

National Assessment Governing Board

Executive Committee

February 28, 2013

AGENDA

4:30 pm	Welcome, Introductions, and Agenda Overview <i>David Driscoll, Chair</i>	
4:35 pm	Committee Issues and Challenges <i>Committee Chairs</i>	
4:45 pm	Report of Ad Hoc Committee on NAEP Background Information <i>Terry Holliday, Chair, Ad Hoc Committee</i>	See Ad Hoc Committee Tab
4:50 pm	Governing Board 25th Anniversary Planning <i>Alan Friedman, 25th Anniversary Planning Committee Chair</i>	Attachment A
4:55 pm	Updating Governing Board Policy: Reviewing the Past, Looking to the Future <i>Ray Fields, Assistant Director for Policy and Research</i>	Attachment B
5:15 pm	NAEP Schedule of Assessments: Long-term Considerations <i>Cornelia Orr, Executive Director</i>	Attachment C
5:25 pm	ACTION ITEM Status of FY 2013 Appropriation for NAEP and the Governing Board <i>Ray Fields</i>	Attachment D
	<i>Tentative: CLOSED SESSION 5:30 – 6:00 p.m.</i>	
5:30 pm	NAEP Contracts, Budget, and Schedule for 2013 and Beyond <i>Cornelia Orr</i> <i>Peggy Carr, Associate Commissioner, NCES</i>	

Governing Board 25th Anniversary Planning Committee

Current Board Members

Alan J. Friedman, 25th Anniversary Committee Chair
Consultant
Museum Development and Science Communication
New York, New York

Shannon Garrison
Fourth-Grade Teacher
Solano Avenue Elementary School
Los Angeles, California

Terry Mazany
President and CEO
The Chicago Community Trust
Chicago, Illinois

Tonya Miles
General Public Representative
Mitchellville, MD

B. Fielding Rolston
Chairman
Tennessee State Board of Education
Kingsport, Tennessee

Board Alumni*

Amanda Avallone
8th Grade Teacher
Boulder, Colorado

Michael Guerra
Non-Public School Representative
Washington, DC

Mark Musick
General Public Representative
Atlanta, Georgia

Michael Nettles
Testing and Measurement Expert
Ann Arbor, Michigan

Mary Frances Taymans, SND
Non-Public School Representative
Washington, DC

Eileen Weiser
General Public Representative – 2nd term
State School Board Member – 1st term
Ann Arbor, Michigan

*Alumni are listed with their category and geographic location while serving on the Board.

Updating Governing Board Policy: Reviewing the Past, Looking to the Future

At its November 29, 2012 meeting, the Executive Committee reviewed a “track-changes” version of the Governing Board policy entitled “Redesigning the National Assessment of Educational Progress” (included in this attachment), having decided earlier that the policy should be updated.

The Executive Committee directed staff to present recommendations for changes to the policy for discussion by the Committee and the full Board at the February 28-March 2, 2013 quarterly meeting. Accordingly, staff prepared the side-by-side analysis that appears on the immediately following pages.

The side-by-side analysis presents a distillation of the current policy in the left-hand column with proposed changes highlighted in the right-hand column. The purpose is to focus the discussion on key ideas in this initial step, and incorporate the results of the discussion in a draft policy statement that would be presented at the May 2013 Board meeting.

In addition to the materials in this attachment, please also consider the materials in the tab for the NCES Workshop on the Future of NAEP.

Please note that the purpose of the proposed policy is quite different from the current policy. The intent of the current policy, adopted in 1996, when state assessments, achievement level reporting, and the relationship between the Governing Board and NCES were in early stages of development, was to make NAEP more efficient in order to provide achievement results to the public more regularly. Hence, “redesigning” is the operative term in the title.

In contrast, the intent of the proposed changes is to build on the experience of the last 25 years to provide a general policy for NAEP that acknowledges its core functions and associated values, and that will serve as an enduring compass for the conduct of the program long into the future.

The proposed policy is based on the current NAEP legislation. Therefore, no changes are proposed that would require authorization, such as provisions of the law affecting mandated subjects or the grade levels assessed.

