provisionally proficient.

Overall, most of the growth models approved by ED thus far are based upon supplementing the number of pupils scoring at a proficient or higher level with those who are projected to be at a proficient level within a limited number of years. Nine of the eleven approved models follow this general approach. Among these states, a distinction may be made between seven states (North Carolina, Arkansas, Florida, Alaska, Arizona, Missouri, and Michigan) that combine currently proficient pupils with those not proficient who are “on track” toward proficiency, and two states (Tennessee and Ohio) that consider only projected proficiency levels for all pupils (i.e., currently proficient pupils who are not on track to remain proficient are counted as not proficient). In contrast, the models used by two other states — Delaware and Iowa — focus on awarding credit for movement of pupils among achievement categories up to proficiency.

Pupils with Disabilities. The most substantial of ED’s recent AYP policy changes involves pupils with disabilities. First, regulations addressing the application of the Title I-A standards and assessment requirements to certain pupils with disabilities were published in the *Federal Register* on December 9, 2003 (pp. 68698-68708). The purpose of these regulations is to clarify the application of standard, assessment, and accountability provisions to pupils “with the most significant cognitive disabilities.” Under the regulations, states and LEAs may adopt alternate assessments based on alternate achievement standards — aligned with the state’s academic content standards and reflecting “professional judgment of the highest achievement standards possible” — for a limited percentage of pupils with disabilities.²⁶ The number of pupils whose proficient or higher scores on these alternate assessments may be considered as proficient or above for AYP purposes is limited to a maximum of 1.0% of all tested pupils (approximately 9% of all pupils with disabilities) at the state and LEA level (there is no limit for individual schools). SEAs may request from the U.S. Secretary of Education an exception allowing them to exceed the 1.0% cap statewide, and SEAs may grant such exceptions to LEAs within their state. According to ED staff, three states in 2003-2004 (Montana, Ohio, and Virginia), and four states in 2004-2005 (the preceding three states plus South Dakota), received waivers to go marginally above the 1.0% limit statewide. In the absence of a waiver, the number of pupils scoring at the “proficient or higher” level on alternate assessments, based on alternate achievement standards, in excess of the 1.0% limit is to be added to those scoring “below proficient” in LEA or state-level AYP determinations.

A new ED policy affecting an additional group of pupils with disabilities was announced initially in April 2005, with final regulations based on it published in the *Federal Register* on April 9, 2007. The new policy is divided into short-term and

²⁶ This limitation does *not* apply to the administration of alternate assessments based on the same standards applicable to all students, for other pupils with (non-cognitive or less severe cognitive) disabilities.

long-term phases. It is focused on pupils with disabilities whose ability to perform academically is assumed to be greater than that of the pupils with “the most significant cognitive disabilities” discussed in the above paragraph, and who are capable of achieving high standards, but may not reach grade level within the same time period as their peers. In ED’s terminology, these pupils would be assessed using alternate assessments based on modified achievement standards.

The short-term policy may apply, with the approval of the Secretary, to states until they develop and administer alternative assessments under the long-term policy (described below).²⁷ Under this short-term policy, in eligible states that have not yet adopted modified achievement standards, schools may add to their proficient pupil group a number of pupils with disabilities equal to 2.0% of all pupils assessed (in effect, deeming the scores of all of these pupils to be at the proficient level).²⁸ This policy would be applicable only to schools and LEAs that would otherwise fail meet AYP standards due solely to their pupils with disabilities group. According to ED staff, as of the date of this report, 28 states are currently exercising this flexibility. Alternatively, in eligible states that have adopted modified achievement standards (currently six states), schools and LEAs may count proficient scores for pupils with disabilities on these assessments, subject to a 2.0% (of all assessed pupils) cap at the LEA and state levels.

The long-term policy is embodied in final regulations published in the *Federal Register* on April 9, 2007. These regulations affect standards, assessments, and AYP for a group of pupils with disabilities who are unlikely to achieve grade level proficiency within the current school year, but who are not among those pupils with the most significant cognitive disabilities (whose situation was addressed by an earlier set of regulations, discussed above). For this second group of pupils with disabilities, states would be authorized to develop “modified academic achievement standards” and alternate assessments linked to these. The modified achievement standards must be aligned with grade-level content standards, but may reflect reduced breadth or depth of grade-level content in comparison to the achievement standards applicable to the majority of pupils. The standards must provide access to grade-level curriculum, and not preclude affected pupils from earning a regular high school diploma.

As with the previous regulations regarding pupils with the most significant cognitive disabilities, there would be no direct limit on the number of pupils who take alternate assessments based on modified achievement standards. However, in AYP determinations, pupil scores of proficient or advanced on alternate assessments based on modified achievement standards may be counted only as long as they do not exceed a number equal to 2.0% of all pupils tested at the state or LEA level (i.e., an

²⁷ Under current regulations, the short-term policy cannot be extended beyond the 2008-2009 school year.

²⁸ This would be calculated on the basis of statewide demographic data, with the resulting percentage applied to each affected school and LEA in the state. In making the AYP determination using the adjusted data, no further use may be made of confidence intervals or other statistical techniques. (The actual, not just the adjusted, percentage of pupils who are proficient must also be reported to parents and the public.)

estimated 20% of pupils with disabilities); such scores in excess of the limit would be considered “non-proficient.” As with the 1.0% cap for pupils with the most significant cognitive disabilities, this 2.0% cap does not apply to individual schools. In general, LEAs or states could exceed the 2.0% cap only if they did not reach the 1.0% limit with respect to pupils with the most significant cognitive disabilities. Thus, in general, scores of proficient or above on alternate assessments based on alternate *and* modified achievement standards may not exceed a total of 3.0% of all pupils tested at a state or LEA level.²⁹ In particular, states are no longer allowed to request a waiver of the 1.0% cap regarding pupils with the most significant cognitive disabilities.

The April 9, 2007, proposed regulations also include provisions that are widely applicable to AYP determinations. First, states are no longer allowed to use varying minimum group sizes (“n”) for different demographic groups of pupils. This prohibits the previously common practice of setting higher “n” sizes for pupils with disabilities or LEP pupils than for other pupil groups. Second, when pupils take state assessments multiple times, states and LEAs may use the highest score for pupils who take tests more than once. Finally, as with LEP pupils, states and LEAs may include the test scores of former pupils with disabilities in the disability subgroup for up to two years after such pupils have exited special education.³⁰

In summary, there are now five groups of pupils with disabilities with respect to achievement standards, assessments, and the use of scores in AYP determinations. These groups are summarized below in **Table 1**.

