

CRS Report for Congress

Received through the CRS Web

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

Updated April 6, 2004

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

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

Summary

The No Child Left Behind Act of 2001 (NCLBA) contains several new requirements related to pupil assessments for states and local educational agencies (LEAs) participating in Elementary and Secondary Education Act (ESEA) Title I-A (Education for the Disadvantaged). These expand upon less extensive requirements that were adopted under the Improving America's Schools Act (IASA) of 1994.

Under the NCLBA, in addition to the IASA requirements for standards and assessments in reading and mathematics at three grade levels, all states participating in Title I-A will be required to implement standards-based assessments for pupils in *each* of grades 3-8 in reading and mathematics by the 2005-2006 school year. States will also have to develop and implement assessments at three grade levels in *science* by the 2007-2008 school year. Pupils 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) pupils. Assessments must be of "adequate technical quality," and grants are authorized for development of enhanced assessments. Grants to states for assessment development are authorized, and \$390 million has been appropriated for FY2004.

In addition, the NCLBA requires all states receiving grants under ESEA 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 pupils primarily in grades 4, 8, and 12. Beginning in 1990, NAEP has conducted a limited number of state-level assessments wherein the sample of pupils tested in each participating state is increased in order to provide reliable estimates of achievement scores for pupils 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 NCLBA to maximize consistency with the NCLBA requirements and prohibit the use of NAEP assessments by agents of the federal government to influence state or LEA instructional programs or assessments.

Issues regarding the expanded ESEA Title I-A pupil assessment requirements, which may be addressed by the 108th Congress, include: What types of assessments will meet the ESEA Title I-A requirements? How strict will the Department of Education be in reviewing and approving state assessment systems, and will states meet the expanded assessment requirements on schedule? What will be the cost of developing and implementing the assessments, and to what extent will federal grants be available to pay for them? What might be the impact of the requirement for annual assessment of the English language proficiency of LEP pupils? What might be the impact on NAEP of requiring state participation, as well as the impact of NAEP on state standards and assessments? And what are the likely major benefits and costs of the expanded ESEA Title I-A pupil assessment requirements?

Contents

Introduction	1
Current State Testing Policies and Practices	1
Testing Program Costs	3
Federal Policies or Activities Regarding Pupil Assessments Under the	
No Child Left Behind Act	4
ESEA Title I-A Requirements for Standards and Assessments	4
Limits on ED Influence Over State Standards and Assessments	9
State Assessment Grants	9
National Assessment of Educational Progress	10
State NAEP	12
NAEP Provisions in the No Child Left Behind Act	12
Status of Implementation of the Assessment Requirements	14
ED Review of Evidence Regarding Assessments to Meet the	
“1994 Requirements” Under Title I-A	14
Common Problem Areas Found in Reviews of State Assessment	
Systems with Respect to the “1994 Requirements”	16
Interpretation by ED of the Expanded Standard and Assessment	
Requirements of the No Child Left Behind Act	17
Title I-A Standard and Assessment Requirements	17
Steps Toward Implementation of the NAEP Requirements	21
Issues Regarding the Expanded ESEA Title I-A Pupil Assessment	
Requirements	22
What Types of Assessments Would Meet the Expanded	
Assessment Requirements?	22
How Strict Will Be ED’s Review of State Assessment Systems, and Will	
States Meet Requirements on Schedule?	24
What Will Be the Cost of Developing and Implementing the	
Required Assessments, and to What Extent Will Federal Grants	
Be Available to Pay for Them?	26
What Might Be the Impact of the Requirement for Annual Assessment	
of English Language Proficiency of LEP Pupils?	28
What Might Be the Impact of Requiring State Participation in NAEP? ...	29
Possible Influence on State Standards and Assessments Arising	
from (Marginally) Increased Stakes	29
Voluntary Participation by LEAs, Schools, and Pupils	30
Can NAEP Results Be Used to “Confirm” State Test Score Trends? .	31
What Are the Likely Benefits and Costs of the Expanded Title I-A	
Assessment Requirements?	32

Glossary of Selected Terms Used in This Report

Criterion-referenced test (CRT): “Criterion-referenced” tests measure the extent to which pupils have mastered specified content (content standard) to a predetermined degree (achievement standard). A typical criterion-referenced test result is that a 4th grade pupil’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 pupil performance is measured against that of other pupils, 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 pupils. A typical norm-referenced test result is that a 4th grade pupil’s achievement in mathematics is at the 55th percentile, meaning that her or his performance is better than that of 55% of a nationally representative sample of 4th grade pupils who have taken the test under the same conditions, but worse than that of the other 45% of tested pupils 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.

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

Introduction

The No Child Left Behind Act of 2001 (NCLBA, P.L. 107-110), signed on January 8, 2002, contains a number of new requirements related to pupil 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 new 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).

This report provides background information on state pupil assessment programs and policies, a description of the ESEA Title I-A assessment requirements as expanded by the NCLBA, a review of the implementation status of these requirements, and an analysis of related issues which may be addressed by the 108th Congress. This report will be updated occasionally, when major developments occur in the process of implementing the expanded ESEA Title I-A pupil assessment requirements.

Current State Testing Policies and Practices

The academic achievement of pupils in public elementary and secondary schools is assessed using many types of tests. Pupils may take tests developed by individual teachers or schools, commercially published tests selected by their LEA, or assessments selected and/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 pupils in selected grades who attend a state's public K-12 schools — because such tests are the primary focus of federal policies regarding pupil assessment.

