



Educational Testing: Implementation of ESEA Title I-A Requirements Under the No Child Left Behind Act

Rebecca R. Skinner
Specialist in Education Policy

Erin D. Caffrey
Analyst in Education Policy

February 5, 2010

Congressional Research Service

7-5700

www.crs.gov

RL31407

Summary

The No Child Left Behind Act of 2001 (NCLB) contains several requirements related to student assessments for states and local educational agencies (LEAs) participating in Elementary and Secondary Education Act (ESEA) Title I-A (Education for the Disadvantaged). Under the NCLB, in addition to previous requirements for standards and assessments in reading and mathematics at three grade levels, all states participating in Title I-A were required to implement standards-based assessments for students in *each* of grades 3-8 in reading and mathematics by the end of the 2005-2006 school year. States must also implement assessments at three grade levels in *science* by the end of the 2007-2008 school year. Students who have been in U.S. schools for at least three years must be tested (for reading) in English, and states must annually assess the English language proficiency of their limited English proficient (LEP) students. The grants to states program for assessment development was appropriated \$410.7 million for FY2010.

In addition, the NCLB requires all states receiving grants under Title I-A to participate in National Assessment of Educational Progress (NAEP) tests in 4th and 8th grade reading and mathematics to be administered every two years, with all costs to be paid by the federal government. NAEP is a series of ongoing assessments of the academic performance of representative samples of students primarily in grades 4, 8, and 12. Beginning in 1990, NAEP has conducted a limited number of state-level assessments wherein the sample of students tested in each participating state is increased in order to provide reliable estimates of achievement scores for students in the state. Previously, all participation in state NAEP was voluntary, and additional costs associated with state NAEP were borne by participating states. The statutory provisions authorizing NAEP are amended by the NCLB to maximize consistency with the NCLB requirements and prohibit the use of NAEP assessments by agents of the federal government to influence state or LEA instructional programs or assessments.

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. Issues regarding expanded ESEA Title I-A student assessment requirements that are being addressed by the 111th Congress include the following: Are states meeting the expanded assessment requirements on schedule? Will federal grants be sufficient to pay the costs of meeting the assessment requirements? What might be the impact on NAEP of requiring state participation, as well as the impact of NAEP on state standards and assessments? What are the likely major benefits and costs of the expanded ESEA Title I-A student assessment requirements? And should the assessment requirements be expanded further?

This report will be updated regularly to reflect major legislative developments and available information.

Contents

Introduction	1
Pre-NCLB State Testing Policies and Practices	1
Testing Program Costs	2
Federal Policies or Activities Regarding Student Assessments Under the No Child Left Behind Act	3
ESEA Title I-A Requirements for Standards and Assessments.....	4
Schedule for Implementation of All Assessment Requirements	7
Limits on ED Influence Over State Standards and Assessments	7
State Assessment Grants.....	8
National Assessment of Educational Progress.....	9
State NAEP.....	10
NAEP Provisions in the No Child Left Behind Act	11
Interpretation by ED of the Expanded Standard and Assessment Requirements of the No Child Left Behind Act	12
Title I-A Standard and Assessment Requirements	12
Evolution of ED Policy Regarding Participation Rates Plus Treatment of Limited English Proficient Students and Certain Students With Disabilities in Assessments and AYP Determinations.....	15
Regulations Published in October 2008 on Title I-A Assessments and Accountability	18
ED Review to Determine Whether States Meet 2005-2006 Assessment Requirements.....	19
Implementation of the NAEP Requirements	20
Issues Regarding the ESEA Title I-A Student Assessment Requirements	21
What Types of Assessments Meet the Expanded Assessment Requirements?	21
How Strict Is ED’s Review of State Assessment Systems?.....	23
What Is the Cost of Developing and Implementing the Required Assessments, and to What Extent Will Federal Grants Be Available to Pay for Them?.....	24
What Might Be the Impact of the Requirement for Annual Assessment of English Language Proficiency of LEP Students?	26
What Might Be the Impact of Requiring State Participation in NAEP?	27
Possible Influence on State Standards and Assessments Arising from (Marginally) Increased Stakes.....	27
Voluntary Participation by LEAs, Schools, and Students.....	28
Can NAEP Results Be Used to “Confirm” State Test Score Trends?	28
Likely Benefits and Costs of the Expanded Title I-A Assessment Requirements.....	29
Glossary of Selected Terms Used in This Report	31

Contacts

Author Contact Information	32
Acknowledgments	32

Introduction

The No Child Left Behind Act of 2001 (NCLB, P.L. 107-110), signed into law on January 8, 2002, contains a number of new requirements related to student assessments for states and local educational agencies (LEAs) participating in Title I-A (Education for the Disadvantaged) of the Elementary and Secondary Education Act (ESEA). These assessment requirements expand upon an earlier series of requirements for participating states to adopt curriculum content standards, academic achievement standards, and assessments linked to these at three grade levels, which were adopted under the Improving America's Schools Act (IASA) of 1994 (P.L. 103-382).

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 provides background information on state student assessment programs and policies, a description of the ESEA Title I-A assessment requirements as expanded by the NCLB, a review of the implementation status of these requirements, and an analysis of related issues likely to be addressed by the 111th Congress. This report will be updated regularly to reflect major legislative developments and available information.

Pre-NCLB State Testing Policies and Practices

The academic achievement of students in public elementary and secondary schools is assessed using many types of tests. Students may take tests developed by individual teachers or schools, commercially published tests selected by their LEA, or assessments selected or developed by their state educational agency (SEA). *This report will focus almost entirely on state-mandated assessments*—tests which must be administered to virtually all students in selected grades who attend a state's public K-12 schools—because such tests are the primary focus of federal policies regarding student assessment.

According to published surveys,¹ every state except one (Iowa) now requires its LEAs to administer specified assessments to all students attending public schools in one or more grades.² The number of grades and subjects in which state-mandated assessments are administered varies widely, from only one grade and subject (e.g., the only state-mandated assessment in Nebraska currently is a writing test for students in grade 4) to tests in multiple subjects and most K-12 grades (e.g., Alabama requires students in each of grades 3-11 to take state-selected tests in English, mathematics, science, and history). Few state-mandated tests are administered to students below grade 3, because of a variety of concerns about administering standardized tests to very young students, or in grade 12, in part because most assessment activity for these students is focused on college entrance tests. With respect to grades 3-8 in particular, 15 states plus the District of Columbia currently administer assessments in mathematics and reading to students in

¹ Much of the data in this section is derived from *No State Left Behind: The Challenges and Opportunities of ESEA 2001*, by the Education Commission of the States, available at <http://www.ecs.org>; and *Assessment and Accountability Systems: 50 State Profiles*, by the Consortium for Policy Research in Education, available at http://www.cpre.org/Publications/Publications_Accountability.htm.

² While Iowa does not mandate participation in any specific assessment, tests developed by the Iowa Testing Programs at the University of Iowa and published nationwide by Riverside Publishing are administered to a large majority of students attending public K-12 schools in Iowa, on the basis of voluntary decisions by each LEA.

each of these grades; however, it is unclear how many of these assessments are linked to state content and achievement standards.

State-mandated assessments have been developed in one of three basic patterns. They are either: (a) developed by the states themselves, usually with technical assistance from commercial firms employing assessment specialists; (b) developed almost completely by commercial test publishers, either as generic tests sold in the same form throughout the nation,³ or special versions of such tests which are customized to be more consistent with the curriculum content and achievement standards of a state; or (c) developed through multi-state consortia.⁴

Some state-mandated assessments, whether developed by the states themselves or in cooperation with other states or commercial firms, are “criterion-referenced” tests, or CRTs (see Glossary) designed to determine the extent to which students have mastered specific curriculum content and skills. Other state-mandated tests are either generic or customized “norm-referenced” tests, or NRTs (see Glossary)—tests designed primarily to rank students’ achievement level in comparison to a nationally representative sample of students—purchased by states from commercial test publishers. These two types of tests vary *primarily* regarding how test results are analyzed, but also typically differ to some degree with respect to such characteristics as the range of questions included.⁵

As of spring 2000, immediately preceding consideration of the NCLB, two states (Montana and South Dakota) administered only NRTs, 17 administered only CRTs, and 29 administered both kinds of tests in different grades and/or subject areas, with six of the latter states (Alabama, Idaho, Montana, South Dakota, West Virginia, and Wisconsin) using NRTs as their primary assessment instruments. In addition, six states (California, Delaware, Indiana, Missouri, New Mexico, and Tennessee) had developed state tests that are designed to produce *both* achievement results linked to state standards (criterion-referenced results) and nationally normed results (norm-referenced results).

Testing Program Costs

Complete information on the costs associated with state-mandated student testing programs is not available. There are many potential sources of such costs, both direct and indirect, at the state, LEA, and school levels, and there are unresolved debates over how to estimate and whether to consider certain types of costs, especially indirect ones.⁶

³ Three of the largest such commercial test publishers are: (1) CTB/McGraw-Hill, at <http://www.ctb.com/>; (2) Riverside (Houghton Mifflin) Publishing, at <http://www.riverpub.com/>; and (3) Harcourt Assessment, at <https://harcourtassessment.com/hai/International.aspx>.

⁴ One example of such a consortium is the New Standards Project, a joint effort of several states and LEAs, the National Center on Education and The Economy, and the Learning Research and Development Center at the University of Pittsburgh. Another is a consortium for assessment development formed by three New England states—New Hampshire, Rhode Island, and Vermont.

⁵ For example, in order to clarify distinctions between high- and low-achieving students, a norm-referenced test will typically include some very difficult questions that only a few students can answer, and some very easy questions that almost all students can answer correctly. Test content and questions are selected largely on the basis of how efficiently they rank students. In contrast, a CRT would be focused solely on the relevant content standards, with no direct emphasis on distinguishing the highest- from the lowest-achieving students.

⁶ Direct expenditures include those for such activities and services as development and field testing of assessments, purchase of test materials, scoring, or dissemination of results. Indirect expenditures might include those for time spent (continued...)

A survey of *direct, state-level* expenditures for state-mandated assessment programs was conducted in early 2001 by the Pew Center on the States.⁷ These data combine state-level expenditure for both test development and administration for FY2001 (FY2000 for North Dakota and Vermont). The figures do not include any LEA-level expenditures, either direct or indirect, nor possible indirect state-level expenditures for state-mandated testing programs.