²⁹ The 3.0% limit might be exceeded for LEAs, but only if — and to the extent that — the SEA waives the 1.0% cap applicable to scores on alternate assessments based on alternate achievement standards.

³⁰ In such cases, the former pupils with disabilities would *not* have to be counted in determining whether the minimum group size was met for the disability subgroup.

**Table 1. Categories of Pupils with Disabilities
with Respect to Achievement Standards, Assessments,
and AYP Determinations Under ESEA Title I-A**

Type of Content Standards	Type of Achievement Standards	Type of Assessment	Cap on # of Proficient or Advanced Scores That May Be Included in AYP Determinations
Grade-level content standards	Grade-level academic achievement standards	Regular (i.e., the same as that applicable to pupils generally)	None
Grade-level content standards	Grade-level academic achievement standards	Regular with accommodations (e.g., special assistance for those with sight or hearing disabilities)	None
Grade-level content standards	Grade-level academic achievement standards	Alternate assessments based on regular, grade-level achievement standards (e.g., portfolios or performance assessments)	None
Grade-level content standards	Modified academic achievement standards	Alternate assessments based on modified academic achievement standards	In general, 2.0% of all pupils assessed
Alternate content standards	Alternate academic achievement standards	Alternate assessments based on alternate achievement standards	In general, 1.0% of all pupils assessed

Participation Rates. On March 29, 2004, ED announced that schools could meet the requirement that 95% or more of pupils (all pupils as well as pupils in each designated demographic group) participate in assessments (in order for the school or LEA to make AYP) on the basis of average participation rates for the last two or three years, rather than having to post a 95% or higher participation rate each year. In other words, if a particular demographic group of pupils in a public school has a 93% test participation rate in the most recent year, but had a 97% rate the preceding year, the 95% participation rate requirement would be met. In addition, the new guidance would allow schools to exclude pupils who fail to participate in assessments due to a “significant medical emergency” from the participation rate calculations. The new guidance further emphasizes the authority for states to allow pupils who miss a primary assessment date to take make-up tests, and to establish a minimum size for demographic groups of pupils to be considered in making AYP determinations (including those related to participation rates). According to ED, in some states, as many as 20% of the schools failing to make AYP did so on the basis of assessment participation rates alone. It is not known how many of these schools would meet the new, somewhat more relaxed standard.

LEP Pupils. In a letter dated February 19, and proposed regulations published on June 24, 2004, ED officials announced two new policies with respect to LEP pupils.³¹ First, with respect to assessments, LEP pupils who have attended schools in the United States (other than Puerto Rico) for less than 10 months must participate in English language proficiency and mathematics tests. However, the participation of such pupils in reading tests (in English), as well as the inclusion of any of these pupils’ test scores in AYP calculations, is to be optional (i.e., schools and LEAs need not consider the scores of first year LEP pupils in determining whether schools or LEAs meet AYP standards). Such pupils are still considered in determining whether the 95% test participation has been met.

Second, in AYP determinations, schools and LEAs may continue to include pupils in the LEP demographic category for up to two years after they have attained proficiency in English. However, these formerly LEP pupils need not be included when determining whether a school or LEA’s count of LEP pupils meets the state’s minimum size threshold for inclusion of the group in AYP calculations, and scores of formerly LEP pupils may not be included in state, LEA, or school report cards. Both these options, if exercised, should increase average test scores for pupils categorized as being part of the LEP group, and reduce the extent to which schools or LEAs fail to meet AYP on the basis of LEP pupil groups.

AYP Determinations for Targeted Assistance Schools. ED has released a February 4, 2004, letter to a state superintendent of education providing more flexibility in AYP determinations for targeted assistance schools.³² Title I-A services are provided at the school level via one of two basic models: targeted assistance schools, where services are focused on individual pupils with the lowest levels of academic achievement, or schoolwide programs, in which Title I-A funds

³¹ See *Federal Register*, June 24, 2004, pp. 35462-35465; and [<http://www.ed.gov/nclb/accountability/schools/factsheet-english.html>].

³² See [<http://www.ed.gov/policy/elsec/guid/stateletters/asaypnc.html>].

may be used to improve academic instruction for all pupils. Currently, most Title I-A programs are in targeted assistance schools, although the number of schoolwide programs has grown rapidly in recent years, and most pupils served by Title I-A are in schoolwide programs.

This policy letter gives schools and LEAs the option of considering only pupils assisted by Title I-A for purposes of making AYP determinations for individual schools. LEA and state level AYP determinations would still have to be made on the basis of all public school pupils. The impact of this authority, if utilized, is unclear. In schools using this authority, there would be an increased likelihood that pupil demographic groups would be below minimum size to be considered. At the same time, if Title I-A participants are indeed the lowest-performing pupils in targeted assistance schools, it seems unlikely that many schools would choose to base AYP determinations only on those pupils, especially given the current structure of the primary AYP requirements under NCLB (i.e., a status model, not a growth model).

Flexibility for Areas Affected by the Gulf Coast Hurricanes.

Following the damage to school systems and dispersion of pupils in the wake of Hurricanes Katrina and Rita in August and September 2005, interest has been expressed by officials of states and LEAs that were damaged by the storms, or that enrolled pupils displaced by these storms, in the possibility of waiving some of NCLB's assessment, AYP, or other accountability requirements. In a series of policy letters to chief state school officers (CSSOs), the Secretary of Education has emphasized forms of flexibility already available under current law and announced a number of policy revisions and potential waivers that might be granted in the future.