According to recent surveys,¹ every state except one (Iowa) now requires its LEAs to administer specified assessments to all pupils 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 pupils in grade 4) to tests in multiple subjects and most K-12 grades (e.g., Alabama requires pupils in each of grades 3-11 to take state-selected tests in English, mathematics, science, and history). Few state-mandated tests are administered to pupils below grade 3, because of a variety of concerns about administering standardized tests to very young pupils, or in grade 12, in part because most assessment activity for these pupils 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 pupils in 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 pupils 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 pupils’ achievement level in comparison to a nationally representative sample of pupils — purchased by states from commercial test publishers. These two types of tests vary *primarily* regarding how test results are

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 pupils attending public K-12 schools in Iowa, on the basis of voluntary decisions by each LEA.

³ Three of the largest such commercial test publishers are: (1) CTB/McGraw-Hill, at [<http://www.ctb.com/>]; Riverside (Houghton Mifflin) Publishing, at [<http://www.riverpub.com/products/groupindex.html>]; and Harcourt Assessment, at [<https://marketplace.psychcorp.com/PsychCorp/International.aspx>].

⁴ An 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.

analyzed, but also typically differ to some degree with respect to such characteristics as the range of questions included.⁵

As of spring 2000, 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) have developed state tests which 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 pupil 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

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 expenditures 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 pupil 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 pupil), New Mexico (\$3.21 per pupil), and Utah (\$3.16 per pupil). In contrast, per pupil costs were found to be highest for several states which

⁵ For example, in order to clarify distinctions between high- and low-achieving pupils, a norm-referenced test will typically include some very difficult questions that only a few pupils can answer, and some very easy questions that almost all pupils can answer correctly. Test content and questions are selected largely on the basis of how efficiently they rank pupils. 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 pupils.

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 by teachers and other staff preparing pupils 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 on the Internet at [<http://www.stateline.org>].

rely primarily or solely on state-specific CRTs, such as Alaska, Wyoming (\$78.34 per pupil), Virginia (\$68.90 per pupil) and Massachusetts (\$68.02 per pupil).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 Pupil Assessments Under the No Child Left Behind Act

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

ESEA Title I-A Requirements for Standards and Assessments

The provisions of ESEA Title I-A, as amended by the NCLBA, 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 NCLBA, these standards and assessment provisions are linked to receipt of financial assistance under ESEA Title I-A — i.e., 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 \$11.7 billion for FY2003, it is generally considered unlikely that many states would choose not 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 pupils, 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 pupil

⁸ See Education Commission of the States, *Estimated Per-Student Spending on Statewide Testing Programs*, Oct. 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, Feb. 1996.

testing that was of little instructional or diagnostic value, and was not linked to the curriculum to which pupils 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 NCLBA 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 NCLBA 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 pupils statewide. One major exception to this general policy is that if no agency or entity in a state has authority to establish statewide standards or assessments (as is the case for Iowa and possibly Nebraska), then the state may adopt either: (a) statewide standards and assessments applicable only to Title I-A pupils 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 pupils 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:

1. Standards and assessments must be developed or adopted at least in the subjects of mathematics and reading/language arts by the 2000-2001 school year.¹¹ Standards are to be adopted in science by the 2005-2006 school year, and assessments in science by 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 pupils in the state (with the two possible exceptions discussed above).
3. The content standards are to specify what pupils are expected to know and be able to do, and are to be “coherent and rigorous.”

¹⁰ See CRS Report RS21094, *Adequate Yearly Progress Under ESEA Title I: Estimates for Three States Based on Specifications in H.R. 1 Conference Agreement*, by David P. Smole and Wayne C. Riddle.

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

4. Achievement standards must establish *at least* three performance levels for all pupils — 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 must be administered to pupils in *each* of grades 3-8 beginning in the 2005-2006¹² school year, if certain minimum levels of annual federal funding are provided for state assessment grants.¹³
7. All pupils in the relevant grades who have attended schools in the LEA for at least 1 year must participate in the assessments.¹⁴
8. LEP pupils are to be assessed in a valid and reliable manner and provided with “reasonable” accommodations. To the extent practicable, LEP pupils 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, pupils who have attended schools in the United States (excluding Puerto Rico) for 3 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.
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

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 1 or 2 of grades 3-8.

¹⁴ Separately, the provisions regarding AYP provide that at least 95% of the pupils in each demographic group within each school must be included in the assessments in order for the school to meet AYP requirements. Pupils 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 pupils 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.

- 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 pupils 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 pupils — including pupils’ oral, reading, and writing skills — beginning in the 2002-2003 school year.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 table.

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, 21 states fully met these requirements.)

School Year 2001-2002

- No new waivers of the deadlines for meeting the “1994 requirements” could be granted after April 8, 2002.

School Year 2002-2003

- States were required to begin to annually assess the English language proficiency of LEP pupils (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 new assessment and related requirements under the NCLBA.

May 1, 2003

- States were required to include in their ESEA consolidated application academic content standards in reading/language arts and mathematics for each of grades 3-8, as well as a detailed timeline for meeting the standard and assessment requirements listed below.

School Year 2005-2006

- Standards-based assessments in reading and mathematics must be administered to pupils in each of grades 3-8 (possible waivers if minimum amounts are not appropriated for state assessment grants or for “exceptional or uncontrollable circumstances”).
- States must adopt content standards at three grade levels in science.

School Year 2007-2008

- States must begin to administer assessments at three grade levels in science.