According to this survey, state-level, direct expenditures for K-12 student assessment programs in FY2001 totaled \$422.8 million. The expenditures per state varied from zero for Iowa and \$0.2 million for North Dakota, to \$44.0 million for California and \$26.7 million for Texas. On a per-pupil basis, these costs were found to vary from \$1.46 per pupil in West Virginia to \$82.55 per pupil in Alaska. Per pupil costs of state-mandated assessments tend to be low in states which rely primarily on versions of commercially-published NRTs, such as West Virginia, Alabama (\$7.80 per student), New Mexico (\$3.21 per student), and Utah (\$3.16 per student). In contrast, per pupil costs were found to be highest for several states which rely primarily or solely on state-specific CRTs, such as Alaska, Wyoming (\$78.34 per student), Virginia (\$68.90 per student) and Massachusetts (\$68.02 per student).⁸

More detailed, but less comprehensive or current, information may be found in a study of the costs of developing and initially implementing assessments aligned with curriculum standards in two states—Kentucky and North Carolina. According to this study,⁹ the total five-year state-level costs of developing and implementing a new assessment aligned with state standards for Kentucky were \$9.55 million (\$1.9 million per year) for test development and \$33.3 million (\$6.67 million per year) in total (including development, administration, etc.). For North Carolina, the total three-year state-level costs were found to be \$4.0 million (\$1.34 million per year) for test development and \$27.5 million (\$5.5 million per year) in total. The costs for these two states are not necessarily representative of the costs for all states. For example, costs might be lower for states which develop tests jointly with a group of other states, or which contract with a commercial test publisher for a customized version of a test which is marketed nationwide in a generic form.

Federal Policies or Activities Regarding Student Assessments Under the No Child Left Behind Act

The following section of this report describes the major student assessment-related provisions of the ESEA *as amended* by the NCLB.

(...continued)

by teachers and other staff preparing students for or administering assessments or overhead costs. For a review of related issues, see Richard P. Phelps, "Estimating the Costs of Standardized Student Testing in the United States," *Journal of Education Finance*, winter 2000, pp. 343-380.

⁷ Available at <http://www.stateline.org/live/ViewPage.action?siteNodeId=136&languageId=1&contentId=14274>.

⁸ See Education Commission of the States, *Estimated Per-Student Spending on Statewide Testing Programs*, October 2001, available at <http://www.ecs.org>.

⁹ Lawrence O. Picus, *Estimating the Costs of Student Assessment in North Carolina and Kentucky: A State-Level Analysis*, CRESST Technical Report 408, February 1996.

ESEA Title I-A Requirements for Standards and Assessments

The provisions of ESEA Title I-A, as amended by the NCLB, regarding standards and assessments reinforce and expand upon provisions initially adopted in the Improving America's Schools Act of 1994 (IASA). Whether under the IASA or the NCLB, these standards and assessment provisions are linked to receipt of financial assistance under ESEA Title I-A—that is, they apply *only* to states wishing to maintain eligibility for Title I-A grants. However, since Title I-A is the largest federal K-12 education program, funded at \$14.5 billion for FY2010, it is generally considered unlikely that many states would decline to participate in the program in order to avoid implementing the expanded assessment requirements.

The IASA of 1994 attempted to raise the instructional standards of Title I-A programs, and the academic expectations for participating students, by tying Title I-A instruction to state-selected curriculum content and academic achievement standards. These provisions were adopted in response to concerns that Title I-A programs had not been sufficiently challenging academically; had not been well integrated with the “regular” instructional programs of participants; and had required extensive student testing that was of little instructional or diagnostic value, and was not linked to the curriculum to which students were exposed. Further, the legislation attempted to make Title I-A tests more meaningful by using state assessments to determine whether schools and LEAs are making “adequate yearly progress” (AYP) toward meeting state achievement standards.¹⁰

States were given several years to meet the IASA requirements. In particular, the full system of standards and assessments was not required to be in place until the 2000-2001 school year and, as is discussed in detail below, only a minority of states met that deadline. Thus, in its debates on the NCLB in 2001, the Congress considered not only the expanded assessment requirements proposed by the Bush Administration, but also the implementation status of requirements adopted in 1994.

Under the ESEA, as amended first by the IASA of 1994 and later by the NCLB of 2001, states wishing to remain eligible for Title I-A grants are required to develop or adopt curriculum content standards as well as academic achievement standards and assessments tied to the standards. In general, these standards and assessments are to be applicable to all LEAs, schools, and students statewide. One major exception to this general policy is that if no agency or entity in a state has the authority to establish statewide standards or assessments (as is generally assumed to be the case for Iowa and Nebraska), then the state may adopt either: (a) statewide standards and assessments applicable only to Title I-A students and programs, or (b) a policy providing that each LEA receiving Title I-A grants will adopt standards and assessments which meet the requirements of Title I-A and are applicable to all students served by each such LEA. Another possible exception, which is discussed further below, is that ED regulations would allow local variation in the assessments used for at least some grade levels. Thus, it should be kept in mind that “state systems of standards and assessments,” as referred to frequently below, may not in some cases be uniform statewide.

In order to comply with the provisions of ESEA Title I-A, state systems of standards and assessments are required to meet a number of specific statutory requirements, as follows:

¹⁰ See CRS Report RL32495, *Adequate Yearly Progress (AYP): Implementation of the No Child Left Behind Act*, by Rebecca R. Skinner.

1. Standards and assessments at 3 grade levels were to be developed or adopted at least in the subjects of mathematics and reading/language arts by the 2000-2001 school year.¹¹ Standards were to be adopted in science by the end of the 2005-2006 school year, and assessments in science by the end of the 2007-2008 school year.
2. The standards and assessments used to meet the Title I-A eligibility requirements must be the same as those applied to all public school students in the state (with the two possible exceptions discussed above).
3. The content standards are to specify what students are expected to know and be able to do, and are to be “coherent and rigorous.”
4. Achievement standards must establish *at least* three performance levels for all students—advanced, proficient, and partially proficient (or basic).
5. Assessments must be aligned with state content and achievement standards.
6. Assessments in mathematics, reading and, beginning in 2007-2008, science must be administered annually to students in at least one grade in each of three grade ranges—grades 3-5, grades 6-9, and grades 10-12. In addition, assessments in mathematics and reading were to be administered to students in *each* of grades 3-8 by the end of the 2005-2006¹² school year.¹³
7. All students in the relevant grades who have attended schools in the LEA for at least one year must participate in the assessments.¹⁴
8. LEP students are to be assessed in a valid and reliable manner and provided with “reasonable” accommodations. To the extent practicable, LEP students are to be assessed in the language and form most likely to yield accurate and reliable information on what they know and can do in academic content areas (in subjects other than English itself). However, students who have attended schools in the United States (excluding Puerto Rico) for three or more consecutive school years are to be assessed in English.¹⁵
9. “Reasonable” adaptations and accommodations are to be provided for students with disabilities, consistent with the provisions of the Individuals with Disabilities Education Act (IDEA) where such adaptations or accommodations are necessary to measure the achievement of those students relative to state standards.

¹¹ As is discussed later in this report, most states did not meet this deadline, established in the 1994 IASA.

¹² There is explicit authority for a one-year delay of this requirement in cases of exceptional or uncontrollable circumstances.

¹³ There is some obvious overlap in these requirements—e.g., states meeting the requirement for assessments in reading and math at three grade levels already meet the requirements for one or two of grades 3-8.

¹⁴ Separately, the provisions regarding AYP provide that at least 95% of the students in each demographic group within each school must be included in the assessments in order for the school to meet AYP requirements. Students may be excluded from school-level score reporting and accountability if they have attended a specific school for less than one year.

¹⁵ LEAs may continue to administer assessments to students in non-English languages for up to five years if, on a case-by-case basis, they determine that this would likely yield more accurate information on what the students know and can do.

10. The assessment system must involve multiple approaches with up-to-date measures of student performance, including measures that assess higher order thinking skills and understanding.

11. Assessments must be used for purposes for which they are valid and reliable, and they must meet relevant, nationally recognized, professional and technical standards. In particular, the state educational agency (SEA) must provide evidence from a test publisher or other relevant source that the assessments are of adequate technical quality for the purposes required under Title I-A.

12. The assessment system must produce individual student interpretive and diagnostic reports that are provided to parents, teachers, and principals as soon as is “practically possible” after the assessments are administered. It must also enable “itemized score analyses” to be produced and reported to LEAs and schools, so that specific academic needs may be identified.

13. The assessment system must enable results for each state, LEA, and school, to be disaggregated (i.e., reported separately) by gender, major racial and ethnic groups, English proficiency status, migrant status, students with disabilities as compared to students without disabilities, and economically disadvantaged students as compared to students who are not economically disadvantaged. However, such disaggregation is not required in cases where the number of students in a group would be too small to yield statistically reliable information or where personally identifiable information would be revealed.

14. Assessments must objectively measure academic achievement, knowledge, and skills, and not assess personal or family beliefs and attitudes, or disclose personally identifiable information.

15. Assessment results must be provided to LEAs, schools, and teachers before the beginning of the subsequent school year.

16. In addition to the general assessment system described in 1-15 above, states are to provide that their LEAs will annually assess the English language proficiency of their LEP students—including students’ oral, reading, and writing skills.¹⁶

Finally, as is discussed further below, states receiving grants under ESEA Title I-A must participate in biennial state-level administrations of the National Assessment of Educational Progress in 4th and 8th grade reading and mathematics, beginning in the 2002-2003 school year. The timing of several of the key requirements listed above is summarized in the following schedule.

¹⁶ A one-year waiver of this requirement is specifically authorized in cases of exceptional or uncontrollable circumstances.

Schedule for Implementation of All Assessment Requirements

School Year 2000-2001

- States were to have adopted content and performance standards, plus assessments linked to these, at three grade levels in mathematics and reading. These requirements were included in the 1994 reauthorization of the ESEA. (As of the date of this report, 39 states fully met these requirements.)

School Year 2002-2003

- States were required to begin to annually assess the English language proficiency of LEP students (possible one-year waiver for “exceptional or uncontrollable circumstances”).
- States were first required to participate in biennial administration of NAEP.
- Annual report cards on state and LEA school systems and schools were required to be published (with a possible one year waiver authorized for “exceptional or uncontrollable circumstances”).
- States were required to begin reporting annually to ED on progress toward meeting new assessment and related requirements under the NCLB.

School Year 2005-2006

- Standards-based assessments in reading and mathematics were to be administered to students in each of grades 3-8 by the end of this year.
- States were required to adopt content and achievement standards at three grade levels in science by the end of this year.

School Year 2007-2008

- States were to begin to administer assessments at three grade levels in science by the end of this year.