In a September 29, 2005, letter to all CSSOs,³³ the Secretary of Education noted that they could exercise existing natural disaster provisions of NCLB [§1116(b)(7)(D) and (c)(10)(F)] to postpone the implementation of school or LEA improvement designations and corrective actions for schools or LEAs failing to meet AYP standards that are located in the major disaster areas in Louisiana, Alabama, Mississippi, Texas, or Florida, without a specific waiver being required. In addition, waivers of these requirements will be considered for other LEAs or schools heavily affected by enrolling large numbers of evacuee pupils. Further, all affected LEAs and schools could establish a separate subgroup for displaced students in AYP determinations on the basis of assessments administered during the 2005-2006 school year. Pupils would appear only in the evacuee subgroup, not other demographic subgroups (e.g., economically disadvantaged or LEP). Waivers could be requested in 2006 to allow schools or LEAs to meet AYP requirements if only the test scores of the evacuee subgroup would prevent them from making AYP. In any case, all such students must still be assessed and the assessment results reported to the public.³⁴

³³ See [<http://www.ed.gov/policy/elsec/guid/secletter/050929.html>].

³⁴ For additional information on this topic, see CRS Report RL33236, *Education-Related Hurricane Relief: Legislative Act*, by Rebecca Skinner, et al.

State Revisions of Their Accountability Plans. Over the period following the initial submission and approval of state accountability plans for AYP and related policies in 2003 through the present, many states have proposed a number of revisions to their plans. Sometimes these revisions seem clearly intended to take advantage of new forms of flexibility announced by ED officials, such as those discussed above, while in other cases states appear to be attempting to take advantage of options or forms of flexibility that reportedly been approved for other states previously.

The proposed changes in state accountability plans have apparently almost always been in the direction of increased flexibility for states and LEAs, with reductions anticipated in the number or percentage of schools or LEAs identified as failing to make AYP. Issues that have arisen with respect to these changes include a lack of transparency, and possibly inconsistencies (especially over time), in the types of changes that ED officials have approved; debates over whether the net effect of the changes is to make the accountability requirements more reasonable or to undesirably weaken them; concern that the changes may make an already complicated accountability system even more complex; and timing — whether decisions on proposed changes are being made in a timely manner by ED.

The major aspects of state accountability plans for which changes have been proposed and approved include the following: (a) changes to take advantage of revised federal regulations and policy guidance regarding assessment of pupils with the most significant cognitive disabilities, LEP pupils, and test participation rates; (b) limiting identification for improvement to schools that fail to meet AYP in the same subject area for two or more consecutive years, and limiting identification of LEAs for improvement to those that failed to meet AYP in the same subject area and across all three grade spans for two or more consecutive years; (c) using alternative methods to determine AYP for schools with very low enrollment; (d) initiating or expanding use of confidence intervals in AYP determinations, including “safe harbor” calculations; (e) changing (usually effectively increasing) minimum group size; and (f) changing graduation rate targets for high schools. Accountability plan changes that have frequently been requested but not approved by ED include (a) identification of schools for improvement only if they failed to meet AYP with respect to the same pupil group *and* subject area for two or more consecutive years, and (b) retroactive application of new forms of flexibility to recalculation of AYP for previous years.³⁵

³⁵ See Center on Education Policy, *Rule Changes Could Help More Schools Meet Test Score Targets for the No Child Left Behind Act*, October 22, 2004, available at [<http://www.cep-dc.org/nclb/StateAccountabilityPlanAmendmentsReportOct2004.pdf>]; Title I Monitor, *Changes in Accountability Plans Dilute Standards, Critics Say*, November 2004; Council of Chief State School Officers, *Revisiting Statewide Educational Accountability Under NCLB*, September 2004, available at [<http://www.ccsso.org>]; and “Requests Win More Leeway Under NCLB,” *Education Week*, July 13, 2005, p. 1.

Data on Schools and LEAs Identified as Failing to Meet AYP

A substantial amount of data has become available on the number of schools and LEAs that have failed to meet the AYP standards of the NCLB on the basis of assessments administered during the 2002-2003 through 2005-2006 school years, and several states are currently releasing preliminary data based on 2006-2007 school year assessment results. A basic problem with these data is that they frequently have been incomplete and subject to change. Currently available compilations of state AYP data are discussed below in two categories: reports focusing on the number and percentage of schools failing to meet AYP standards for one or more years versus reports on the number and percentage of public schools and LEAs identified for improvement — that is, they had failed to meet AYP standards for at least two consecutive years.

Schools Failing to Meet AYP Standards for One or More Years

Beginning with the 2002-2003 school year, data on the number of schools in each state that made or did not make AYP have been reported by the states to ED, in a series of Consolidated State Performance Reports. Until recently, these Reports were not disseminated by ED; however, the Consolidated State Performance Reports for the 2004-2005 and 2005-2006 school years have been made available by ED.³⁶

According to these Consolidated State Performance Reports,³⁷ for the nation overall, 28% of all public schools failed to make adequate yearly progress on the basis of assessment scores for the 2006-2007 school year. The percentage of public schools failing to make adequate yearly progress for 2006-2007 varied widely among the states, from 4% for Wisconsin and 6% for Wyoming to 75% for the District of Columbia and 66% for Florida. **Table 2** provides the percentage of schools failing to make adequately yearly progress, on the basis of 2006-2007 assessment results, for each state.

According to the “National Assessment of Title I: Final Report,” published by ED in October 2007, of schools failing to make AYP in the 2004-2005 school year, 43% did so with respect to achievement in reading or math (or both) for the “all pupils” group. In contrast, 40% of schools failing to make AYP did so on the basis of achievement in reading or math (or both) for one or more subgroups while making AYP with respect to achievement of the “all pupils” group. The remaining 17% of schools failing to make AYP that year did so with respect to test participation rates only (3%), “other academic indicator” only (4%), or other combinations of AYP criteria (10%). Among schools with numbers of pupils in each of the designated categories to meet the minimum group size criterion for their state, the percentage of schools failing to make AYP with respect to math or reading achievement in 2004-

³⁶ See [<http://www.ed.gov/admins/lead/account/consolidated/index.html>].

³⁷ For one state, Maine, these data were not available in the Consolidated State Performance Report, and were obtained directly from the state educational agency.

2005 was found to vary from 3% for the Asian or White pupil groups, 18% for Hispanic pupils, 23% for pupils from low-income families, 24% for LEP pupils, 26% for African-American pupils, and 38% for pupils with disabilities.