Limits on ED Influence Over State Standards and Assessments.

Several statutory constraints have been placed on the authority of the Secretary of Education in enforcing 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 1905¹⁷). 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 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 NCLBA (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 pupils. 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 FY2003-2007.

The state assessment requirements that were newly adopted under the NCLBA 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 1 year for each year that the following minimum amounts are *not* appropriated: FY2002 — \$370 million, FY2003 — \$380 million, FY2004 — \$390 million, and each of FY2005-2007 — \$400 million. For example, if an amount less than \$400 million were 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 move from 2005-2006 to 2006-2007. For FY2002 and FY2003, the minimum amounts were appropriated for these grants (\$370 and \$380 million, respectively). For FY2004, the minimum (\$390 million) will be appropriated, assuming that the Department is able to follow through on a proposed transfer of \$710,000 from a different account (i.e., the amount initially appropriated was \$710,000 below the threshold).

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

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

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; for FY2002, \$17 million was available for this purpose. In February 2003, grants to nine states were announced. While each of these grants was made to a single state, each grantee has a number of "collaborators," including other states as well as, in some cases, LEAs, universities, or educational research organizations. The grants range from \$1.4 to \$2.3 million. The amount available for assessment enhancement grants was \$4,484,000 under the FY2003 appropriation; no funds will be available for assessment enhancement grants under the FY2004 appropriation.

Finally, the NCLBA 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 pupil achievement, instructional practices, dropout and graduation rates, and school staff turnover rates; effects on different groups of pupils, such as LEP pupils, pupils from low-income families, or pupils with disabilities; and relationships between accountability systems and exclusion of pupils from state assessments.

National Assessment of Educational Progress¹⁸

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 pupils 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

¹⁸ For additional information on NAEP, see CRS Report 98-348, *National Assessment of Educational Progress: Background and Reauthorization Issues*, by Wayne Riddle.

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 pupil scores in relation to *performance levels* — standards for pupil 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 pupils 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.¹⁹ The tests in each subject area have not changed in content or structure since they were originally developed in 1969, 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 pupils are actually taught with the passage of decades of time.²⁰ 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 pupils, and the tests are designed so that no pupil takes an entire NAEP test. The use of sampling is intended to minimize both the costs of NAEP and test burdens on pupils. It also makes it possible to include a broad range of items in each test. Since no individual pupil takes an entire NAEP test, it is impossible for NAEP to report individual pupil scores.²¹ It is intended that NAEP tests be administered to a representative sample of all pupils in public and private schools, although there has been ongoing debate over whether LEP pupils 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 pupils 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

¹⁹ 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.

²⁰ An NAGB policy adopted in May 2002 addresses this concern with respect to the science assessment, and some changes will be made to the content of the science assessment before its next administration.

²¹ The Voluntary National Test proposal of the Clinton Administration was to develop individual versions of the NAEP 4th grade reading and 8th grade math tests (see CRS Report 97-774, *National Tests: Administration Initiative*, by Wayne Riddle). Activity related to this proposal has been terminated.

not intended to specifically reflect any particular set of standards. In addition, pupils 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 pupil race, ethnicity, socioeconomic status, disability, LEP status, and gender.

State NAEP. While NAEP, as currently structured, cannot provide assessment results for individual pupils, the levels at which scores could be provided — the Nation overall, states, LEAs, or schools — depend on the size and specificity of the sample group of pupils 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. Only the main NAEP, not the long-term trend assessment, is administered at the state level. Under state NAEP, the sample of pupils tested in a state is increased in order to provide reliable estimates of achievement scores for pupils in each participating state.

Until enactment of the NCLBA (see below), participation in NAEP was voluntary for states,²² 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 state-level NAEP assessments 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, FY2002 appropriations legislation provided for a Trial Urban Assessment of achievement in reading and writing — experimental administration of NAEP to expanded pupil samples in a limited number of large urban LEAs. The assessment was administered to extended samples of pupils 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.²³

NAEP Provisions in the No Child Left Behind Act. The NCLBA 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 pupil samples in the states will now be paid by the federal government. An unstated, but implicit, purpose of this new requirement is to “confirm” trends in pupil

²² 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.

²³ For a description of the Trial Urban Assessment, and available results, see [<http://nces.ed.gov/nationsreportcard/reading/results2002/districtresults.asp>].

achievement, as measured by state-selected assessments.²⁴ The results from the initial state NAEP assessment in 4th and 8th grade reading and mathematics involving all 50 states were released in November 2003.

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 NCLBA. While 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 pupils 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 which may be selected for NAEP test administration. The NCLBA itself explicitly provides that participation in NAEP tests is voluntary for all *pupils* and *schools*, but it contains conflicting provisions regarding voluntary participation by *LEAs*. The NAEP authorization statute (recently redesignated 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*, Dec. 2, 2002) require both LEAs which receive Title I-A grants, and schools within such LEAs, to participate in NAEP if selected to be among the sample tested (34 CFR 200.11(b)).

The NCLBA authorizes funds specifically for state NAEP tests for fiscal years 2002-2007 — \$72 million for FY2002 and “such sums as may be necessary” for the succeeding years. The NCLBA did not extend the authorization for NAEP overall. However, Title III of P.L. 107-279, the National Assessment of Educational Progress Authorization Act, has recently 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 FY2004-08. P.L. 107-279 also redesignates NAEP’s

²⁴ 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. 1 a 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>].

statutory language as Title III of the Education Sciences Reform Act of 2002 (ESRA), but does not otherwise directly or substantially amend the provisions.²⁵

For FY2002, the total amount appropriated for all NAEP and NAGB activities was \$111.553 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 Assessment described above. The total amount appropriated for NAEP and NAGB was \$94.767 million for FY2003 and \$94.763 million for FY2004.