Limits on ED Influence Over State Standards and Assessments

Several statutory constraints have been placed on the authority of the Secretary of Education to enforce these standard and assessment requirements. First, the ESEA contains a provision—similar to others found in the Department of Education Organization Act and the General Education Provisions Act—stating that nothing in ESEA Title I shall be construed to authorize any federal official or agency to “mandate, direct, or control a State, local educational agency, or school’s specific instructional content, academic achievement standards and assessments, curriculum, or program of instruction” (Section 1905).¹⁷ Second, states may not be required to submit their standards to the U.S. Secretary of Education (Section 1111(b)(1)(A)) or to have their content or achievement standards approved or certified by the federal government (Section

¹⁷ Similar, although somewhat less specific, language may be found in ESEA Section 9526(b)(1) and Section 9527(a).

9527(c)) in order to receive funds under the ESEA, other than the (limited) review necessary in order to determine whether the state meets the Title I-A requirements. Finally, no state plan may be disapproved by ED on the basis of specific content or achievement standards or assessment items or instruments (Section 1111(e)(1)(F)).

State Assessment Grants

The ESEA authorizes (in Title VI-A-1) annual grants to the states to help pay the costs of meeting the Title I-A standard and assessment requirements added by the NCLB (i.e., the newly required assessments in science at three grade levels and at grades 3-8 in mathematics and reading). These grants may be used by states for development of standards and assessments or, if these have been developed, for assessment administration and such related activities as developing or improving assessments of the English language proficiency of LEP students. The amount authorized to be appropriated for these state assessment grants, plus grants for development of enhanced assessment instruments (see below), is \$490 million for FY2002 and “such sums as may be necessary” for each of FY2003-FY2008.

The state assessment requirements that were newly adopted under the NCLB are contingent upon the appropriation of minimum annual amounts for these state assessment grants. The administration, but not the development, of grade 3-8 and science assessments may be delayed by one year for each year that the following minimum amounts are *not* appropriated: FY2002, \$370 million; FY2003, \$380 million; FY2004, \$390 million; and each of FY2005-FY2008, \$400 million. For example, if an amount less than \$400 million had been appropriated for state assessment grants for FY2005, the deadline for state administration of tests in reading and mathematics for each of grades 3-8 would have moved from 2005-2006 to 2006-2007. For each of FY2002-FY2008, at least the minimum amounts have been appropriated for these grants.

The state assessment grants are to be allocated as follows: after reservation of 0.5% of the total for the Outlying Areas and 0.5% for the Bureau of Indian Affairs, each state will first receive \$3 million. Remaining funds will be allocated among the states in proportion to their number of children and youth aged 5-17 years. This allocation formula reflects an implicit assumption that costs of assessment development are partially similar for all states, regardless of their size, and partially related to the size of the state’s school age population.

The ESEA also authorizes competitive grants to states for the development of enhanced assessment instruments. Aided activities may include efforts to improve the quality, validity, and reliability of assessments beyond the levels required by Title I-A, to track student progress over time, or to develop performance or technology-based assessments. Funds appropriated each year for state assessment grants which are in excess of the “trigger” amounts for assessment development grants listed above are to be used for enhanced assessment grants. The amounts available for assessment enhancement grants thus far are \$17 million for FY2002, \$4.5 million for FY2003, none for FY2004, \$11.7 million for FY2005, \$7.6 million for each of FY2006 and FY2007, \$8.7 million for FY2008, and \$10.7 million for each of FY2009 and FY2010.

Finally, the NCLB authorizes a study of the impact of the expanded Title I-A assessment requirements. The Secretary of Education is authorized to use the lesser of 15% of total appropriations for Title I, Part E (National Assessment of Title I) or \$1.5 million per year to contract for an independent study of “assessments used for State accountability purposes,” including the correlations between such assessments and student achievement, instructional practices, dropout and graduation rates, and school staff turnover rates; effects on different groups

of students, such as LEP students, students from low-income families, or students with disabilities; and relationships between accountability systems and exclusion of students from state assessments.

National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP) is a federally funded series of assessments of the academic performance of elementary and secondary students in the United States. NAEP tests generally are administered to public and private school students in grades 4, 8, and 12 in a variety of subjects, including reading, mathematics, science, writing and, less frequently, geography, history, civics, social studies, and the arts. NAEP assessments have been conducted since 1969.

NAEP is administered by the National Center for Education Statistics (NCES), with oversight and several aspects of policy set by the National Assessment Governing Board (NAGB), both within the U.S. Department of Education. Since 1983, the assessment has been developed primarily under a cooperative agreement with the Educational Testing Service (ETS), a private, non-profit organization which also develops and administers such assessments as the SAT. A private business firm, Westat, Inc., carries out much of the test administration activities. Two other private firms, National Computer Systems and American Institutes for Research, distribute and score the assessments and develop the background questionnaires, respectively.

NAEP consists of two separate groups of tests. One is the *main assessment*, in which test items (questions) are revised over time in both content and structure to reflect more current views and practices. The main assessment also reports student scores in relation to *performance levels*—standards for student achievement that are based on score thresholds set by NAGB. The performance levels are considered to be “developmental,” and are intended to place NAEP scores into context. They are based on determinations by NAGB of what students should know and be able to do at a basic (“partial mastery”), proficient (“solid academic performance”), and advanced (“superior performance”) level with respect to challenging subject matter.

The second group of NAEP tests form the long-term trend assessment, which monitors trends in math and reading achievement.¹⁸ The tests in each subject area have not changed in content or structure since they were originally developed in 1969, purportedly making it possible to reliably compare results from year to year. However, many have expressed concerns that the long-term trend assessment questions may be increasingly disconnected from what students are actually taught with the passage of decades of time.¹⁹ Since the long-term trend assessment is not involved with the ESEA Title I-A assessment requirements, it will not be discussed further.

All NAEP tests are administered to only a sample of students, and the tests are designed so that no student takes an entire NAEP test. The use of sampling is intended to minimize both the costs of NAEP and test burdens on students. It also makes it possible to include a broad range of items in each test. Since no individual student takes an entire NAEP test, it is impossible for NAEP to

¹⁸ Additional long-term trend assessments in writing and science were last administered in 1999. There is no current plan to administer the writing assessment in the future; revised science assessment test items are being developed, and may be administered in the future.

¹⁹ An NAGB policy adopted in May 2002 addresses this concern with respect to the science assessment, and changes were to be made to the content of the science assessment before its next administration.

report individual student scores. It is intended that NAEP tests be administered to a representative sample of all students in public and private schools, although there has been ongoing debate over whether LEP students or those with disabilities are adequately represented and whether appropriate accommodations or adaptations are being provided for them.

The frameworks for NAEP tests provide a broad outline of the content on which students are to be tested. Frameworks are developed by NAGB through a national consensus approach involving teachers, curriculum specialists, policymakers, business representatives, and the general public. In developing the test frameworks, national and various standards are taken into consideration, but the frameworks are not intended to specifically reflect any particular set of standards. In addition, students and school staff fill out background questionnaires. The NAEP statute limits the range of background information that may be collected to data “directly related to the appraisal of academic achievement, and to the fair and accurate presentation of such information” (Section 303(b)(5)(B)) plus demographic data on student race, ethnicity, socioeconomic status, disability, LEP status, and gender.

State NAEP

While NAEP, as currently structured, cannot provide assessment results for individual students, the levels at which scores could be provided, whether the nation overall, states, LEAs, or schools, depend on the size and specificity of the sample group of students tested. NAEP has always provided scores for the Nation as a whole and four multistate regions. Beginning in 1990, NAEP has conducted a limited number of state-level assessments in 4th and 8th grade mathematics and reading. In addition, state science assessments have been administered to 4th and 8th grade students in 1996 (8th grade only), 2000, and 2005. Only the main NAEP, not the long-term trend assessment, is administered at the state level. Under state NAEP, the sample of students tested in a state is increased in order to provide reliable estimates of achievement scores for students in each participating state.

Until enactment of the NCLB (see below), participation in NAEP was voluntary for states,²⁰ the additional cost associated with state NAEP administration was borne by the states and, after participating in any state NAEP test, states could separately decide whether to allow release of NAEP results for their state. As with other main NAEP tests, state NAEP scores are reported with respect to performance levels—basic, proficient, and advanced—developed by NAGB. In general, approximately 40 states participated in each state-level NAEP assessment administered between 1990 and 2000, and all states except one (South Dakota) participated in state NAEP at least once during this period.

In addition to this administration of NAEP at a state level, the FY2002 appropriations provided for a Trial Urban District Assessment of achievement in reading and writing: experimental administration of NAEP to expanded student samples in a limited number of large urban LEAs. The assessment was administered to extended samples of students in 2002 in Atlanta, Chicago, the District of Columbia, Houston, Los Angeles, and New York City, as part of the regular state and national assessment activities.²¹ Additional trial urban assessments were conducted in 2003,

²⁰ Once states decided to participate they were not prohibited from mandating participation by LEAs or schools under state and local law, although it appears that most states have always attempted to obtain LEA and school participation through voluntary recruitment.

²¹ For a description of the Trial Urban Assessment, and available results, see <http://nationsreportcard.gov/> (continued...)

2005, and 2007. In 2009, 18 metropolitan school districts participated in the Trial Urban District Assessment: Atlanta, Austin, Baltimore, Boston, Charlotte-Mecklenburg, Chicago, Cleveland, Detroit, District of Columbia, Fresno, Houston, Jefferson County (KY), Los Angeles, Miami-Dade, Milwaukee, New York City, Philadelphia, and San Diego.

NAEP Provisions in the No Child Left Behind Act

The NCLB provides that all states wishing to remain eligible for grants under ESEA Title I-A will be *required* to participate in state NAEP tests in 4th and 8th grade reading and mathematics, which are to be administered every two years. The costs of testing expanded student samples in the states in these subjects are now paid by the federal government. An unstated, but implicit, purpose of this new requirement is to “confirm” trends in student achievement, as measured by state-selected assessments.²² The results from the initial state NAEP assessment in 4th and 8th grade reading and mathematics involving all 50 states were released in 2003, with subsequent rounds of results released in 2005 and 2007.

In addition, the authorizing statute for NAEP (at that time, Sections 411-412 of the National Education Statistics Act, or NESAs) was almost completely rewritten in the NCLB. Although most of the new provisions are essentially the same as previous law, the statute has been amended in several respects. It is explicitly provided that students in home schools may not be required to participate in NAEP tests. Agents of the federal government are prohibited from using NAEP assessments to influence state or LEA instructional programs or assessments. Mechanisms are provided for limited public access to NAEP questions and test instruments and for review of complaints about NAEP tests. Provisions regarding NAGB are revised to specify that at least two members must be parents who are not employed by any educational agency. Regarding the release of state NAEP results, participating states still may choose not to allow such release but *only* with respect to state NAEP tests *other than* those required for Title I-A purposes.