DRAFT
Updating Governing Board Policy for NAEP
Side-by-Side Analysis of Current Policy and Proposed Changes, by Section

Current Policy: Redesigning the National Assessment of Educational Progress	Proposed Changes: Policy for the National Assessment of Educational Progress
<p>Foreword: Explains changes in the context since the 1996 adoption relevant to the policy statement.</p> <p>Introduction: “A Better Way to Measure Educational Progress in America” Explains what NAEP is and argues that it needs to be made more efficient to better serve the public by providing more timely achievement data on a more frequent basis.</p> <p>Argues that NAEP cannot do all that some would ask; tradeoffs must be made among desirable activities; NAEP should “stick to its knitting.”</p> <p>National Assessment Redesign: Explains why NAEP needs to be changed, as well as the underlying values and intended impact of the policy statement:</p> <ol style="list-style-type: none"> 1. To provide the public with more frequent information in more subjects 2. Reliability, validity, and quality of data to remain hallmarks of NAEP 3. Intent of policy is to guide current operations, new requests for proposals, and the activities and structure of the Governing Board 	<p>Foreword: Should explain the purpose of the policy statement—To serve as the general policy for the conduct and reporting of the National Assessment of Educational Progress.</p> <p>Introduction: Should explain why NAEP as a national and state indicator of student achievement is needed:</p> <ol style="list-style-type: none"> 1. Education is a foundation for a vibrant democracy, economic competitiveness and national security, important for the nation, states and individuals 2. Education achievement is a measure of human capital 3. Knowing how well students are achieving nationally and state-by-state is essential for the public and policymakers 4. NAEP is uniquely positioned to provide information regularly about student achievement at grades 4, 8, and 12 across a wide range of important subjects <p>Background: History and Evolution of NAEP: Should provide a brief overview of the forces that brought NAEP into existence; how it has evolved to reflect changes in curricula, assessment technology, reporting methods, and policy context; and that NAEP continues to evolve to better serve the public.</p> <p>Should describe core elements of NAEP:</p> <ol style="list-style-type: none"> 1. National, state and urban district samples 2. Broad coverage of a wide range of subjects 3. Trends over time 4. Group, not individual results 5. Demographic subgroup and gap reporting 6. Grade-based reporting: 4, 8, and 12 7. Reliability, validity, data quality, innovation, integrity, freedom from ideology& partisanship 8. Continuing, informed responsiveness to changing policy & technical environment while maintaining role of independent monitor

Prepared by: Ray Fields; February 8, 2013

Current Policy	Proposed Change
<p>Purpose and Objectives of NAEP: Purpose--“...to provide a fair and accurate presentation of educational achievement in reading, writing, and the other subjects included in the third National Education Goal, regarding student achievement and citizenship.”</p> <p>Objectives:</p> <ol style="list-style-type: none"> 1. To measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations; 2. To develop, through a broadly inclusive process, sound assessments to measure what students know and can do as well what students should know and be able to do; and 3. To help states and others link their assessments with the National Assessment and use National Assessment data to improve education performance. 	<p>Purpose and Objectives of NAEP: Purpose: to measure and report, in a timely manner, fair and accurate information on student academic achievement in reading, mathematics, and other subject matter, including achievement trends over time.</p> <p>Objectives:</p> <ol style="list-style-type: none"> 1. To provide timely, fair, and accurate data about student achievement at grades 4, 8, and 12 in important subject matter at the national level in public and private schools, among the states, for selected urban districts, and, as feasible, in comparison with other nations; 2. To develop, through a broadly inclusive process, technically sound, relevant assessments, with breadth and depth of coverage in each subject area, designed to measure what students know and can do, what students <u>should</u> know and be able to do, and trends over time; 3. To serve as a consistent external, independent measure by which state achievement results can be compared; 4. To bring attention to achievement gaps where they may exist between demographic subgroups and the urgency of closing those achievement gaps; 5. To continue to set and report achievement levels for NAEP results, with appropriate caveats, continuing validity studies, and in association with other relevant statistics, such as percentile scores across the full range of student performance on each assessment and average scores; 6. To help states and others use National Assessment data and resources to improve education performance and close achievement gaps; 7. To continue to innovate in NAEP, and incorporate technology, as appropriate, in framework development, assessment design, item development, assessment administration, data collection, test security, scoring, analysis and reporting.