Status of Implementation of the Assessment Requirements

The scheduled deadlines for implementation of major assessment requirements under ESEA Title I-A are outlined earlier in this report. Thus far, almost all implementation activity has taken place with respect to requirements adopted initially in the 1994 IASA and continued under the NCLBA. The process of implementing the 1994 requirements is still incomplete.

ED Review of Evidence Regarding Assessments to Meet the “1994 Requirements” Under Title I-A

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 — i.e., they are not reviewing the assessments themselves.²⁶ 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 pupils, or alternate assessments or other accommodations for pupils with disabilities; provisions for reporting scores by disaggregated pupil groups; or data on the extent of actual participation in assessments of LEP pupils or pupils with disabilities.

²⁵ See CRS Report RL31353, *Educational Research, Statistics, and Evaluation: Legislation in the 107th Congress*, by Paul M. Irwin.

²⁶ 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 NCLBA, it remains applicable, at least for the present, mainly because most applicable underlying requirements are essentially unchanged.

All of the reviews conducted thus far focus primarily on the “1994 requirements” of ESEA Title I-A, because the deadlines for meeting the expanded requirements under the NCLBA have not yet occurred. However, aside from the specifics regarding grade levels and subject areas, these reviews are applicable to ESEA Title I-A as amended by the NCLBA, since the “1994 requirements” also must be met under the revised statute. In addition, these “1994 reviews” are the best currently available indication of how ED staff will implement the expanded assessment requirements as those deadlines approach.

**Status of Review of State Assessment Systems with Respect to the “1994 Requirements” Under ESEA Title I-A
(as of the publication date of this report)**

Assessments Fully Approved (21): Colorado, Connecticut, Delaware, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Missouri, New Hampshire, New York, North Carolina, Oregon, Pennsylvania, Rhode Island, Texas, Vermont, Virginia, Wyoming

Timeline Waiver Granted (26): Alaska, Arizona, Arkansas, California, Florida, Georgia, Hawaii, Illinois, Iowa, Michigan, Minnesota, Mississippi, Nebraska, Nevada, New Jersey, New Mexico, New York, North Dakota, Ohio, Oklahoma, Puerto Rico, South Carolina, South Dakota, Tennessee, Utah, Washington, Wisconsin

Compliance Agreement Negotiated (5): Alabama, District of Columbia, Idaho, Montana, West Virginia

As of the publication date of this report, 21 states fully met all of the pre-NCLBA requirements regarding standards and assessments (see box above). Twenty-six of the states that had not met these requirements have been granted waivers until fully meeting them within a specified period of time. These waivers have been granted under authority in Title XIV-D of the 1994 version of the ESEA or Title IX-D of the current ESEA.²⁷ For the remaining five states, which were found to have made the least progress toward meeting the “1994 requirements,” ED has negotiated compliance agreements.²⁸ According to ED, “A compliance agreement allows a state to continue to receive federal education funds while it comes into

²⁷ The pre-NCLBA version also specifically authorized the Secretary of Education to waive the deadline for assessment implementation for an additional year if necessary to correct problems that were identified in field testing; the current ESEA authorizes specific waivers for up to 1 year for full implementation of the new requirements for annual assessments in grades 3-8, and assessment of the English language proficiency of LEP pupils, in cases of “exceptional and unforeseen circumstances,” and authorizes states to delay implementation of the grade 3-8 requirement if minimum amounts are not appropriated for state assessment grants.

²⁸ The compliance agreements for 4 of the 5 states have been published in the *Federal Register* on Feb. 20 (Alabama and Idaho), Feb. 21 (West Virginia), and Mar. 24 (Montana), 2003.

compliance under specific terms, conditions, and a timeline spelled out in the written agreement, with three years as the maximum time allowed.”²⁹

Both before and after the NCLBA, 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 NCLBA, the ESEA provides that the Secretary *shall* withhold 25% of funds otherwise available for state administration and program improvement activities from states which fail to meet the 1994 requirements, and *may* withhold additional state administration funds for failure to meet new assessment requirements adopted under the NCLBA. In addition, states which persistently and thoroughly fail to meet the standard 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. In spite of these authorized or potential sanctions, no state has yet experienced any reduction in funds or other negative consequences (at least with respect to the Title I-A program specifically) due to failure to meet any deadline related to the Title I-A standard and assessment requirements.

Common Problem Areas Found in Reviews of State Assessment Systems with Respect to the “1994 Requirements”. The peer reviews of state assessment systems conducted thus far have identified a number of common problem areas, as indicated in “decision letters” from ED officials to the states.³⁰ These are: (a) lack of adequate inclusion, accommodation, and incorporation of alternate assessments for LEP and disabled pupils; (b) insufficient documentation of the technical quality of assessments (i.e., their reliability, alignment, validity, etc.), especially the degree of alignment of assessments with content and pupil performance/achievement standards; and (c) inadequate timelines for completion and implementation of the assessments.