There are conflicting statutory and regulatory provisions regarding participation in NAEP tests by LEAs and schools that may be selected for NAEP test administration. The NCLB itself explicitly provides that participation in NAEP tests is voluntary for all *students* and *schools*, but it contains conflicting provisions regarding voluntary participation by *LEAs*. The NAEP authorization statute (redesignated in 2002 as Section 303 of the Education Sciences Reform Act by P.L. 107-279) states that participation is voluntary for LEAs as well, but ESEA Title I-A provides that the plans of LEAs receiving aid under that program must include an assurance that they will participate in state NAEP tests if selected (Section 1112(b)(1)(F)). Finally, program regulations published by the U.S. Department of Education (*Federal Register*, December 2, 2002) require both LEAs that

(...continued)

tuda_reading_2007/ and http://nationsreportcard.gov/tuda_math_2007/, accessed January 8, 2008.

²² The role of NAEP in “confirming” state test score trends is not explicitly stated in the final statute, but is explicitly mentioned in ED documents, such as the following:

Confirming Progress—Under H.R. 1a small sample of students in each state will participate in the 4th and 8th grade National Assessment of Educational Progress (NAEP) in reading and math every other year in order to help the U.S. Department of Education verify the results of statewide assessments required under Title I to demonstrate student performance and progress.

See *Using the National Assessment of Educational Progress to Confirm State Test Results*, prepared by an Ad Hoc Committee on Confirming Test Results, National Assessment Governing Board, at <http://www.nagb.org>.

receive Title I-A grants, and schools within such LEAs, to participate in NAEP if selected to be among the samples tested (34 C.F.R. § 200.11(b)).

The NCLB authorizes funds specifically for state NAEP tests for FY2002-FY2007: \$72 million for FY2002 and “such sums as may be necessary” for the succeeding years. The NCLB did not extend the authorization for NAEP overall. However, Title III of P.L. 107-279, the National Assessment of Educational Progress Authorization Act, extended the general NAEP authorization through FY2008. The authorization level is \$107.5 million for all NAEP activities (including state assessments), plus \$4.6 million for NAGB, for FY2003, and “such sums as may be necessary” for each of FY2004-FY2008. P.L. 107-279 also redesignated NAEP’s statutory language as Title III of the Education Sciences Reform Act of 2002 (ESRA), but did not otherwise directly or substantially amend the provisions.

For FY2002, the total amount appropriated for all NAEP and NAGB activities was \$111.6 million. This was a large increase over the FY2001 level of \$40 million, primarily as a result of the shift in responsibility for state NAEP costs from states to the federal government. The FY2002 appropriation also included \$2.5 million for the Trial Urban District Assessment described above. The total amount appropriated for NAEP and NAGB was \$94.8 million for each of FY2003 and FY2004, \$94.1 million for FY2005, \$93.1 million for each of FY2006 and FY2007, \$104.1 million for FY2008, \$138.8 million for FY2009, and \$138.8 million for FY2010.

Interpretation by ED of the Expanded Standard and Assessment Requirements of the No Child Left Behind Act

Title I-A Standard and Assessment Requirements

In their reviews of state systems of standards and assessments, peer reviewers (specialists in the areas of standards and assessments who are not federal employees) and ED staff have been considering only various forms of “evidence” submitted by the states which are intended to document that state standards and assessments meet the specific Title I-A requirements outlined earlier in this report; that is, they are not reviewing the assessments themselves.²³ Examples of such “evidence” include results from studies, by test publishers or others, of the degree of alignment between state standards and assessments; evaluations of the validity, reliability, or other aspects of the technical quality of state assessments; state policies on providing native language testing or other accommodations for LEP students, or alternate assessments or other accommodations for students with disabilities; provisions for reporting scores by disaggregated student groups; or data on the extent of actual participation in assessments of LEP students or students with disabilities.

²³ Peer reviewers have relied primarily upon the Department’s *Peer Reviewer Guidance for Evaluating Evidence of Final Assessments Under Title I of the Elementary and Secondary Education Act* (available at <http://www.ed.gov/policy/elsec/guid/cpg.pdf>) to guide their activities. While this document was published before enactment of the NCLB, it remains applicable, at least for the present, mainly because most applicable underlying requirements are essentially unchanged.

On July 5, 2002, ED published regulations on the Title I-A assessment requirements newly adopted under the NCLB.²⁴ Under the provisions of ESEA Title I, Part I, ED was required to establish a “negotiated rulemaking” procedure, as authorized under the Negotiated Rulemaking Act of 1990, in developing regulations regarding the Title I-A standards and assessments requirements.

Under negotiated rulemaking, ED solicits advice from “representatives of Federal, State, and local administrators, parents, teachers, paraprofessionals, and members of local school boards or other organizations involved with the implementation and operation of” Title I-A programs (Section 1901(b)(1)), after which an initial draft of proposed regulations is prepared. ED selects representatives of these organizations to participate in a negotiated rulemaking process, to include persons “from all geographic regions of the United States, in such numbers as will provide an equitable balance between representatives of parents and students and representatives of educators and education officials” (Section 1901(b)(3)(B)).

The selected representatives are to discuss the Department’s draft of proposed regulations, and make any changes to this, consistent with the authorizing statute, on which they can reach consensus. The NCLB provides that “published proposed regulations shall conform to agreements that result from negotiated rulemaking” *unless* “the Secretary reopens the negotiated rulemaking process or provides a written explanation to the participants involved in the process explaining why the Secretary decided to depart from, and not adhere to, such agreements” (ESEA Title I, Section 1902(a)). Thus, ED is encouraged, but not required, to follow the recommendations of the negotiated rulemaking panel, and the process may be viewed primarily as an additional mechanism, beyond publication for comments in the *Federal Register*, of obtaining input on proposed regulations from concerned organizations.²⁵

Significant features of the Department’s final regulations, developed through the negotiated rulemaking process²⁶ and published in the *Federal Register* on July 5, 2002, are described below. In general, the regulations repeat statutory requirements, while clarifying the following points:

(a) content standards can cover multiple grades, but they must include grade-specific “content expectations,” and achievement standards must be grade-specific;

²⁴ *Federal Register*, July 5, 2002, pp. 45038-45047. As is discussed below, proposed amendments to these regulations were published in the *Federal Register* on March 20, 2003.

²⁵ ED’s implementation of the negotiated rulemaking requirement was challenged in federal court. Four organizations (the Center on Law and Education, the National Coalition for the Homeless, the National Law Center on Homelessness, and Designs for Change) and an individual parent charged that parents and students were inadequately represented in the process, particularly in view of the language requiring an “equitable balance between representatives of parents and students and representatives of educators and education officials.” The negotiated rulemaking panel included 17 persons; while only 2 of the 17 persons represented parents specifically, several of the others were parents in addition to representing other groups. On May 22, 2002, the United States District Court for the District of Columbia ruled in favor of the Department of Education and the case was dismissed. An analysis of the legal issues associated with this suit is beyond the scope of this report.

²⁶ In the negotiated rulemaking process, which took place in mid-March 2002, the initial draft proposed regulations were changed in very few significant respects. The primary changes: (a) it was further clarified that the assessment requirements apply only to public schools and their students, not to private (or home) schools; (b) for purposes of disaggregated score reporting, “students with disabilities” would be only those identified under the IDEA (this would exclude students identified only under Section 504 of the Rehabilitation Act); and (c) the criteria to be met by varying local assessments was changed from “equivalent content, rigor, and quality” and “concurrent validity” to “equivalent to one another in their content coverage, difficulty, and quality,” and “comparable validity and reliability.” These changes constituted essentially fine-tuning of certain points of clarification in the draft proposed regulations.

- (b) high school standards must cover what *all* high school students are expected to know and be able to do;
- (c) assessments may include extended or essay response items or ask a student to analyze text or express opinions;
- (d) assessments may include either CRTs or NRTs, although any NRTs used must be augmented to “measure accurately the depth and breadth of” the state’s content standards, provide results expressed in terms of the state’s achievement standards, and be “designed to provide a coherent system across grades and subjects”;
- (e) state assessment systems may include assessments which vary by LEA in some grades,²⁷ and any LEA-selected assessments used to meet the Title I-A requirements must be “equivalent to one another and to state assessments, where they exist, in their content coverage, difficulty, and quality,” “have comparable validity and reliability,” provide “consistent determinations of the annual progress of schools and LEAs within the state,” and produce results which are sufficiently comparable that they can be aggregated;
- (f) LEP, migrant, and homeless students are to be included in the assessment system at all times;
- (g) states are to determine the minimum number of students from specific demographic groups to include in public reports or accountability calculations, to maintain statistical reliability and protect privacy;
- (h) the requirement for dissemination of “itemized score analyses” does not require the release of individual test items;
- (i) states must provide evidence, from test publishers or other “relevant sources,” that their assessment systems are of adequate technical quality to meet each purpose required under Title I-A, and this information can be made available by ED to the public, consistent with applicable federal laws on disclosure of information;
- (j) the assessment requirements apply only to public schools and their students, not to private (or home) schools, although the achievement of private school students who participate in Title I-A must be assessed in *some* manner;
- (k) while states must develop achievement (as well as content) standards in science by 2005-2006, they need not develop specific cut scores for the achievement levels until 2007-2008, when the assessments must be implemented; and
- (l) for purposes of disaggregated score reporting, “students with disabilities” are only those identified under the IDEA,²⁸ although all students with disabilities, whether identified under the IDEA or Section 504 of the Rehabilitation Act, are to be included in assessments and provided with appropriate accommodations.

²⁷ In states that lack authority to require the use of the same assessments statewide (only), the assessment system may consist *entirely* of locally selected assessments.

²⁸ This would exclude students identified only under Section 504 of the Rehabilitation Act.

Evolution of ED Policy Regarding Participation Rates Plus Treatment of Limited English Proficient Students and Certain Students With Disabilities in Assessments and AYP Determinations

ED published supplementary “non-regulatory draft guidance” on the standard and assessment requirements, as well as those related to NAEP participation, on March 10, 2003.²⁹ This document was intended to provide guidance consistent with that in the regulations discussed above, but it is more detailed. This guidance specifically provided that states were to include in their ESEA consolidated application/plan academic content standards in reading/language arts and mathematics for each of grades 3-8, as well as a detailed timeline for meeting subsequent deadlines for the development and implementation of assessments in these subjects and grades, plus standards and assessments at three grade levels in science, by May 1, 2003.