Schools Failing to Meet AYP Standards for Two Consecutive Years or More

ED, in its “National Assessment of Title I: Final Report,” published in October 2007, reported that 11,648 public schools, including 9,808 Title I-A schools, were identified for improvement during the 2005-2006 school year, based on assessment results through the 2004-2005 school year. These constituted 12% of all public schools or 18% of all Title I-A schools. Schools most likely to be identified were those in large, urban LEAs, schools with high pupil poverty rates, and schools with large minority enrollment. The percentage of both all and of Title I-A schools identified varied widely among the states, from less than 1% (of all)/1% (of Title I-A) schools in Nebraska to more than 40% of all schools in Hawaii, New Mexico, and Puerto Rico, or more than 50% of all Title I-A schools in Florida, New Mexico, and Puerto Rico.

LEAs Failing to Meet AYP Standards

Although most attention, in both the statute and implementation activities, thus far has focused on application of the AYP concept to schools, a limited amount of information is becoming available about LEAs that fail to meet AYP requirements, and the consequences for them. According to the Consolidated State Performance Reports referred to above, approximately 30% of all LEAs failed to meet AYP standards on the basis of assessment results for the 2006-2007 school year (see **Table 2**). Among the states, there was even greater variation for LEAs than for schools. Three states — Alabama, Wisconsin, and Wyoming — reported that 1% or less of their LEAs failed to make adequate yearly progress, while 97% of the LEAs in North Carolina and 91% of those in West Virginia failed to meet AYP standards.

In its “National Assessment of Title I: Final Report,” ED has reported that 1,578 LEAs, representing approximately 10% of all LEAs, were identified for improvement for the 2005-2006 school year. A large number of states have recently adopted policies under which LEAs would be identified as needing improvement only if they failed to make AYP in the same subject (reading or mathematics) in each of three grade levels (elementary, middle, and high) for two or more consecutive school years. According to a recent study of NCLB implementation in six states by the Harvard Civil Rights Project, this has substantially increased the proportion of LEAs identified for improvement that serve central city areas and racially diverse or high-poverty pupil populations.³⁸

³⁸ Harvard Civil Rights Project, “Changing NCLB Accountability Standards: Implications for Racial Equity,” June 2005, available at [<http://www.civilrightsproject.harvard.edu>].

Table 2. Reported Percentage of Public Schools and Local Educational Agencies (LEAs) Failing to Make Adequate Yearly Progress (AYP) on the Basis of Spring 2007 Assessment Results

State	Reported Percentage of Rated Schools Not Making AYP, 2007	Reported Percentage of LEAs Not Making AYP, 2007
Alabama	16	1
Alaska	34	54
Arizona	28	42
Arkansas	38	18
California	33	47
Colorado	27	43
Connecticut	32	19
Delaware	30	32
District of Columbia	75	84
Florida	66	na ^a
Georgia	18	61
Hawaii	35	na ^a
Idaho	73	73
Illinois	24	28
Indiana	48	21
Iowa	7	2
Kansas	12	12
Kentucky	22	47
Louisiana	12	na ^a
Maine	30	5
Maryland	23	71
Massachusetts	48	70
Michigan	18	3
Minnesota	38	47
Mississippi	21	69
Missouri	46	63
Montana	10	15
Nebraska	12	21
Nevada	33	6
New Hampshire	42	31
New Jersey	26	7
New Mexico	55	74
New York	20	27
North Carolina	55	97
North Dakota	9	14
Ohio	38	70
Oklahoma	12	14
Oregon	22	52
Pennsylvania	23	9
Rhode Island	21	33
South Carolina	63	na ^a
South Dakota	18	3
Tennessee	13	10

State	Reported Percentage of Rated Schools Not Making AYP, 2007	Reported Percentage of LEAs Not Making AYP, 2007
Texas	9	11
Utah	23	17
Vermont	12	17
Virginia	26	55
Washington	35	50
West Virginia	19	91
Wisconsin ^b	4	0
Wyoming	6	10
Puerto Rico	47	na ^a
National Average	28	30^a

Source: State Consolidated Performance Reports [http://www.ed.gov/admins/lead/account/consolidated/sy06-07/index.html].

a. NA = Not available. Thus, the national total percentage for LEAs excludes these.

b. Wisconsin reports 2 LEAs as failing to make AYP out of a total of 425 LEAs.

Issues in State Implementation of NCLB Provisions

Introduction

The primary challenge associated with the AYP concept is to develop and implement school, LEA, and state performance measures that are: (a) challenging, (b) provide meaningful incentives to work toward continuous improvement, (c) are at least minimally consistent across LEAs and states, and (d) focus attention especially on disadvantaged pupil groups. At the same time, it is generally deemed desirable that AYP standards should allow flexibility to accommodate myriad variations in state and local conditions, demographics, and policies, and avoid the identification of so many schools and LEAs as failing to meet the standards that morale declines significantly systemwide and it becomes extremely difficult to target technical assistance and corrective actions on low-performing schools. The AYP provisions of NCLB are challenging and complex, and have generated substantial criticism from several states, LEAs, and interest groups. Many critics are especially concerned that efforts to direct resources and apply corrective actions to low-performing schools would likely be ineffective if resources and attention are dispersed among a relatively large proportion of public schools. Others defend NCLB's requirements as being a measured response to the weaknesses of the pre-NCLB AYP provisions, which were much more flexible but, as discussed above, had several weaknesses.

The remainder of this report provides a discussion and analysis of several specific aspects of NCLB's AYP provisions that have attracted significant attention and debate. These include the provision for an ultimate goal, use of confidence intervals and data-averaging, population diversity effects, minimum pupil group size (n), separate focus on specific pupil groups, number of schools identified and state

variations therein, the 95% participation rule, state variations in assessments and proficiency standards, and timing.

It should be noted that this report focuses on issues that have arisen in the implementation of NCLB provisions on AYP. As such, it generally does not focus on alternatives to the current statutory provisions of NCLB.

Ultimate Goal

The required incorporation of an ultimate goal — of all pupils at a proficient or higher level of achievement within 12 years of enactment — is one of the most significant differences between the AYP provisions of NCLB and those under previous legislation. Setting such a date is perhaps the primary mechanism requiring state AYP standards to incorporate annual increases in expected achievement levels, as opposed to the relatively static expectations embodied in most state AYP standards under the previous IASA. Without an ultimate goal of having all pupils reach the proficient level of achievement by a specific date, states might simply establish relative goals (e.g., performance must be as high as the state average) that provide no real movement toward, or incentives for, significant improvement, especially among disadvantaged pupil groups.