The first of these three problem areas has received the greatest attention. The revised ESEA, ED’s “Summary Guidance on the Inclusion Requirement for Title I Final Assessments,” as well as other letters and policy guidance documents, indicate that the only students who should be excluded from assessments are those who have attended public schools in a LEA for less than one year. Otherwise, *all* pupils should be included in both the assessments and associated accountability systems.³¹ Where appropriate, accommodations (for example, extended time to complete an assessment) or alternate assessments³² should be provided for pupils with disabilities.

²⁹ U.S. Department of Education press release dated Apr. 8, 2002.

³⁰ These are available at [http://www.ed.gov/offices/OESE/saa/state_chart.html].

³¹ Pupils who have attended schools in a LEA for one year or more, but who have attended a particular school for less than one year, may be excluded from accountability determinations for the school (but not for the LEA overall).

³² Section 612 (a)(17) of the Individuals with Disabilities Education Act (IDEA) requires
(continued...)

LEP pupils should be assessed in the language most likely to yield valid results, except that those who have attended schools in the United States (other than Puerto Rico) for 3 or more years must generally be assessed in English, and they should be provided with other accommodations (e.g., extended time or use of bilingual word lists or dictionaries) where appropriate, as determined on an individual basis. With respect to inclusion of LEP pupils and those with disabilities, ED is reviewing “evidence” not only of state policies but also practices (i.e., actual rates of participation by LEP and disabled pupils). Many of the states whose assessments have not yet been approved have been informed that they need to make changes regarding assessment of or reporting of scores for LEP and/or disabled pupils.

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

Title I-A Standard and Assessment Requirements. On July 5, 2002, ED published regulations on the Title I-A assessment requirements newly adopted under the NCLBA.³³ 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 NCLBA 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

³² (...continued)

states to develop guidelines for the administration of alternate assessments for pupils with disabilities who cannot participate in state- and LEA-wide assessment programs.

³³ *Federal Register*, July 5, 2002, pp. 45038-45047. As is discussed below, proposed amendments to these regulations were published in the *Federal Register* on Mar. 20, 2003.

additional mechanism, beyond publication for comments in the *Federal Register*, of obtaining input on proposed regulations from concerned organizations.³⁴

Significant features of the Department’s final regulations — developed through the negotiated rulemaking process³⁵ 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; (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 pupil 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,³⁶ 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 pupils 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

³⁴ ED’s implementation of the negotiated rulemaking requirement was challenged in federal court. Four organizations (The Center on Law and Education, National Coalition for the Homeless, 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.

³⁵ 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 included: (a) it was further clarified that the assessment requirements apply only to public schools and their pupils, not to private (or home) schools; (b) for purposes of disaggregated score reporting, “pupils with disabilities” would be only those identified under the IDEA (this would exclude pupils 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.

³⁶ 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.

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 pupils, not to private (or home) schools, although the achievement of private school pupils 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, “pupils with disabilities” are only those identified under the IDEA,³⁷ although all pupils 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.

Recent ED Policy Developments Regarding Participation Rates Plus Treatment of Limited English Proficient Pupils and Certain Pupils With Disabilities in Assessments and AYP Determinations. ED officials have recently published regulations and other policy guidance on participation rates plus the treatment of limited English proficient pupils and certain pupils 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 pupils (all pupils as well as pupils in each designated demographic group) participate in assessments (in order for the school or LEA to make AYP) on the basis of *average* participation rates for the last two or three years, rather than having to post a 95% or higher participation rate each year. In other words, if a particular demographic group of pupils in a school has a 93% test participation rate in the most recent year, but had a 97% rate the preceding year, the 95% participation rate requirement would be met. In addition, the new guidance would allow schools to exclude pupils who fail to participate in assessments due to a “significant medical emergency” from the participation rate calculations. The new guidance further emphasizes the authority for states to allow pupils who miss a primary assessment date to take make-up tests, and to determine the minimum size for demographic groups of pupils to be considered in making AYP determinations (including those related to participation rates). According to ED, in some states, as many as 20% of the schools failing to make AYP did so on the basis of assessment participation rates alone. It is not known how many of these schools would meet the new, somewhat more relaxed standard.

On February 19, 2004, ED officials announced two new policies with respect to LEP pupils.³⁸ First, with respect to assessments, LEP pupils in their *first year* of attending schools in the United States must participate in English language proficiency and mathematics tests. However, the participation of such pupils in

³⁷ This would exclude pupils identified only under Section 504 of the Rehabilitation Act.

³⁸ See [<http://www.ed.gov/nclb/accountability/schools/factsheet-english.html>].

reading tests, as well as the inclusion of these pupils' test scores in AYP calculations, is to be optional (i.e., schools and LEAs need not consider the scores of first year LEP pupils in determining whether schools or LEAs meet AYP standards). Second, in AYP determinations, schools and LEAs may continue to include pupils in the LEP demographic category for up to two years after they have attained proficiency in English. Both these options, if exercised, should increase average test scores for LEP pupil groups, and reduce the extent to which schools or LEAs fail to meet AYP on the basis of such pupils.

Additional regulations, addressing the application of the Title I-A standard and assessment requirements to certain pupils with disabilities, were published in the *Federal Register* on December 9, 2003 (pp. 68698-68708). These regulations amend the regulations published on July 5, 2002, discussed above, as well as final regulations on other aspects of Title I-A accountability requirements, published on December 2, 2002.³⁹ The purpose of these amendments is to clarify the application of standard, assessment, and accountability provisions to pupils “with the most significant cognitive disabilities.” Under the regulations, states and LEAs may adopt alternative assessments *based on alternative achievement standards* — aligned with the state’s academic content standards and reflecting “professional judgment of the highest achievement standards possible” — for a limited percentage of pupils with disabilities.⁴⁰ When making AYP determinations, in general no more than 1.0% of all pupils (approximately 9% of all pupils with disabilities) counted as having achievement levels of proficient or above may consist of pupils taking such alternative assessments based on alternative achievement standards at the state and LEA level; there is no such limitation 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.