Assessment Participation Rates. More recently, ED officials have published regulations and other policy guidance on participation rates plus the treatment of limited English proficient students and certain students with disabilities in assessments and the calculation of AYP for schools and LEAs, in an effort to provide additional flexibility and reduce the number of schools and LEAs identified as failing to make AYP. On March 29, 2004, ED announced that schools could meet the requirement that 95% or more of students (all students as well as students 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 students in a 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 students 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 students who miss a primary assessment date to take make-up tests, and to determine the minimum size for demographic groups of students 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 Students. In a letter dated February 19, and proposed regulations published on June 24, 2004, ED officials announced two new policies with respect to LEP students.³⁰ First, with respect to assessments, LEP students who have attended schools in the United States (other than Puerto Rico) for less than 12 months must participate in English language proficiency and mathematics tests. However, the participation of such students in reading tests (in English), as well as the inclusion of any of these students’ test scores in AYP calculations, is to be optional (i.e., schools and LEAs need not consider the scores of first year LEP students in determining whether schools or LEAs meet AYP standards). Such students are still considered in determining whether the 95% test participation has been met.

²⁹ See <http://www.ed.gov/topics/topicsTier2.jsp?&top=Policy&subtop=Policy+guidance&subtop2=Elementary+%26+secondary+education&type=T>.

³⁰ See 69 *Federal Register*, pp. 35462-35465, June 24, 2004; and <http://www.ed.gov/nclb/accountability/schools/factsheet-english.html>.

Second, in AYP determinations, schools and LEAs may continue to include students in the LEP demographic category for up to two years after they have attained proficiency in English. However, these formerly LEP students need not be included when determining whether a school or LEA's count of LEP students meets the state's minimum size threshold for inclusion of the group in AYP calculations, and scores of formerly LEP students may not be included in state, LEA, or school report cards. Both these options, if exercised, should increase average test scores for students 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 student groups.³¹ Finally, it was reported in August 2005 that the Secretary of Education had formed a working group to consider better ways to assess the achievement of LEP students for purposes of accountability under the NCLB.³²

Students With Disabilities. Regulations addressing the application of the Title I-A standards and assessment requirements to certain students 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 students "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 students with disabilities.³³ The number of students 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 students (approximately 9% of all students 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 students 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.

ED policy affecting an additional group of students 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 *long-term* phases. It is focused on students with disabilities whose ability to perform academically is assumed to be greater than that of the students 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 students 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).³⁴ Under this

³¹ A bill introduced in the 108th Congress, H.R. 3049, would have authorized the exclusion of scores of LEP students who have resided in the United States for less than *three* years, and would allow formerly LEP students to be included in that group for AYP calculation purposes indefinitely.

³² "Task Force to Gauge Progress of English Language Learners," *Education Daily*, August 10, 2005, p. 1.

³³ This limitation does *not* apply to the administration of alternate assessments *based on the same standards applicable to all students*, for other students with (non-cognitive or less severe cognitive) disabilities.

³⁴ Under current regulations, the short-term policy cannot be extended beyond the 2008-2009 school year.

short-term policy, in eligible states that have not yet adopted modified achievement standards, schools may add to their proficient student group a number of students with disabilities equal to 2.0% of all students assessed (in effect, deeming the scores of all of these students to be at the proficient level).³⁵ This policy would be applicable only to schools and LEAs that would otherwise fail meet AYP standards due solely to their students 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 students with disabilities on these assessments, subject to a 2.0% (of all assessed students) 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 students with disabilities who are unlikely to achieve grade level proficiency within the current school year, but who are not among those students with the most significant cognitive disabilities (whose situation was addressed by an earlier set of regulations, discussed above). For this second group of students 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 students. The standards must provide access to grade-level curriculum, and not preclude affected students from earning a regular high school diploma.

As with the previous regulations regarding students with the most significant cognitive disabilities, there would be no direct limit on the number of students who take alternate assessments based on modified achievement standards. However, in AYP determinations, student 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 students tested at the state or LEA level (i.e., an estimated 20% of students with disabilities); such scores in excess of the limit would be considered “non-proficient.” As with the 1.0% cap for students 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 students 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 students tested at a state or LEA level.³⁶ In particular, states are no longer allowed to request a waiver of the 1.0% cap regarding students 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 students. This prohibits the previously common practice of setting higher “n” sizes for students with disabilities or LEP students than for other student groups. Second, when students take state assessments multiple times, states and LEAs may use the highest score for students who take tests more than once. Finally, as with LEP students, states

³⁵ This would be calculated based on 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 students who are proficient must also be reported to parents and the public.)

³⁶ 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.

and LEAs may include the test scores of former students with disabilities in the disability subgroup for up to two years after such students have exited special education.³⁷

Thus, eligible states and LEAs will be allowed to count as “proficient or above” in AYP determinations the proficient or higher scores of up to 1.0% of all tested students on “alternate assessments based on alternate achievement standards,” and of up to an additional 2.0% of all tested students on “alternate assessments based on modified achievement standards.” For both groups, there is no limit for individual schools on the percentage of students in either of these categories, and there is no limit on the number or percentage of students to whom either type of alternate assessment may be administered.

Regulations Published in October 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). Many of the regulations deal with policy areas other than assessments and related accountability topics. Many of the proposed regulations clarify previous regulations or codify as regulations policies that had previously been established through less formal mechanisms (such as policy guidance or peer reviewer guidance). The regulations relevant to assessments are briefly described below.

The October 2008 regulations clarify that assessments required under Title I-A may include multiple formats as well as multiple 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 through inclusion of a variety of indicators *other than* standards-based assessments in reading, mathematics, and science. Also, states are required to include the latest results from the most recent National Assessment of Educational Progress (NAEP) assessments on their state and LEA performance report cards. Further, ED policies regarding growth models of AYP are codified in regulations (previously they were published only in policy guidance and peer reviewer guidance documents).

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 student groups in AYP determinations, especially at the school level. States must also report on the number of students 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.

³⁷ In such cases, the former students with disabilities would *not* have to be counted in determining whether the minimum group size was met for the disability subgroup.

ED Review to Determine Whether States Meet 2005-2006 Assessment Requirements

State educational agencies (SEAs) must provide evidence from a test publisher or other relevant source that their assessments are of adequate technical quality for the purposes required under Title I-A.³⁸ Ongoing peer reviews of state assessment programs are being conducted to determine if they meet NCLB requirements to test students in each of grades 3-8 in reading and mathematics, and to adopt content and achievement standards in science. A letter sent to chief state school officers in April 2006 by the Assistant Secretary for Elementary and Secondary Education³⁹ describes the current categories of results from the state reviews. These categories, and the number of states in each category as of January 7, 2009, with respect to the *mathematics and reading assessments*, include the following:

- *Full Approval.* Meets all statutory and regulatory requirements (39 states: Alabama, Alaska, Arizona, Arkansas, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Montana, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Virginia, Washington, West Virginia, and Wisconsin).
- *Approval Pending.* A limited number (generally one to three) of fundamental components of the state assessment system fail to meet the statutory or regulatory requirements (11 states: California, Hawaii, Mississippi, Nebraska, Nevada, New Hampshire, New Jersey, Oregon, Texas, Vermont, and Wyoming, plus the District of Columbia and Puerto Rico).

Peer reviews are continuing for the states whose assessment systems for reading and mathematics have not yet been fully approved. With respect to *science*, the assessments of only 11 states (Delaware, Indiana, Kentucky, Louisiana, Massachusetts, New York, Ohio, Tennessee, Utah, Washington, and Wisconsin) had been approved as of January 7, 2009.

Both before and after the NCLB, the ESEA authorized sanctions for states failing to meet the deadlines for adopting standards and assessments. The 1994 version provided that the Secretary of Education *may* withhold funds for state administration plus program improvement from states failing to meet any of the Title I-A state plan requirements, including those related to standards and assessments (Section 1111(d)(2)). As amended by the NCLB, the ESEA provides that the Secretary *shall* withhold 25% of funds otherwise available for state administration and program improvement activities from states that fail to meet the 1994 requirements, and *may* withhold

³⁸ Several statutory constraints have been placed on the authority of the U.S. Secretary of Education to enforce these standard and assessment requirements. First, the ESEA states that nothing in Title I shall be construed to authorize any federal official or agency to “mandate, direct, or control a State, local educational agency, or school’s specific instructional content, academic achievement standards and assessments, curriculum, or program of instruction” (Sections 1905, 9526, and 9527). Second, states may not be required to submit their standards to the Secretary (Section 1111(b)(1)(A)) or to have their content or achievement standards approved or certified by the federal government (Section 9527(c)) in order to receive funds under the ESEA, other than the (limited) review necessary in order to determine whether the state meets the Title I-A technical requirements. Finally, no state plan may be disapproved by ED on the basis of specific content or achievement standards or assessment items or instruments (Section 1111(e)(1)(F)).

³⁹ See <http://www.ed.gov/admins/lead/account/saapr3.pdf>.

additional state administration funds for failure to meet new assessment requirements adopted under the NCLB. States in the Approval Pending category face the possibility of the loss of Title I-A administrative funds, plus the additional sanctions of limitations on approval of flexibility requests, and heightened oversight by ED. According to ED, funds withheld (from the SEA) would be distributed to LEAs in the state. In addition, states that persistently and thoroughly fail to meet the standards and assessment requirements over an extended period of time potentially may be subject to elimination of their Title I-A grants altogether, since they would be out of compliance with a basic program requirement.

Implementation of the NAEP Requirements

In the period since enactment of the NCLB, a number of steps have been taken toward implementation of the new requirements for state participation in NAEP. First, the schedule for test administration has been revised to provide for administration of state NAEP tests in 4th and 8th grade reading and mathematics every two years, beginning with the 2002-2003 school year (spring 2003). Initial NAEP 4th and 8th grade reading and mathematics results for all states were released in November 2003. Subsequent rounds of NAEP tests was administered in all states in 2005 and 2007. Further, as is discussed in a later section of this report, the NAGB has published a report, “Using the National Assessment of Educational Progress to Confirm State Test Results,” which examines issues related to the possible use of state NAEP results to “confirm” trends in state assessment results.

Several changes to NAEP policies and practices have been implemented that are supportive of, or were adopted primarily in response to, the expanded role for NAEP under the NCLB.⁴⁰ In recognition of the increased emphasis on measurement of performance gaps among different demographic groups of students in the NCLB, more questions are being added at the upper and lower ends of the difficulty range, so that achievement gaps among student groups can be more reliably measured. In addition, studies are being conducted of possible ways to adjust sampling strategy in order to assure adequate numbers of students in the various demographic groups referenced in the NCLB.

At the same time, a number of administrative adjustments are being implemented that are intended to reduce required student sample sizes in the aggregate (e.g., the main NAEP state and national student samples will be combined for the first time), although samples of students will likely be increased in small and/or sparsely populated states in order to enhance the precision of results. Efforts are being made to minimize time demands, with a goal of reporting results of reading and mathematics assessments within six months of test administration.