Nevertheless, a goal of having all pupils at a proficient or higher level of achievement, within 12 years or any other specified period of time, may be easily criticized as being “unrealistic,” if one assumes that “proficiency” has been established at a challenging level. Proponents of such a demanding ultimate goal argue that schools and LEAs frequently meet the goals established for them, even rather challenging goals, if the goals are very clearly identified, defined, and established, if they are attainable, and if it is made visibly clear that they will be expected to meet them. This is in contrast to a pre-NCLB system under which performance goals were often vague, undemanding, and poorly communicated, with few, if any, consequences for failing to meet them. A demanding goal might maximize efforts toward improvement by state public school systems, even if the goal is not met. Further, if a less ambitious goal were to be adopted, what lower level of pupil performance might be acceptable, and for which pupils?

At the same time, by setting deadlines by which all pupils must achieve at the proficient or higher level, the AYP provisions of NCLB create an incentive for states to weaken their pupil performance standards to make them easier to meet. In many states, only a minority of pupils (sometimes a small minority) are currently achieving at the proficient or higher level on state reading and mathematics assessments. Even in states where the percentage of all pupils scoring at the proficient or higher level is substantially higher, the percentage of those in many of the pupil groups identified under NCLB’s AYP provisions is substantially lower. It would be extremely difficult for such states to reach a goal of 100% of their pupils at the proficient level, even within 10-12 years, without reducing their performance standards.

There has thus far been some apparent movement toward lowering proficiency standards in a small number of states. Reportedly, a few states have redesignated lower standards (e.g., “basic” or “partially proficient”) as constituting a “proficient” level of performance for Title I-A purposes, or established new “proficient” levels

of performance that are below levels previously understood to constitute that level of performance, and other states have considered such actions.³⁹ For example, in submitting its accountability plan (which was approved by ED), Colorado stated that it would deem students performing at both its “proficient” and “partially proficient” levels, as defined by that state, as being “proficient” for NCLB purposes.⁴⁰ In its submission, the state argued that “Colorado’s standards for all students remain high in comparison to most states. Colorado’s basic proficiency level on CSAP is also high in comparison to most states.” Similarly, Louisiana decided to identify its “basic” level of achievement as the “proficient” level for NCLB purposes, stating that “[t]hese standards have been shown to be high; for example, equipercentile equating of the standards has shown that Louisiana’s ‘Basic’ is somewhat more rigorous than NAEP’s ‘Basic.’ In addition, representatives from Louisiana’s business community and higher education have validated the use of ‘Basic’ as the state’s proficiency goal.”⁴¹

This is an aspect of NCLB’s AYP provisions on which there will likely be continuing debate. It is unlikely that any state, and few schools or LEAs of substantial size and a heterogeneous pupil population, will meet NCLB’s ultimate AYP goal, unless state standards of proficient performance are significantly lowered or states aggressively pursue the use of such statistical techniques as setting high minimum group sizes and confidence intervals (described below) to substantially reduce the range of pupil groups considered in AYP determinations or effectively lower required achievement level thresholds.

Some states have addressed this situation, at least in the short run, by “backloading” their AYP standards, requiring much more rapid improvements in performance at the end of the 12-year period than at the beginning. These states have followed the letter of the statutory language that requires increases of “equal increments” in levels of performance after the first two years, and at least once every three years thereafter.⁴² However, they have “backloaded” this process by, for example, requiring increases only once every two-three years at the beginning, then requiring increases of the same degree every year for the final years of the period leading up to 2013-2014. For example, both Indiana and Ohio established incremental increases in the threshold level of performance for schools and LEAs that are equal in size, and that are to take effect in the school years beginning in 2004, 2007, 2010, 2011, 2012, and 2013. As a result, the required increases per year are three times greater during 2010-2013 than in the 2004-2009 period. These states may

³⁹ See, for example, “States Revise the Meaning of ‘Proficient’,” *Education Week*, October 9, 2002.

⁴⁰ See [<http://www.ed.gov/admins/lead/account/stateplans03/cocsa.pdf>], p. 7.

⁴¹ See [<http://www.ed.gov/admins/lead/account/stateplans03/lacsa.doc>], p 12.

⁴² According to Section 1111(b)(2)(H), “Each State shall establish intermediate goals for meeting the requirements, ... of this paragraph and that shall — (i) increase in equal increments over the period covered by the State’s timeline....” The program regulations also would seem to require increases in equal increments: “Each State must establish intermediate goals that increase in equal increments over the period covered by the timeline....” (34 C.F.R. § 200.17).

be trying to postpone required increases in performance levels until NCLB provisions are reconsidered, and possibly revised, by Congress.

Confidence Intervals and Data-Averaging

Many states have used one or both of a pair of statistical techniques to attempt to improve the validity and reliability of AYP determinations. Use of these techniques also tends to have an effect, whether intentional or not, of reducing the number of schools or LEAs identified as failing to meet AYP standards.

The averaging of test score results for various pupil groups over two- or three-year periods is explicitly authorized under NCLB, and this authority is used by many states. In some cases, schools or LEAs are allowed to select whether to average test score data, and for what period (two years or three), whichever is most favorable for them. As discussed above, recent policy guidance also explicitly allows the use of averaging for participation rates.

The use of another statistical technique was not explicitly envisioned in the drafting of NCLB's AYP provisions, but its inclusion in the accountability plans of several states has been approved by ED. This is the use of "confidence intervals," usually with respect to test scores, but in a couple of states also to the determination of minimum group size (see below). This concept is based on the assumption that any test administration represents a "sample survey" of pupils' educational achievement level. As with all sample surveys, there is a degree of uncertainty regarding how well the sample results — average test scores for the pupil group — reflect pupils' actual level of achievement. As with surveys, the larger the number of pupils in the group being tested, the greater the probability that the group's average test score will represent their true level of achievement, all else being equal. Put another way, confidence intervals are used to evaluate whether achievement scores are below the required threshold to a statistically significant extent.