In addition, ED published supplementary “non-regulatory draft guidance” on these standard and assessment requirements, as well as those related to NAEP participation, on March 10, 2003.⁴¹ This document is intended to provide guidance which is consistent with that in the regulations discussed above, but is more detailed. This guidance specifically provides that states must 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*.

³⁹ These are discussed in CRS Report RL31487, *Education for the Disadvantaged: Overview of ESEA Title I-A Amendments Under the No Child Left Behind Act*, by Wayne Riddle.

⁴⁰ This limitation does *not* apply to the administration of alternative assessments *based on the same standards applicable to all students*, for other pupils with (non-cognitive or less severe cognitive) disabilities.

⁴¹ [<http://www.ed.gov/topics/topicsTier2.jsp?&top=Policy&subtop=Policy+guidance&subtop2=Elementary+%26+secondary+education&type=T>].

Steps Toward Implementation of the NAEP Requirements. In the period since enactment of the NCLBA, 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. 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.

At the same time, NAGB has established, and NCES has begun to implement, several changes to NAEP policies and practices that are supportive of, or were adopted primarily in response to, the expanded role for NAEP under the NCLBA.⁴² In recognition of the increased emphasis on measurement of performance gaps among different demographic groups of pupils in the NCLBA, more questions are being added at the upper and lower ends of the difficulty range, so that achievement gaps among pupil 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 pupils in the various demographic groups referenced in the NCLBA.

At the same time, a number of administrative adjustments are being implemented that are intended to reduce required pupil sample sizes in the aggregate (e.g., the main NAEP state and national pupil samples will be combined for the first time), although samples of pupils 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 has not previously participated in state NAEP tests. A basic problem is that selected NAEP mathematics tests have been translated into and administered (to LEP pupils) in Spanish, but there are no Spanish versions of the NAEP reading tests. Aside from the task of translation, questions arise about the comparability of tests administered in different languages, especially in subjects such as reading. Currently, NAGB plans to administer NAEP mathematics tests to 4th and 8th grade pupils in Puerto Rico beginning in 2003, but policy will not be established regarding development and administration of NAEP reading tests there until a future date.

Finally, state NAEP tests will now be administered by contractors, rather than (as in the past) local teachers; there will be a full-time NAEP coordinator in every state, and a State Service Center will be established to support these coordinators; and NAGB has established procedures for limited public access to NAEP test items,

⁴² 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].

and for submission, review, and resolution of complaints about NAEP tests by parents and other members of the public.

Issues Regarding the Expanded ESEA Title I-A Pupil Assessment Requirements

What Types of Assessments Would Meet the Expanded Assessment Requirements?

As described above, the NCLBA includes explicit reference to a number of criteria which 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 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 which are the same statewide (except in states which 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 pupils 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 pupil 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 pupils, not for the purpose of determining whether pupils 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.⁴³

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

⁴³ 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.

results to determine whether schools or LEAs meet state standards of adequate yearly progress (AYP). The best way to assure that assessments of the extent to which pupils 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 recent 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.”⁴⁴ 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 which may lack authority to establish statewide assessments, no states have been allowed to meet the 1994 standard and assessment requirements through 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 pupils 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 CFR 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 CFR 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

⁴⁴ National Research Council, *Uncommon Measures: Equivalence and Linkage Among Educational Tests*, 1998, p. 5-4.

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 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 Will Be ED's Review of State Assessment Systems, and Will States Meet Requirements on Schedule?

While distinct, this issue is related to the topic discussed immediately above, since the likelihood that states will meet statutory deadlines will in many cases be significantly influenced by the degree of flexibility to include a variety of tests (NRTs and CRTs, locally varying and statewide tests) in state assessment systems. The greater the allowed degree of flexibility, the easier it will be for states to modify existing assessment systems to meet the expanded Title I-A requirements. However, the rigor of ED's review of state assessment systems is a somewhat more independent factor — a high or low degree of strictness may accompany either a broad or narrow range of flexibility in the requirements being enforced.

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 "1994 requirements" initiated by the IASA. 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 pupils 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 pupils,⁴⁵ and requirements under the IDEA for alternate assessments where necessary for pupils 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.

⁴⁵ 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>].

While 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 pupils. If, for example, significant numbers of LEP pupils 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 pupil 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 pupils with lowest achievement in a participating school.⁴⁶

While 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 assessment items or instruments. Nevertheless, the degree of federal influence over at least the broad parameters of state pupil assessment systems — such as grades and subject areas tested, inclusion of special needs pupil groups, disaggregated reporting of results — will increase as the NCLBA provisions are fully implemented.