Special issues arise with respect to Puerto Rico, which is treated as a state under ESEA Title I-A but did not participate in state NAEP tests prior to the enactment of the NCLB. Questions have been raised about the comparability of tests administered in different languages, especially in reading. NAEP tests in mathematics were administered to 4th- and 8th-grade students in Puerto Rico in 2003 and 2005, and results from both test administrations have been recently released.⁴¹

⁴⁰ See *NAGB Adopts Policies to Implement the No Child Left Behind Act of 2001* at <http://www.nagb.org/>, plus <http://nces.ed.gov/nationsreportcard/about/current.asp>.

⁴¹ See <http://nces.ed.gov/nationsreportcard/puertorico/>, visited on April 16, 2007.

Finally, state NAEP tests are now administered by contractors, rather than (as in the past) local teachers; there is a full-time NAEP coordinator in every state, and a State Service Center has been established to support these coordinators; and NAGB has established procedures for limited public access to NAEP test items, and for submission, review, and resolution of complaints about NAEP tests by parents and other members of the public.

Issues Regarding the ESEA Title I-A Student Assessment Requirements

What Types of Assessments Meet the Expanded Assessment Requirements?

As described above, the NCLB includes explicit reference to a number of criteria that state assessments must meet in order to comply with the ESEA Title I-A requirements. However, the statute does not appear to directly or explicitly address two major issues with respect to the assessments: (a) whether qualifying state assessment systems must include only CRTs or whether they may include a mix of CRTs and NRTs, as long as the latter are modified to provide the required linkage to state content and achievement standards; and (b) whether qualifying state assessment systems must include only assessments that are the same statewide (except in states that lack authority to require statewide assessments) or whether they may include a mixture of statewide and locally varying assessments, as long as the latter are deemed to be “equivalent” and adequately linked to state content and achievement standards. It is stated that assessments must “be the same academic assessments used to measure the achievement of all children” (Section 1111(b)(3)(C)(i)), but the implications of this provision are ambiguous in cases where a state has no assessment to measure the achievement of all children in certain grades.

Arguably, criterion-referenced assessments which are administered to all public school students statewide in the relevant grades are most fully consistent with the requirements which are explicitly stated in Title I-A. Only CRTs are designed comprehensively and “from the ground up” to measure student achievement with respect to specific content and academic achievement standards. While certain NRTs may be somewhat related to state standards in their generic form, with substantial overlap in test items with CRTs, and more closely related if modified specifically for this purpose—as would be required under the regulations—they are nevertheless initially designed primarily for the purpose of ranking and sorting students, not for the purpose of determining whether students meet state-determined achievement levels. In fact, it is not yet clear whether modified versions of assessments designed *initially* as NRTs can indeed meet the Title I-A requirements for linkage with state content and performance standards; some states, such as California, have attempted to meet the 1994 assessment requirements through use of modified NRTs, but no such assessments have yet been fully approved by ED.⁴²

Similarly, assessments that are the same statewide would seem to most fully meet the purposes of Title I-A, especially with respect to the use of assessment results to determine whether schools or

⁴² However, ED has approved the assessment systems of three other states (Delaware, Indiana, Missouri) where state-specific tests were reportedly designed from the beginning to produce both criterion-referenced and norm-referenced results.

LEAs meet state standards of adequate yearly progress (AYP). The best way to assure that assessments of the extent to which students meet state achievement standards are equivalent and consistent statewide is to use the same assessments throughout the state. This is especially important in view of the use of assessment results to determine whether schools or LEAs meet AYP standards, and the need to aggregate local results to determine whether states overall meet such requirements. Establishing equivalence among varying local tests might be possible, but is likely to be very difficult. According to a National Research Council report, “Under limited conditions it may be possible to calculate a linkage between two tests, but multiple factors affect the validity of inferences that may be drawn from the linked scores. These factors include the context, format, and margin of error of the tests; the intended and actual uses of the tests; and the consequences attached to the results of the tests.”⁴³ Further, there is no precedent for allowing states to meet Title I-A assessment requirements through use of different assessments in different LEAs—except for the two states that may lack authority to establish statewide assessments, no states have been allowed to meet the 1994 standard and assessment requirements through the use of locally varying assessments.

Articulation between the tests used in different grades, and coherence of the overall assessment system, are also important concerns. If, for example, statewide tests are used in some grades but locally varying tests in other grades, or if CRTs are used in some grades and modified NRTs in others, this would likely create significant articulation difficulties, with variations from grade to grade in the proportion of students meeting state standards which result solely from the assessment instrument used, separate from any underlying differences in achievement levels.

Criteria established in the regulations published by ED for mixed state assessment systems are relatively demanding. Any NRTs used must be augmented to “measure accurately the depth and breadth of the State’s academic content standards” (34 C.F.R. § 200.3(a)(2)(ii)(A)), and have results expressed in terms of the state’s achievement standards; and any LEA-selected assessments used to meet the Title I-A requirements must be of “equivalent to one another ... in their content coverage, difficulty and quality,” have “comparable validity and reliability,” and produce results which can be aggregated (34 C.F.R. § 200.3(c)(2)). If these criteria were to be strictly interpreted by ED in the assessment review process, it is likely to be very difficult for mixed state assessment systems to be approved. However, opponents of proposals to allow states to meet the Title I-A requirements through mixed assessment systems are concerned that ED’s review process may not be very strict, and that in some states, systems may be approved which are not well aligned with state standards or are not consistent among LEAs statewide, at least in certain grades, with the result that the standards for determining whether schools are meeting AYP standards would significantly vary among LEAs.

In contrast, proponents of a relatively high degree of state flexibility in meeting the Title I-A requirements through mixed assessment systems argue that this will minimize federal influence and intrusion, recognize state primacy in selecting assessment systems which meet their needs, minimize costs, and still meet the purposes of Title I-A because of the criteria which such systems would have to meet. Proponents of allowing the use of modified NRTs to meet the requirements, at least for some grades, argue that the differences between NRTs and CRTs have more to do with how test results are analyzed and presented than with the test items themselves. The fact that several states currently use a mix of statewide CRTs in some grades and NRTs in others, or

⁴³ National Research Council, *Uncommon Measures: Equivalence and Linkage Among Educational Tests* (Washington: National Academies Press, 1998), p. 5-4.

statewide tests of either type in some grades and locally varying tests in others, may indicate that such mixed assessment systems meet important educational needs and goals, as perceived by the states themselves.

How Strict Is ED's Review of State Assessment Systems?

As indicated by the relevant policy guidance and the published communications to states, peer reviewers and ED staff appear to have been conducting relatively rigorous and detailed reviews of the "evidence" submitted by states regarding whether their assessment systems meet the ESEA's requirements. The features which the Title I-A statute requires state assessment systems to exhibit are themselves numerous and relatively detailed, and a substantial implementation of them is likely to involve somewhat exhaustive review. The assessment reviews have focused especially on issues regarding testing, score reporting, and inclusion in accountability systems for LEP students and those with disabilities. While there are complex issues and considerations in these areas, they are not being raised solely, and possibly not even primarily, because of the Title I-A requirements. For example, while there are general guidelines, applicable under Title VI of the Civil Rights Act of 1964 to any LEA receiving federal grants, regarding the use of an appropriate language and/or other accommodations for assessment of LEP students,⁴⁴ and requirements under the IDEA for alternate assessments where necessary for students with disabilities, it is largely in the context of Title I-A that such requirements are having an impact because of the scrutiny currently being given to whether state assessments meet the Title I-A requirements.

Although it may be questioned whether ED should be reviewing state assessment systems in such detail, this scrutiny may be necessary to enforce Title I-A's statutory requirements, and might also be necessary to establish outcome accountability for all major groups of disadvantaged students. If, for example, significant numbers of LEP students or those with disabilities were excluded from state assessments, or were not provided with appropriate accommodations, then it would be impossible to determine whether they, along with the student population in general, are adequately meeting state performance goals. Such inclusive assessment, combined with disaggregated score reporting, becomes increasingly important as focus shifts toward outcome measures to assure accountability for use of federal aid funds, and Title I-A programs are increasingly conducted in a schoolwide program format, in which services are not targeted on the individual students with lowest achievement in a participating school.⁴⁵

Although detailed review by ED of state assessment systems may raise concerns about undue federal influence over this fundamental aspect of state and local public education systems, there are many statutory limitations on the review process. As noted earlier, the federal government is prohibited from mandating, directing, or controlling a state's, LEA's, or school's standards, assessments, or curriculum; states may not be required to submit their standards to ED; and no state plan may be disapproved by ED on the basis of specific content or achievement standards or

⁴⁴ See U.S. Department of Education, Office for Civil Rights, "Testing the Academic Educational Achievement of Limited English Proficient Students," in *The Use of Tests When Making High-Stakes Decisions for Students: A Resource Guide for Educators and Policymakers*, a draft document dated July 6, 2000, available on the Internet at <http://www.ed.gov/legislation/FedRegister/other/2000-4/121500b.html>.

⁴⁵ There are two basic types of Title I-A programs. *Schoolwide programs* are authorized when 40% or more of the students in a school are from low-income families. In these programs, Title I-A funds may be used to improve the performance of all students in a school, and there is no requirement to focus services on only the most disadvantaged students. The other major type of Title I-A service model is the *targeted assistance school* program, under which services are generally limited to the lowest achieving students in the school.

assessment items or instruments. Nevertheless, the degree of federal influence over at least the broad parameters of state student assessment systems—such as grades and subject areas tested, inclusion of special needs student groups, disaggregated reporting of results—has increased under the NCLB.

The rigor of ED's assessment review process, and the flexibility of the assessment regulations, will also likely influence the extent to which states meet the expanded requirements on schedule. A Government Accountability Office report published in 2002 identified four additional factors which have influenced the pace of state compliance with Title I-A assessment requirements: "(1) the efforts of state leaders to make Title I compliance a priority; (2) coordination between staff of different agencies and levels of government; (3) obtaining buy-in from local administrators, educators, and parents; and (4) the availability of state level expertise."⁴⁶

What Is the Cost of Developing and Implementing the Required Assessments, and to What Extent Will Federal Grants Be Available to Pay for Them?

The addition of requirements to conduct annual reading and mathematics assessments in at least four more grades than required previously, and to include standards and assessments at three grade levels in science, has required most states to significantly increase their expenditures for standard and test development and administration. As indicated earlier, it is very difficult, if not impossible, to specify all of these potential costs with precision.