Current Policy	Proposed Change
<p>The Audience for the National Assessment The primary audience is the American public.</p> <p>Results should be released within 6 months of testing. Reports should be understandable, free of jargon, easy to use and widely disseminated. Reports will not trade accuracy for simplicity; validity and reliability will be maintained. Test questions and student responses should be given more prominence in reports, to illustrate the meaning of the NAEP achievement levels.</p> <p>Principal users of NAEP data are national and state policymakers and educators. NAEP data will be available to help them interpret their results to the public, improve education performance, and perform secondary analysis.</p> <p>Limitations: What the National Assessment Is Not NAEP describes how well students perform, but does not explain why; provides group results, not individual results; does not collect information on personal values or attitudes; is independent of particular curricula; does not promote specific ideas, ideologies, or teaching techniques; is not, by itself, appropriate for improving achievement in particular classrooms, evaluating specific teaching practices, or determining whether particular approaches to curricula are working.</p> <p>Objective 1: To measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations.</p> <p>Assess all subjects specified by Congress: reading, writing, mathematics, science, history, geography, civics, the arts, foreign language, and economics.</p> <p>The National Assessment shall be conducted annually, two or three subjects per year, in order to cover all required subjects at least twice a decade.</p>	<p>The Audience for the National Assessment The primary audience is the American public.</p> <p>Results should be released within 6 months of testing. Reports should be understandable, free of jargon, easy to use and widely disseminated. Reports will not trade accuracy for simplicity; validity and reliability will be maintained. Test questions and student responses should be given prominence in reporting results, to illustrate the meaning of the NAEP scale and achievement levels.</p> <p>Principal users of NAEP data are national and state policymakers and educators. NAEP data will be available to help them interpret their results to the public, improve education performance, and perform secondary analysis.</p> <p>Limitations: What the National Assessment Is Not NAEP describes how well students perform, but does not explain why; provides group results, not individual results; does not collect information on personal values or attitudes; is independent of particular curricula; does not promote specific ideas, ideologies, or teaching techniques; is not, by itself, appropriate for improving achievement in particular classrooms, evaluating specific teaching practices, or determining whether particular approaches to curricula are working.</p> <p>Objective 1: To provide timely, fair, and accurate data about student achievement at grades 4, 8, and 12 in important subject matter at the national level in public and private schools, at the state level, for selected urban districts, and, as feasible, in comparison with other nations.</p> <p>As specified by Congress, conduct assessments in reading and mathematics, and, as time and resources allow: in science, writing, history, civics, geography, the arts, foreign language, economics, and other areas as determined by the National Assessment Governing Board.</p> <p>The National Assessment shall be conducted annually, two or three subjects per year, with the goal for the frequency of each subject area assessment of at least twice in a ten-year period.</p>

Current Policy:	Proposed Change
<p>The National Assessment shall assess all subjects listed in the third National Educational Goal—reading, writing, mathematics, science, history, geography, civics, the arts, foreign language and economics—according to a publicly released schedule adopted by the National Assessment Governing Board, covering eight to ten years, with reading, writing, mathematics, and science tested more frequently than the other subjects.</p> <p>The National Assessment Governing Board shall consult with technical experts and with education policymakers, in conjunction with the development of assessment frameworks, to determine the feasibility, desirability, and costs of combining several related subjects into a single assessment.</p> <p>Provide National Assessment results for states</p> <p>National Assessment state-level assessments shall be conducted on a reliable, predictable schedule according to an eight to ten year plan adopted by the National Assessment Governing Board.</p> <p>Reading, writing, mathematics, and science at grades 4 and 8 shall be given priority for National Assessment state-level assessments.</p> <p>States shall have the option to use National Assessment tests in other subjects and at grade 12 by assuming a larger share of the costs and adhering to requirements that protect the integrity of the National Assessment program. However, the National Assessment Governing Board shall seek ways to make such use of National Assessment tests attractive and financially feasible.</p> <p>Where possible, changes in national and state sampling procedures shall be made that will reduce burden on states, increase efficiency, and save costs.</p>	<p>The National Assessment Governing Board, in consultation with the National Center for Education Statistics, shall establish a dependable, publicly announced assessment schedule of at least ten years in scope, specifying the subject or topic (e.g., High School Transcript Study), grades, year, and sampling levels (e.g., national, state) for the conduct of each assessment.</p> <p>To ensure that nationally representative samples are achieved, the National Assessment Governing Board and the National Center for Education Statistics shall develop and implement a plan to foster state cooperation in securing school participation in non-mandated subjects and grades.</p> <p>Provide National Assessment results for states</p> <p>Note: see first paragraph above on this page in this column.</p> <p>Reading, writing, mathematics, science, U.S. history, and civics shall be given priority for National Assessment state-level assessments, with the assessments to be administered at grades 4, 8, and/or 12, as appropriate.</p> <p>States shall have the option to use National Assessment tests not administered below the national level by assuming the costs and adhering to requirements that protect the integrity of the National Assessment program. The National Assessment Governing Board and NCES shall seek ways to ensure that state decision makers receive timely notice of this option and that the cost to states is minimized.</p> <p>On a continuing basis, the National Center for Education Statistics shall seek innovations in national and state sampling procedures to reduce burden on states, increase efficiency, and minimize costs.</p>