A recent paper by a former ED official who was largely responsible for the initial stages of review of state compliance with the “1994 requirements” under ESEA Title I-A encourages continuation of a “vigorous” approach by ED to enforcement of these requirements — “being prepared to use a full range of enforcement strategies — from jawboning to compliance agreements to withholding administrative or program funds if necessary” even though ED has not had “a strong record of compliance monitoring in ESEA programs.”⁴⁷ In contrast, others have urged ED staff to emphasize flexibility in reviewing state compliance with assessment and related requirements under the NCLBA — “We would advise those involved in the rulemaking and guidance process to proceed cautiously, for the very vagueness of the law ... is actually an asset, as it leaves each state room to experiment within its own strengths and limitations.”⁴⁸ It is not yet clear whether, or how, the

⁴⁶ There are two basic types of Title I-A programs. *Schoolwide programs* are authorized when 40% or more of the pupils in a school are from low-income families. In these programs, Title I-A funds may be used to improve the performance of all pupils in a school, and there is no requirement to focus services on only the most disadvantaged pupils. 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 pupils in the school.

⁴⁷ Michael Cohen, “Implementing Title I Standards, Assessments And Accountability: Lessons From The Past, Challenges For The Future,” *No Child Left Behind: What Will It Take?*, Thomas B. Fordham Foundation, Feb. 2002. pp. 84, 87.

⁴⁸ Lisa Graham Keegan, et al., “Adequate Yearly Progress: Results, not Process,” *No Child* (continued...)

nature and rigor of this review process may change with more complete implementation of the NCLBA.

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 recent General Accounting Office report 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."⁴⁹

Given the experience thus far with the "1994 requirements," as described above, it seems likely that a significant number of states may fail to meet the future assessment deadlines in the NCLBA. ED is prohibited from granting additional waivers, or negotiating additional compliance agreements, with respect to the 1994 requirements, and the NCLBA requires ED to apply sanctions to states which fail to meet their modified deadlines for those requirements. However, there are no similar limits on the possible granting of waivers (under the general ESEA waiver authority in Title IX, Part D), and no *required* sanctions (although the sanction of reducing state administration grants is authorized), for states which fail to meet the *new* assessment requirement deadlines in the NCLBA.

What Will Be 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 assessments in at least four more grades than required previously, and to include standards and assessments at three grade levels in science, will require 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. Existing estimates of actual state and/or local expenditures for assessment programs are incomplete, and they refer only to direct, state-level costs for current testing programs, not to the expanded assessments required under the NCLBA.

The NCLBA conference report directs the General Accounting 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; however, no information is yet available from that study. At least two organizations have attempted during 2001-2002 to estimate costs for states of meeting assessment requirements similar to those of the NCLBA. In 2001, the National Association of State Boards of Education (NASBE) estimated that the new

⁴⁸ (...continued)

Left Behind: What Will It Take?, Thomas B. Fordham Foundation, Feb. 2002, p. 22.

⁴⁹ *Title I, Education Needs to Monitor States' Scoring of Assessments*, GAO-02-393, Apr. 2002, p. 13.

grade 3-8 assessments (only) would cost states between \$2.7 and \$7.0 billion in the aggregate over a seven-year period.⁵⁰ 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, recently estimated that the annual cost of meeting all of the new assessment requirements in the NCLBA would be approximately \$312 million for each of 2002-2003 and 2003-2004, \$388 million for each of 2004-2005 and 2005-2006, and \$328 million for each of 2006-2007 and 2007-2008.⁵¹

Each of these estimates is based on a number of assumptions which may or may not prove to be valid in practice, and the NASBE study was based on portions of the original Bush Administration proposal, not the final version of the NCLBA. For example, the NASBE study considers only the requirement for annual standards-based assessments in mathematics and reading in each of grades 3-8, not the science assessment requirement; is based on the original proposal that grade 3-8 assessments must be implemented in 2004-2005, rather than 2005-2006; does not seem to incorporate an assumption that development costs will decline at any point of the life cycle of assessments; and incorporates an assumption that test development costs will rise continuously with the number of pupils in a state, with no adjustment for economies of scale. The Accountability Works study assumes that all states already meet the assessment requirements under the 1994 IASA (whereas only 21 do so currently); incorporates assumed levels of annual test development and administration costs which appear to be based on data from only two states; excludes costs for SEA staff and overhead; and incorporates an assumed amount for annual test administration costs (\$10 per pupil) which would not be consistent with use of essay or other extended response (and relatively expensive to score) questions. Further, *neither* of these studies accounts for variation among the states in the extent to which they already administer standards-based assessments in reading and mathematics in each of grades 3-8 plus science assessments at three grade levels; and the basis for some of the assumptions in each of the studies is unclear.

The NCLBA authorizes \$400 million for FY2002, and “such sums as may be necessary” for a number of subsequent years, for state assessment development and administration grants. The administration, although not the development, of the grade 3-8 assessments may be delayed by 1 year for each year that the minimum amounts (e.g., \$390 million for FY2004) are *not* appropriated (the minimum amount has been appropriated for each of FY2002-2004). The available information on *current* levels of direct, state-level expenditures for testing programs indicate that the “trigger” appropriation levels for state assessment grants are, in the aggregate, somewhat lower than these estimates.⁵² The “trigger” amounts are also somewhat above the Accountability Works estimate of *future* costs, approximately the same as the minimum NASBE estimate, and well below the upper NASBE estimate.

⁵⁰ See [<http://www.nasbe.org/Archives/cost.html>].

⁵¹ See [http://www.accountabilityworks.org/publications/no_child_left_behind_test_costs.pdf].

⁵² The \$370 million “trigger” amount (and actual appropriation) for FY2002 is 88% of the estimated aggregate expenditure level for FY2001 (discussed on p. 3 of this report) of \$422.8 million.