The NCLB conference report directed the Government Accountability Office to conduct a study of the costs to each state of developing and administering the assessments required under Title I-A, both overall and for each of fiscal years 2002-2008. In 2003, GAO published a report (*Title I: Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies*, GAO-03-389) that discussed issues related to potential costs of meeting the NCLB assessment requirements, and provided a range of alternative cost projections. GAO based its conclusions on a survey of assessment practices in all states, and a detailed examination of the costs of assessment development and administration in seven states.

According to the GAO, the level of state costs for assessment development and administration, as well as the relationship between those costs and funding provided by the NCLB's assessment development grants, depends primarily on the kinds of test questions states choose to utilize: multiple choice, open-ended (essay questions), or a combination of these. Tests with questions that elicit open-ended responses, which require people who can evaluate students' responses, are much more expensive to administer and score than multiple-choice questions that can be scored by computers. Over the period of FY2002-FY2008, in comparison to a total of the annual minimum assessment development grant appropriations of \$2.7 billion, GAO estimated that it would cost states \$1.9 billion to meet the NCLB assessment requirements using only multiple choice tests, \$5.3 billion using a mixture of multiple choice and open-ended test items in all states, and \$3.9 billion if states use the same mixture of multiple choice and open-ended test items as in the recent past. It should be noted that this study considered only the projected state-level costs of developing standard assessments on reading, mathematics, and science, and not costs for

⁴⁶ U.S. Government Accountability Office (GAO), *Title I, Education Needs to Monitor States' Scoring of Assessments*, GAO-02-393, April 2002, p. 13.

developing alternate assessments for students with disabilities, or English language proficiency assessments for LEP students, or possible increased costs for LEAs.⁴⁷

The NCLB authorizes \$400 million for FY2002, and “such sums as may be necessary” through FY2008, for state assessment development and administration grants. The administration, although not the development, of assessments newly required by the NCLB (grades three through eight reading and mathematics assessments, plus science assessments at three grade levels) may be delayed by one year for each year that the minimum amounts (e.g., \$400 million for FY2007) are *not* appropriated. Thus far, the minimum amount has been appropriated for each of FY2002-FY2008. The available information on direct, state-level expenditures for testing programs indicates that the “trigger” appropriation levels for state assessment grants are, in the aggregate, similar to these estimates.⁴⁸ They are also either similar to, or substantially below, the test development and administration costs projected by GAO (above), depending on assumptions regarding types of test items used.⁴⁹

It is probable that the costs of meeting the expanded assessment requirements have varied widely from state to state, not only because of differences in state size, but also particularly because of substantial differences in the extent to which state-mandated tests in reading and mathematics were already being administered to all students in grades three through eight, or tests in science for students in selected grade ranges, and whether the tests met the Title I-A technical requirements of alignment with state standards, inclusion of all student groups, etc. Assessment development costs may also be reduced through cooperative arrangements among some states to jointly develop certain assessments, such as the New England Common Assessment Program involving New Hampshire, Rhode Island, and Vermont.

With respect to the distribution among the states of funds for test development and administration, the NCLB provides for allocation of a substantial share of these funds in equal amounts to each state, with the remainder allocated in proportion to children and youth aged five to 17 years. The allocation formula does not recognize the substantial variation in the extent to which states may already administer the required assessments, and therefore face varying levels of additional assessment program costs. The allocation of funds by formula to all states, regardless of the current status of their state assessment policies and programs, might recognize that all states face ongoing costs, and might possibly reward states which have already adopted relatively extensive assessment programs. At the same time, the formula does not target funds on the states with the greatest needs.

⁴⁷ Earlier, two organizations attempted during 2001-2002 to estimate costs for states of meeting assessment requirements similar to those of the NCLB. In 2001, the National Association of State Boards of Education (NASBE) estimated that the new grade 3-8 assessments (only) would cost states between \$2.7 and \$7.0 billion in the aggregate over a seven-year period <http://www.nasbe.org/Archives/cost.html>. On an annual basis, if costs were equally distributed across the seven years, this would represent a range of \$386 million to \$1 billion per year. In contrast, Accountability Works, a private consulting firm, estimated that the annual cost of meeting all of the new assessment requirements in the NCLB would range from approximately \$312 million to \$388 million for each of 2002-2003 through 2007-2008 <http://www.schoolreport.com/AWNCLBTestingCostsStudy.pdf>.

⁴⁸ The \$400 million “trigger” amount (and actual appropriation) for FY2007 is 95% of the estimated aggregate expenditure level for FY2001 (discussed earlier in this report) of \$422.8 million.

⁴⁹ Estimates of the state-level costs of developing and administering assessments required by the NCLB are becoming available for a limited number of individual states. For example, a study published in September 2005 for Virginia <http://www.pen.k12.va.us/VDOE/nclb/coststudyreport-state.pdf>, concluded that estimated assessment costs for this state ranged from \$7.3-\$8.2 million for each of the 2004-2005 through 2007-2008 school years. These amounts are somewhat less than the assessment grants to Virginia of \$8.5-\$8.8 million for FY2004-FY2005.

What Might Be the Impact of the Requirement for Annual Assessment of English Language Proficiency of LEP Students?

As noted earlier, the NCLB requires states to provide that their LEAs will annually assess the English language proficiency of their LEP students. This is separate from the requirements regarding treatment of LEP students in states' general assessment systems—that is, the requirement that LEP students be included in such assessments, in which they are to be assessed in a valid and reliable manner and provided with “reasonable” accommodations, in the language and form most likely to yield accurate and reliable information on what they know and can do in academic content areas (in subjects other than English itself), with students who have attended schools in the United States (excluding Puerto Rico) for three or more consecutive school years to be assessed in English.

In contrast to such requirements regarding treatment of LEP students in states' general assessment systems, the separate requirement for annual assessments of English language proficiency lacks specificity. There are no statutory details regarding technical characteristics of the tests—except that the assessment must consider the students' oral, reading, and writing skills—and (thus far) no policy guidance from ED. It is also somewhat ambiguous regarding whether states or LEAs are ultimately or primarily responsible for implementing this requirement.

Depending on possible future regulations or policy guidance from ED, this new requirement may lead to relatively little change in current activities in LEAs. Although comprehensive and detailed surveys of such assessment practices are not currently available, there is substantial evidence that LEAs in general already assess the English language proficiency of LEP students for purposes of placement in instructional programs, determination of needed accommodations in general assessment programs, evaluation of programs targeted on LEP students, and movement of students from special programs to mainstream instruction. While a variety of assessment methods are used, including teacher observation and home language surveys, recent surveys indicate that a large majority of LEAs administer formal English language proficiency tests to their LEP (or potentially LEP) students.⁵⁰ Policy guidance from ED's Office for Civil Rights indicates that such assessments should be undertaken especially, but not only, for purposes of assigning students to instructional programs targeted at LEP students, determining the timing of transition to regular or mainstream instruction for such students, and evaluating the effectiveness of special programs for LEP students; although this guidance is unspecific regarding the type of assessment LEAs should use.⁵¹

In addition, LEAs participating in the new English Language Acquisition program authorized under ESEA Title III, Part A, must report annually the number and percentage of participating students who attain English proficiency, as determined by a “valid and reliable assessment of English proficiency” (Section 3121(a)(3)). If ED's future policy guidance is consistent with the statute's lack of specificity regarding the new Title I-A requirement, there may be little required change in LEA activities as a result of the requirement.

⁵⁰ See National Research Council, *Improving Schooling for Language-Minority Children: A Research Agenda* (Washington: National Academies Press, 1997), pp. 115-116.

⁵¹ See <http://www.ed.gov/about/offices/list/ocr/docs/laumemos.html>.

What Might Be the Impact of Requiring State Participation in NAEP?

Possible Influence on State Standards and Assessments Arising from (Marginally) Increased Stakes

Two key characteristics of the NAEP program since its inception have been: (1) the content frameworks, upon which test items are based, have been independent of the content standards adopted by any state or national organization; and (2) the “stakes” associated with performance on the tests have been extremely low. The NCLB’s requirement for states to participate in NAEP in order to retain eligibility for ESEA Title I-A grants, with the implicit purpose of using the results to “confirm” performance trends on state-selected assessments, has potential implications for both of these characteristics of NAEP.

Previously, the only “stakes” associated with state participation in NAEP have been the symbolic ones arising from public dissemination of NAEP results for states that chose to participate and which allowed their assessment results to be published. Public attention to these results, among persons other than selected policymakers, researchers, and policy analysts, seems to have been limited. The NAEP scores have had no impact on state finances or eligibility for federal programs or services.

While state involvement with NAEP will change significantly under the NCLB, the stakes for states will remain relatively low. State results will be published as an implicit “confirmation” of test score trends on state assessments, but these NAEP scores will still have no direct impact on state eligibility for federal assistance. Provisions of the House-and Senate-passed versions of the NCLB for state bonuses and sanctions based in part on NAEP score trends were eliminated from the conference version. Under the NCLB as enacted, ED is required to establish a peer review process to evaluate whether states have met their statewide AYP goals; states which fail to meet them are to be listed in an annual report to Congress, and technical assistance is to be provided to states that fail to meet their goals for two consecutive years. State NAEP scores will likely be considered in this review process. However, there is no provision for state bonuses or sanctions under this procedure, only publicity and technical assistance. This increases the “stakes” associated with state NAEP performance, but only to a very modest degree.

Nevertheless, even a small increase in the stakes associated with state performance on NAEP tests attracts attention to the possibility that NAEP frameworks and test items might influence state standards and assessments. To the extent that the required participation in NAEP increases attention to state performance on these tests, there might be a basis for concern that states would have an incentive to modify their curriculum content standards to more closely resemble the NAEP test frameworks. To counteract this potential problem, the NCLB prohibits the use of NAEP assessments by agents of the federal government to influence state or LEA instructional programs or assessments. However, subtle, indirect, and/or unintended forms of influence may be impossible to detect or prohibit. A “White Paper” policy statement released by NAGB on May 18, 2002, attempts to distinguish between “active attempts ... to persuade others to adopt NAEP policies, procedures, or content,” which are prohibited, and “influence by good example,” which (according to this document) is not.

Voluntary Participation by LEAs, Schools, and Students

Might a conflict arise between the requirement for NAEP participation by states participating in ESEA Title I-A and the provision that participation in NAEP tests is voluntary for all students, schools, and possibly LEAs? While participation by states, LEAs, schools, and students was voluntary under previous *federal* law and policy, states or LEAs were not prohibited from requiring participation by LEAs, schools, or students under their own laws or policies. However, as noted earlier (see the section of this report titled “NAEP Provisions in the No Child Left Behind Act”), there are conflicting statutory and regulatory provisions regarding participation in NAEP tests by LEAs and schools which may be selected for NAEP test administration.