"Confidence intervals" may be seen as "windows" surrounding a threshold test score level (i.e., the percentage of pupils at the proficient or higher level required under the state's AYP standards).⁴³ The size of the window varies with respect to the number of pupils in the relevant group who are tested, and with the desired degree of probability that the group's average score represents their true level of achievement. This is analogous to the "margin of error" commonly reported along with opinion polls. While test results are not based on a small sample of the relevant population, as are opinion poll results, since the tests are to be administered to the full "universe" of pupils, the results from any particular test administration are considered to be only estimates of pupils' true level of achievement, or of the effectiveness of a school or LEA in educating specified pupil groups, and thus the "margin of error" or "confidence interval" concepts are deemed by many to be relevant to these test scores. The probability, or level of confidence, is most often set at 95%, but in some cases may be as low as 90% or as high as 99% — that is, it is

⁴³ Alternatively, the confidence interval "window" may be applied to average test scores for each relevant pupil group, that would be compared to a fixed threshold score level to determine whether AYP has been met.

95% (or 90% or 99%) certain that the true achievement level for a group of pupils is within the relevant confidence interval of test scores above and below the average score for the group. All other relevant factors being equal, the smaller the pupil group, and the higher the desired degree of probability, the larger is the window surrounding the threshold percentage.

For example, consider a situation where the threshold percentage of pupils at the proficient or higher level of achievement in reading for elementary schools required under a state's AYP standards is 40%. Without applying confidence intervals, a school would simply fail to make AYP if the average scores of all of its pupils, or of any of its relevant pupil groups meeting minimum size thresholds, is below 40%. In contrast, if confidence intervals are applied, windows are established above and below the 40% threshold, turning the threshold from a single point to a variable range of scores. The size of this score range or window will vary depending on the size of the pupil group whose average scores are being considered, and the desired degree of probability (95% or 99%) that the average achievement levels for pupils in each group are being correctly categorized as being “truly” below the required threshold. In this case, a school would fail to make AYP with respect to a pupil group only if the average score for the group is below the lowest score in that range.⁴⁴

The use of confidence intervals to determine whether group test scores fall below required thresholds to a statistically significant degree improves the validity of AYP determinations, and addresses the fact that test scores for any group of pupils will vary from one test administration to another, and these variations may be especially large for a relatively small group of pupils. At the same time, the use of confidence intervals reduces the likelihood that schools or (to a lesser extent) LEAs will be identified as failing to make AYP. Also, for relatively small pupil groups and high levels of desired accuracy (especially a 99% probability), the size of confidence intervals may be relatively large. Ultimately, the use of this technique may mean that the average achievement levels of pupil groups in many schools will be well below 100% proficiency by 2013-2014, yet the schools would still meet AYP standards because the groups' scores are within the relevant confidence interval.

Population Diversity Effects

Minimum Pupil Group Size (n). Another important technical factor in state AYP standards is the establishment of the minimum size (n) for pupil groups to be considered in AYP calculations. NCLB recognizes that in the disaggregation of pupil data for schools and LEAs, there might be pupil groups that are so small that average test scores would not be statistically reliable, or the dissemination of average scores for the group might risk violation of pupils' privacy rights.

⁴⁴ The text above describes the way in which confidence intervals have been used by states for AYP determinations. The concept could be applied in a different way, requiring scores to be at or above the highest score in the “window” in order to demonstrate that a pupil group had met AYP standards to a statistically significant degree. This would reflect confidence (at the designated level of probability) that a school or LEA had met AYP standards, whereas the current usage reflects confidence that the school or LEA had failed to meet AYP standards.

Both the statute and ED regulations and other policy guidance have left the selection of this minimum number to state discretion. While most states have reportedly selected a minimum group size between 30 and 50 pupils, the range of selected values for “n” is rather large, varying from as few as five to as many as 200 pupils⁴⁵ under certain circumstances. One state (North Dakota) has set no specific level for “n,” relying only on the use of confidence intervals (see above) to establish reliability of test results. Although most states have always set a standard minimum size for all pupil groups, some states until recently established higher levels of “n” for pupils with disabilities or LEP pupils.⁴⁶

In general, the higher the minimum group size, the less likely that many pupil groups will actually be separately considered in AYP determinations. (Pupils will still be considered, but only as part of the “all pupils” group, or possibly other specified groups.) This gives schools and LEAs fewer thresholds to meet, and reduces the likelihood that they will be found to have failed to meet AYP standards. In many cases, if a pupil group falls below the minimum group size at the school level, it is still considered at the LEA level (where it is more likely to meet the threshold). In addition, since minimum group sizes for reporting achievement data are typically lower than those used for AYP purposes,⁴⁷ scores are often reported for pupil groups who are not separately considered in AYP calculations. At the same time, relatively high levels for “n” weaken NCLB’s specific focus on a variety of pupil groups, many of them disadvantaged, such as LEP pupils, pupils with disabilities, or economically disadvantaged pupils.

Separate Focus on Specific Pupil Groups. There are several ongoing issues regarding NCLB’s requirement for disaggregation of pupil achievement results in AYP standards, namely the requirement that a variety of pupil groups be separately considered in AYP calculations. The first of these was discussed immediately above: the establishment of minimum group size, with the possible result that relatively small pupil groups will not be considered in the schools and LEAs of states that set “n” at a comparatively high level, especially in states that set a higher level for certain groups (e.g., pupils with disabilities) than others.

A second issue arises from the fact that the definition of the specified pupil groups has been left essentially to state discretion. This is noteworthy particularly with respect to two groups of pupils: LEP pupils and pupils in major racial and ethnic groups. Regarding LEP pupils, many have been concerned about the difficulty of demonstrating that these pupils are performing at a proficient level if this pupil group is defined narrowly to include only pupils unable to perform in regular English-language classroom settings. In other words, if pupils who no longer need special language services are no longer identified as being LEP, how will it be

⁴⁵ In Texas, the minimum group size for pupil groups (other than the “all pupils” group, where the minimum is 40) is the greater of 50 students or 10% of all students in a school or LEA (up to a maximum of 200). In California, the minimum group size is the greater of 50 students or 15% of all students in the school or LEA (up to a maximum of 100).

⁴⁶ Under regulations published on April 9, 2007, this practice is no longer allowed.

⁴⁷ Minimum group sizes for AYP purposes are typically in the range of 30 to 40 pupils, while those for reporting are typically in the range of five to 20 pupils.

possible to bring those who are identified as LEP up to a proficient level of achievement?