It is clear that the costs of meeting the expanded assessment requirements will vary 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 are already administered to all pupils in grades 3-8, or tests in science are administered to pupils in selected grade ranges, and whether the existing tests meet the Title I-A technical requirements of alignment with state standards, inclusion of all pupil groups, etc. As of the 2001-2002 school year, 15 states administered tests in reading and mathematics in each of grades 3-8, and 24 states administered tests in at least three grade levels in science, but it is unclear how many of these tests are adequately aligned with state content and achievement standards.⁵³

With respect to the distribution among the states of funds for test development and administration, the NCLBA provides for allocation of a substantial share (43% for FY2002) of these funds in equal amounts to each state, with the remainder allocated in proportion to children and youth aged 5-17 years. The allocation formula does not recognize the substantial variation in the extent to which states currently administer the required assessments, and therefore face varying increases in 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.

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

As noted earlier, the NCLBA requires states to provide that their LEAs will annually assess the English language proficiency of their LEP pupils, beginning in the 2002-2003 school year. This is separate from the requirements regarding treatment of LEP pupils in states' general assessment systems — i.e., the requirement that LEP pupils 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 pupils who have attended schools in the United States (excluding Puerto Rico) for 3 or more consecutive school years to be assessed in English.

In contrast to such requirements regarding treatment of LEP pupils 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 pupils' 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.

⁵³ Education Commission of the States, *No State Left Behind: The Challenges and Opportunities of ESEA 2001*. Available at [<http://www.ecs.org>].

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 pupils for purposes of placement in instructional programs, determination of needed accommodations in general assessment programs, evaluation of programs targeted on LEP pupils, and movement of pupils 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) pupils.⁵⁴ Policy guidance from ED's Office for Civil Rights indicates that such assessments should be undertaken especially, but not only, for purposes of assigning pupils to instructional programs targeted at LEP pupils, determining the timing of transition to regular or mainstream instruction for such pupils, and evaluating the effectiveness of special programs for LEP pupils; although this guidance is unspecific regarding the type of assessment LEAs should use.⁵⁵

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 pupils 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.

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 NCLBA'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,

⁵⁴ See National Research Council, *Improving Schooling for Language-Minority Children: A Research Agenda*, 1997, pp. 115-116.

⁵⁵ See [<http://www.ed.gov/about/offices/list/ocr/docs/laumemos.html>].

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 NCLBA, 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 NCLBA for state bonuses and sanctions based in part on NAEP score trends were eliminated from the conference version. Under the NCLBA 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 2 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 NCLBA 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 Pupils. 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 pupils, schools, and possibly LEAs? While participation by states, LEAs, schools, and pupils was voluntary under previous *federal* law and policy, states or LEAs were not prohibited from requiring participation by LEAs, schools, or pupils under their own laws or policies. However, as noted earlier (p. 13), 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 pupils nationwide, resistance to participation in NAEP might grow to an extent that it threatens the quality of the national sample of tested pupils 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 pupils to participate in

order to produce valid and reliable assessment results. In the recent past, some states have attempted to participate in NAEP but found themselves unable to induce sufficient numbers of LEAs or schools to do so.⁵⁶

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 pupils, 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 which 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 pupil 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 — and 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 pupils.

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.”⁵⁷ 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;

⁵⁶ 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*, Feb. 16, 2000.

⁵⁷ Mark D. Rebase, “Using NAEP to Confirm State Test Results: An Analysis of Issues,” *Will No Child Truly Be Left Behind?*, published by the Thomas B. Fordham Foundation, Feb. 2002, p. 14.

(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 pupils 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 pupil scores. For example, the scores of several pupils 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 pupils 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.⁵⁸

What Are the 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 pupil 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 pupils, schools, and LEAs, throughout most of the elementary and middle school grades. While a majority of pupils 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 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 pupil 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 NCLBA, which involves increased state and local

⁵⁸ See the report for details. Available at [<http://www.nagb.org>].

flexibility in the use of federal education assistance funds.⁵⁹ Under this strategy, accountability for appropriate use of federal aid funds is to be established more on the basis of pupil 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 pupil 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 a recent ED publication, “Testing for Results, Helping Families, Schools and Communities Understand and Improve Student Achievement,”⁶⁰ 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 most, if not all, states will choose to implement them in order to maintain Title I-A eligibility,⁶¹ then assessment requirements attached to an aid program focused on *disadvantaged pupils* would broadly influence policies regarding standards, assessments, and accountability affecting *all pupils* in the participating states. This would represent a substantial increase in federal influence in the assessment and accountability aspects of K-12 education policy.

In the majority of states which have not administered a consistent series of mandated, standards-based assessments in each of grades 3-8, this policy 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.

⁵⁹ 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)*, by Wayne Riddle. pp. 11-12.

⁶⁰ On the Internet, see [<http://www.ed.gov/nclb/accountability/ayp/testingforresults.html>].

⁶¹ Officials in the state of Vermont are reportedly considering terminating their participation in Title I-A to avoid implementing expanded assessment and accountability requirements. See “Dean Urges Rejection of Federal Funds,” *Burlington Free Press*, Apr. 19, 2002.

The authors of a recent study of the effects of high stakes 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.”⁶² At the least, annual testing of pupils in grades 3-8 would increase the importance of having tests which 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 is and would remain limited in several important respects. With the exception of the limited role of state NAEP tests, the standards and assessments would be totally selected by the states. ED would not have any authority 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 proposed increase in federal influence is constrained, and is balanced by a decrease of federal influence in certain other respects.

⁶² 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/>].