Some have expressed concern that the new provisions regarding voluntary participation in NAEP might lead to two types of difficulties: (a) in a time of likely increased assessment activity for students nationwide, resistance to participation in NAEP might grow to an extent that it threatens the quality of the national sample of tested students and makes it difficult to maintain trend lines; and (b) more specifically, states might be stuck between a requirement to participate in NAEP and an inability to recruit a sufficiently large sample of LEAs, schools, and students to participate in order to produce valid and reliable assessment results. In the past, some states have attempted to participate in NAEP but found themselves unable to induce sufficient numbers of LEAs or schools to do so.⁵²

The primary counter to this concern is that the policies regarding voluntary participation in NAEP have changed only modestly. As far as federal policies are concerned, participation has already been voluntary at all levels. While states or LEAs previously could have mandated participation by LEAs, schools, or students, apparently they generally attempted to avoid doing so. Thus, in practice, little may have changed. There may nevertheless be some cause for concern, with the expansion of NAEP to states that have not previously chosen to participate.

Can NAEP Results Be Used to “Confirm” State Test Score Trends?

An unstated, but clearly implicit, purpose of the state NAEP participation requirement is to “confirm” trends in student achievement, as measured by state-selected assessments by comparing them with trends in NAEP results. Some have questioned whether it is possible or appropriate to use results on one assessment to “confirm” results on another assessment which may have been developed very differently, and what form this “confirmation” might take.

State assessments vary widely in terms of several important characteristics, such as the content and skills which they are designed to assess, their format, and modes of response. They are likely to continue to vary widely, especially as the final assessment regulations allow the use of both CRTs and modified NRTs, as well as locally varying assessments. As a result, some state assessments will be much more similar to NAEP in these important respects than others, and there will be consequent variation in the significance of similarities or differences when comparing trends in NAEP versus state assessment score trends for students.

⁵² In 2000, 48 states (all except Alaska and South Dakota) initially stated their intention of participating in state NAEP, although ultimately only 41 did so. States which intended to participate, but did not do so, reportedly were unable to recruit sufficient number of LEAs and schools. See “Test Weary Schools Balk at NAEP,” *Education Week*, February 16, 2000.

If, for example, a state test is closely aligned to state curriculum content standards which are substantially different from the content embodied in NAEP assessment frameworks, and if instruction is modified to better match the state standards, then it is possible that scores on the state assessment will rise while those on NAEP will be flat or even decline. NAEP frameworks are designed with the intention that they substantially reflect state standards *on average*; according to a recent analysis, “States vary in the amount that their assessment domains [i.e., the content and skills covered by the assessments] overlap with NAEP. For some, there is almost complete overlap. For others, the overlap is modest.”⁵³ Other major differences between NAEP and state assessments include (a) the time of year when tests are administered; (b) relative placement of cut scores for achievement levels; (c) the (often high, but varying) stakes associated with state assessments versus the low stakes associated with NAEP; and (d) test format and modes of response.

As for the form which a comparison of NAEP and state test scores might take, two obvious candidates are average raw scores and the percentages of students at different achievement levels (basic, proficient, etc.). While these are key benchmarks, either alone, or even both, might overlook important changes or differences in the distribution of student scores. For example, the scores of several students might improve but not by enough to raise them above the cut score for the next highest achievement level. As noted above, the NAGB has published a report, “Using the National Assessment of Educational Progress to Confirm State Test Results,” whose authors argue that state NAEP scores can be used as evidence to confirm the general trends in scores on individual state assessments, although such confirmation should not be viewed as, or take the form of, a strict statistical “validation” of state test results. They address the question of whether comparisons should be based on raw scores or percentages of students at various achievement levels by recommending a new method of comparison which considers changes and differences in the overall achievement score distribution, not focusing solely on overall averages or cut scores.⁵⁴

Likely Benefits and Costs of the Expanded Title I-A Assessment Requirements

This report concludes with a review of major potential benefits and costs of the expanded student assessment requirements of ESEA Title I-A. The primary benefit from annual administration of a consistent series of standards-based tests would be the provision of timely information on the performance of students, schools, and LEAs, throughout most of the elementary and middle school grades. While a majority of students have already been taking assessments in many of grades 3-8, these have been typically a mix of CRTs and NRTs, state-mandated and locally selected tests, with no provision that most of these are either equivalent statewide or aligned to state content and achievement standards. Even under the broadest interpretation of ED’s draft policy guidance, which would allow states to use modified NRTs in addition to CRTs, and locally varying tests which are deemed to be equivalent, the resulting state assessment systems would be more coherent, consistent, and well articulated than the current systems in most states. The availability of such consistent, annual assessment results would be of value for both diagnostic

⁵³ Mark D. Rebase, “Using NAEP to Confirm State Test Results: Opportunities and Problems,” in *No Child Left Behind: What Will It Take?* (Washington: Thomas B. Fordham Foundation, February 2002), p. 14.

⁵⁴ See the report for details, available at <http://www.nagb.org>.

and accountability purposes. The resulting assessment systems would also continuously emphasize the importance of meeting state standards as embodied by the assessments.

These expanded requirements regarding student assessments—and school, LEA, and state accountability based on performance on the assessments—have been enacted in the context of a broader strategy, also initiated in the 1994 ESEA amendments and expanded by the NCLB, which involves increased state and local flexibility in the use of federal education assistance funds.⁵⁵ Under this strategy, accountability for appropriate use of federal aid funds is to be established more on the basis of student performance outcomes, and less on prescribed procedures or targeting of resources, than in the past. Such a strategy implicitly relies heavily on high quality, current, detailed, and widely disseminated information on student achievement as a basis for outcome accountability policies and procedures. It is desirable that achievement data be as comparable and current as possible while not compromising the primacy of states and LEAs in setting K-12 education policy.

According to the ED publication, “Testing for Results, Helping Families, Schools and Communities Understand and Improve Student Achievement,”⁵⁶ annual standards-based assessments “will empower parents, citizens, educators, administrators and policymakers with data ... in annual report cards on school performance and on statewide progress.” Further:

The tests will give teachers and principals information about how each child is performing and help them to diagnose and meet the needs of each student. They will also give policymakers and leaders at the state and local levels critical information about which schools and school districts are succeeding and why, so this success may be expanded and any failures addressed.... A good evaluation system provides invaluable information that can inform instruction and curriculum, help diagnose achievement problems and inform decision making in the classroom, the school, the district and the home. Testing is about providing useful information and it can change the way schools operate.”

At the same time, the expanded Title I-A assessment requirements might lead to a variety of costs, or unintended consequences, in both financial and other forms. One such “cost” is expanded federal influence on state and local education policies. Assuming that states will continue to implement them in order to maintain Title I-A eligibility, then assessment requirements attached to an aid program focused on *disadvantaged students* are broadly influencing policies regarding standards, assessments, and accountability affecting *all students* in the participating states. This represents a substantial increase in federal influence in the assessment and accountability aspects of K-12 education policy.

In the majority of states that did not previously mandate standards-based assessments in each of grades 3-8, their policies may have resulted primarily from cost or time constraints, or the states may have determined that annual testing of this sort is not educationally appropriate, or at least that its benefits are not equal to the relevant costs. These costs may include not only the direct costs of test development, administration, scoring, reporting, etc., not all of which may be paid through federal assessment grants, but also an increased risk of “over-emphasis” on preparation for the tests, especially if the tests do not adequately assess the full range of knowledge and skills which schools are expected to impart. The authors of a recent study of the effects of high-stakes

⁵⁵ These provisions are described in CRS Report RL31284, *K-12 Education: Highlights of the No Child Left Behind Act of 2001 (P.L. 107-110)*.

⁵⁶ See <http://www.ed.gov/nclb/accountability/ayp/testingforresults.html>.

assessment policies in 18 states have posited an “Uncertainty Principle,” which may be relevant to such concerns: “The more important that any quantitative social indicator becomes in social decision-making, the more likely it will be to distort and corrupt the social process it is intended to monitor.”⁵⁷ At the least, annual testing of students in grades 3-8 would increase the importance of having tests that are well designed and closely linked to state content and achievement standards which are truly challenging.

Nevertheless, even within the specific realm of standards and assessments, federal influence remains limited in several important respects. With the exception of the limited role of state NAEP tests, the standards and assessments are totally selected by the states. ED is not authorized by the NCLB to review the substance of any state standards, and no state plan may be disapproved by ED on the basis of specific content or achievement standards or test items or instruments.

Ultimately, whether increased federal influence in certain respects, combined with less federal control over certain other aspects of state and local use of federal aid funds, is a “balanced tradeoff” is a subjective political judgment. The key analytical point is that the increase in federal influence is constrained, and is balanced by a decrease of federal influence in certain other respects.

Glossary of Selected Terms Used in This Report

Criterion-Referenced Test (CRT)	“Criterion-referenced” tests measure the extent to which students have mastered specified content (content standard) to a predetermined degree (achievement standard). A typical criterion-referenced test result is that a 4 th grade student’s achievement in mathematics is at the “proficient” level, which is above a “basic” level, but below an “advanced” level. Most state-developed assessments, such as the Connecticut Mastery Test, the North Carolina End-of-Grade Tests, or the Texas Assessment of Academic Skills, are criterion-referenced tests.
Domain (of a test)	The content and skills upon which a test is based.
Item (of a test)	A test question.
Norm-Referenced Test (NRT)	The primary distinguishing characteristic of “norm-referenced” tests is that student performance is measured against that of other students, rather than against some fixed standard of performance. Norm-referenced test results are usually expressed in terms of population percentiles along a bell-shaped distribution of tested students. A typical norm-referenced test result is that a 4 th grade student’s achievement in mathematics is at the 55 th percentile, meaning that her or his performance is better than that of 55% of a nationally representative sample of 4 th grade students who have taken the test under the same conditions, but worse than that of the other 45% of tested students in the sample. Most of the widely administered, commercially published K-12 achievement tests, such as the Iowa Test of Basic Skills, TerraNova, or the Stanford series, are norm-referenced tests, at least in their standard forms.
Standardized Test	Any test for which the test items, as well as the conditions under which the test is administered, are constant. Thus, both CRTs and NRTs may be standardized tests.

⁵⁷ Audrey L. Amrein, and David C. Berliner, *High Stakes Testing, Uncertainty, and Student Learning*, published on the Internet at the Education Policy Analysis Archives, vol. 10, no. 18, at <http://epaa.asu.edu/epaa/v10n18/>.

Author Contact Information

Rebecca R. Skinner
Specialist in Education Policy
rskinner@crs.loc.gov, 7-6600

Erin D. Caffrey
Analyst in Education Policy
ecaffrey@crs.loc.gov, 7-9447

Acknowledgments

Wayne C. Riddle, former Specialist in Education Policy, was the original author of this report.