In developing their AYP standards, some states addressed this concern by including pupils in the LEP category for one or more years after they no longer need special language services. As was discussed above, ED has recently published policy guidance encouraging all states to follow this approach, allowing them to continue to include pupils in the LEP group for up to two years after being mainstreamed into regular English language instruction, and further allowing the scores of LEP pupils to be excluded from AYP calculations for the first year of pupils' enrollment in United States schools. If widely adopted, these policies should reduce the extent that schools or LEAs are identified as failing to meet AYP standards on the basis of the LEP pupil group.

Another aspect of this issue arises from the discretion given to states in defining "major racial and ethnic groups." Neither the statute nor ED has defined this term. Some states defined the term relatively comprehensively (e.g., Maryland includes American Indian, African American, Asian, White, and Hispanic pupil groups) and some more narrowly (e.g., Texas identifies only three groups — White, African American, and Hispanic). A more narrow interpretation may reduce the attention focused on excluded pupil groups. It would also reduce the number of different thresholds some schools and LEAs would have to meet in order to make AYP.

A final, overarching issue arises from the relationship between pupil diversity in schools and LEAs and the likelihood of being identified as failing to meet AYP standards. All other relevant factors being equal (especially the minimum group size criteria), the more diverse the pupil population, the more thresholds a school or LEA must meet in order to make AYP. While in a sense this was an intended result of legislation designed to focus (within limits) on all pupil groups, the impact of making it more difficult for schools and LEAs serving diverse populations to meet AYP standards may also be seen as an unintended consequence of NCLB. This issue has been analyzed in a recent study by Thomas J. Kane and Douglas O. Staiger, who concluded that such "subgroup targets cause large numbers of schools to fail ... arbitrarily single out schools with large minority subgroups for sanctions ... or statistically disadvantage diverse schools that are likely to be attended by minority students.... Moreover, while the costs of the subgroup targets are clear, the benefits are not. Although these targets are meant to encourage schools to focus more on the achievement of minority youth, we find no association between the application of subgroup targets and test score performance among minority youth."⁴⁸ According to the "National Assessment of Title I: Final Report," published by ED in October 2007, among schools with relatively low poverty rates, the percentage of schools failing to make AYP ranged from 3% for those with only 1 subgroup to 25% for those with 3 subgroups, and 32% for those with 4 or 5 subgroups. Among schools with relatively

⁴⁸ Thomas J. Kane and Douglas O. Staiger, "Unintended Consequences of Racial Subgroup Rules," in Paul Peterson and Martin West, eds., *No Child Left Behind? The Politics and Practice of School Accountability* (Washington: Brookings Institution Press, 2003), pp. 152-176.

high poverty rates, the percentage of schools failing to make AYP ranged from 31% for those with only 1 subgroup to 70% for those with 6 or 7 subgroups.

An additional study published by Policy Analysis for California Education (PACE)⁴⁹ found that when comparing public schools in California with similar aggregate pupil achievement levels, schools with larger numbers of different NCLB-relevant demographic groups were substantially less likely to have met AYP standards in the 2002-2003 school year. Similarly when comparing California public schools with comparable percentages of pupils from low-income families, schools with larger numbers of relevant demographic groups of pupils were much less likely to have met AYP.

However, without specific requirements for achievement gains by each of the major pupil groups, it is possible that insufficient attention would be paid to the performance of the disadvantaged pupil groups among whom improvements are most needed, and for whose benefit the Title I-A program was established. Under previous law, without an explicit, specific requirement that AYP standards focus on these disadvantaged pupil groups, most state AYP definitions considered only the performance of all pupils combined. And it is theoretically possible for many schools and LEAs to demonstrate substantial improvements in achievement by their pupils overall while the achievement of their disadvantaged pupils does not improve significantly, at least until the ultimate goal of all pupils at the proficient or higher level of achievement is approached. This is especially true under a “status” model of AYP such as the one in NCLB, under which advantaged pupil groups may have achievement levels well above what is required, and an overall achievement level could easily mask achievement well below the required threshold by various groups of disadvantaged pupils.

One possible alternative to current policy would be to allow states to count each student only once, in net, in AYP calculations, with equal fractions for each relevant demographic category (e.g., a Hispanic LEP pupil from a low-income family would count as one-third of a pupil in each group).

Number of Schools Identified and State Variations Therein

As was discussed earlier, concern has been expressed by some analysts since early debates on NCLB that a relatively high proportion of schools would fail to meet AYP standards. On the basis of assessment results for 2006-2007, approximately 28% of all public schools nationwide failed to make AYP, and approximately 12% of all public schools were identified as needing improvement (i.e., failed to meet AYP standards for two or more consecutive years). Future increases in performance thresholds, as the ultimate goal of all pupils at the proficient or higher level of achievement is approached, may result in higher percentages of schools failing to make AYP.

⁴⁹ John R. Novak and Bruce Fuller, *Penalizing Diverse Schools?* PACE Policy Brief 03-4, December 2003.

In response to these concerns, ED officials have emphasized the importance of taking action to identify and move to improve underperforming schools, no matter how numerous. They have also emphasized the possibilities for flexibility and variation in taking corrective actions with respect to schools that fail to meet AYP, depending on the extent to which they fail to meet those standards. It should also be re-emphasized that many of the schools reported as having failed to meet AYP standards have failed to meet AYP for one year only, while NCLB requires that a series of actions be taken only with respect to schools or LEAs participating in ESEA Title I-A that fail to meet AYP for two consecutive years or more.

Further, some analysts argue that a set of AYP standards that one-quarter or more of public schools fail to meet may accurately reflect pervasive weaknesses in public school systems, especially with respect to the performance of disadvantaged pupil groups. To these analysts, the identification of large percentages of schools is a positive sign of the rigor and challenge embodied in NCLB's AYP requirements, and is likely to provide needed motivation for significant improvement (and ultimately a reduction in the percentage of schools so identified).

Others have consistently expressed concern about the accuracy and efficacy of an accountability system under which such a high percentage of schools is identified as failing to make adequate progress, with consequent strain on financial and other resources necessary to provide technical assistance, public school choice and supplemental services options, as well as other corrective actions. In addition, some have expressed concern that schools might be more likely to fail to meet AYP simply because they have diverse enrollments, and therefore more groups of pupils to be separately considered in determining whether the school meets AYP standards. They also argue that the application of technical assistance and, ultimately, corrective actions to such a high percentage of schools will dilute available resources to such a degree that these responses to inadequate performance would be insufficient to markedly improve performance. A few analysts even speculate that the AYP system under NCLB is intended to portray large segments of American public education as having "failed," leading to proposals for large scale privatization of elementary and secondary education.⁵⁰

The proportion of public schools identified as failing to meet AYP standards is not only relatively large in the aggregate, but also varies widely among the states. As was discussed above, the percentage of public schools identified as failing to make AYP on the basis of assessment results for 2006-2007 ranged from 4% to 75% among the states. This result is somewhat ironic, given that one of the major criticisms of the pre-NCLB provisions for AYP was that they resulted in a similarly wide degree of state variation in the proportion of schools identified, and the more consistent structure required under NCLB was widely expected to lead to greater consistency among states in the proportion of schools identified.

It seems likely that the pre-NCLB variations in the proportion of schools failing to meet AYP reflected large differences in the nature and structure of state AYP

⁵⁰ See Alfie Kohn, "Test Today, Privatize Tomorrow: Using Accountability to 'Reform' Public Schools to Death," *Phi Delta Kappan*, vol. 85, no. 8 (April 2004), pp. 568-577.

standards, as well as major differences in the nature and rigor of state pupil performance standards and assessments. While the basic structure of AYP definitions is now substantially more consistent across states, significant variations remain with respect to the factors discussed in this section of the report (such as minimum group size or use of confidence intervals), and substantial differences in the degree of challenge embodied in state standards and assessments remain. Overall, it seems likely that the key influences determining the percentage of a state's schools that fails to make AYP include (in no particular order): (1) degree of rigor in state content and pupil performance standards; (2) minimum pupil group size (n) in AYP determinations; (3) use of confidence intervals in AYP determinations (and whether at a 95% or 99% level of confidence); (4) extent of diversity in pupil population; (5) extent of communication about, and understanding of, the 95% test participation rule; and (6) possible actual differences in educational quality.

95% Participation Rule

It appears that in many cases, schools or LEAs have failed to meet AYP solely because of low participation rates in assessments, meaning that fewer than 95% of all pupils, or of pupils in relevant demographic groups meeting the minimum size threshold, took the assessments. While, as discussed above, ED recently published policy guidance that relaxes the participation rate requirement somewhat — allowing use of average rates over two- to three-year periods, and excusing certain pupils for medical reasons — the high rate of assessment participation that is required in order for schools or LEAs to meet AYP standards is likely to remain an ongoing focus of debate.

Although few argue against having any participation rate requirement, it may be questioned whether it needs to be as high as 95%. In recent years, the overall percentage of enrolled pupils who attend public schools each day has been approximately 93.5%, and it is generally agreed that attendance rates are lower in schools serving relatively high proportions of disadvantaged pupils. Even though schools are explicitly allowed to administer assessments on make-up days following the primary date of test administration, and it is probable that more schools and LEAs will meet this requirement as they become more fully aware of its significance, it is likely to continue to be very difficult for many schools and LEAs to meet a 95% test participation requirement.

State Variations in Assessments and Proficiency Standards

As noted above, it is likely that state variations in the percentage of schools failing to meet AYP standards are based not only on underlying differences in achievement levels, as well as a variety of technical factors in state AYP provisions, but also on differences in the degree of rigor or challenge in state pupil performance standards and assessments. Particularly now that all states receiving Title I-A grants must also participate in state-level administration of NAEP tests in 4th and 8th grade reading and math every two years, this variation can be illustrated for all states by comparing the percentage of pupils scoring at the proficient level on NAEP versus state assessments.

Such a comparison was conducted by a private organization, Achieve, Inc., based on 8th grade reading and math assessments administered in the spring of 2003.⁵¹ For a variety of reasons (e.g., several states did not administer standards-based assessments in reading or math to 8th grade pupils in 2003), the analysis excluded several states; 29 states were included in the comparison for reading, and 32 states for math. According to this analysis, the percentage of pupils statewide who score at a proficient or higher level on state assessments, using state-specific pupil performance standards, was generally much higher than the percentage deemed to be at the proficient or higher level on the NAEP tests, and employing NAEP's pupil performance standards. Of the states considered, the percentage of pupils scoring at a proficient or higher level on the state assessment was lower than on NAEP (implying a more rigorous state standard) for five states⁵² (out of 32) in math and only two states (out of 29) in reading.

Further, among the majority of states where the percentage of pupils at the proficient level or above was found to be higher on state assessments than on NAEP, the relationship between the size of the two groups varied widely — in some cases only marginally higher on the state assessment, and in others the percentage at the proficient level was more than twice as high on the state assessment as on NAEP. Although some portion of these differences in performance may result from differences in the motivation of pupils to perform well (and of teachers to encourage high performance) on NAEP versus state assessments, comparisons to NAEP results help to illuminate the variations in state proficiency standards. It is not yet clear whether such comparisons will significantly encourage greater consistency in those standards.

A second issue is whether some states might choose to lower their standards of “proficient” performance, in order to reduce the number of schools identified as failing to meet AYP and make it easier to meet the ultimate NCLB goal of all pupils at the proficient or higher level by the end of the 2013-2014 school year. In the affected states, this would increase the percentage of pupils deemed to be achieving at a “proficient” level, and reduce the number of schools failing to meet AYP standards.

Although states are generally free to take such actions without jeopardizing their eligibility for Title I-A grants, because performance standards are ultimately state-determined and have always varied significantly, such actions have elicited public criticism from ED. In a policy letter dated October 22, 2002, the Secretary of Education stated

Unfortunately, some states have lowered the bar of expectations to hide the low performance of their schools. And a few others are discussing how they can ratchet down their standards in order to remove schools from their lists of low performers. Sadly, a small number of persons have suggested reducing standards for defining “proficiency” in order to artificially present the facts.... Those who

⁵¹ Center on Education Policy, *From the Capital to the Classroom, Year 2 of the No Child Left Behind Act* (January 2004), p. 61.

⁵² In two additional states, the percentages were essentially the same.

play semantic games or try to tinker with state numbers to lock out parents and the public, stand in the way of progress and reform. They are the enemies of equal justice and equal opportunity. They are apologists for failure.⁵³

⁵³ See [<http://www.ed.gov/news/pressreleases/2002/10/10232002a.html>].