Current Policy:	Proposed Change
<p>Vary the amount of detail in testing and in reporting results</p> <p>National Assessment testing and reporting shall vary, using standard report cards most frequently, comprehensive reporting in selected subjects about once every ten years, and special, focused assessments.</p> <p>National Assessment results shall be timely, with the goal being to release results within 6 months of the completion of testing for standard report cards and within 9 months for comprehensive reports.</p> <p>Simplify the National Assessment design</p> <p>Options shall be identified to simplify the design of the National Assessment.</p> <p>Simplify the way the National Assessment reports trends in student achievement</p> <p>A carefully planned transition shall be developed to enable "the main National Assessment," to become the primary way to measure trends in reading, writing, mathematics, and science in the National Assessment program.</p>	<p>Use technology to disseminate NAEP reports and to make NAEP data and information easily accessible to various audiences, including educators, policymakers and the public</p> <p>The National Assessment shall rely primarily on electronic copies of reports for dissemination to the public and minimize the number of printed copies of NAEP reports.</p> <p>On a continuing basis, the National Assessment Governing Board and the National Center for Education Statistics shall evaluate audience needs and employ innovations in technology to improve access, usability, and usefulness of NAEP data and information.</p> <p>Provide timely results</p> <p>National Assessment results shall be timely, with the goal of releasing results within 6 months of the completion of testing.</p> <p>Collect and report background information directly related to achievement</p> <p>Note: Board-adopted recommendations of the Ad Hoc Committee on NAEP Background Information will be incorporated here</p> <p>Note: Not applicable</p> <p>Simplify the way the National Assessment reports trends in student achievement</p> <p>The "main National Assessment" shall be the primary way to measure trends in reading and mathematics in the National Assessment program. The frequency of administering the "long-term trend" assessments in reading and mathematics shall be reduced.</p>

Current Policy:	Proposed Change
<p data-bbox="94 247 721 321">Use performance standards to report whether student achievement is "good enough"</p> <p data-bbox="191 359 748 468">The National Assessment shall continue to report student achievement results based on performance standards.</p> <p data-bbox="94 520 509 558">Use international comparisons</p> <p data-bbox="191 590 764 772">National Assessment test frameworks, test specifications, achievement levels, and data interpretations shall take into account, where feasible, curricula, standards, and student performance in other nations.</p> <p data-bbox="191 810 699 919">The National Assessment shall promote "linking" studies with international assessments.</p>	<p data-bbox="824 247 1520 321">Note: see Objective 5 below ("To continue to set and report achievement levels for NAEP results...")</p> <p data-bbox="824 520 1240 558">Use international comparisons</p> <p data-bbox="922 590 1520 772">National Assessment frameworks, specifications, achievement levels, and data interpretations shall take into account, where feasible, curricula, performance standards, and student achievement in other nations.</p> <p data-bbox="922 810 1520 1024">In making decisions about the NAEP schedule of assessments, the National Assessment Governing Board shall consider opportunities to conduct studies linking NAEP with international assessments, in consultation with the National Center for Education Statistics.</p> <p data-bbox="922 1062 1490 1209">To the extent feasible, studies linking NAEP with international assessments shall be conducted by the National Center for Education Statistics.</p>

Current Policy:	Proposed Change
<p data-bbox="94 184 691 216">Emphasize reporting for grades 4, 8, and 12</p> <p data-bbox="191 583 789 726">The National Assessment shall continue to test in and report results for grades 4, 8, and 12; however, in selected subjects, one or more of these grades may not be tested.</p> <p data-bbox="191 842 797 1056">Age-based testing and reporting shall be permitted when deemed appropriate and when necessary for international comparisons and for long-term trends, should the National Assessment Governing Board decide to continue long-term trends in their current form.</p> <p data-bbox="191 1098 789 1276">Grade 12 results shall be accompanied by clear, highlighted statements about school and student participation, student motivation, and cautions, where appropriate, about interpreting 12th grade achievement results.</p> <p data-bbox="191 1350 781 1459">The National Assessment design shall seek to improve school and student participation rates and student motivation at grade 12.</p> <p data-bbox="191 1644 727 1753">The National Assessment shall provide practical incentives for school and district participation at grades 4, 8, and 12.</p>	<p data-bbox="824 184 1523 254">Collect and report valid and reliable data at grades 4, 8, and 12</p> <p data-bbox="922 289 1515 546">All aspects of the National Assessment program, its administration and its governance shall be: conducted in a manner that advances the integrity of the program and the credibility of the results; subject to rigorous statistical standards; and free of partisanship, ideology, and curricular or pedagogical bias.</p> <p data-bbox="922 583 1503 798">The National Assessment shall continue to assess and report results for grades 4, 8, and 12. However, the National Assessment Governing Board may determine that, in selected subjects, one or more of these grades may not be assessed.</p> <p data-bbox="922 842 1528 1056">Age-based testing and reporting shall be permitted when deemed appropriate and when necessary for international comparisons and for long-term trends, should the National Assessment Governing Board decide to continue long-term trends in their current form.</p> <p data-bbox="922 1098 1528 1312">NAEP results shall be accompanied by clear, highlighted statements about school and student participation rates; indicators of student engagement in the assessment when feasible; and cautions, where appropriate, about interpreting achievement results.</p> <p data-bbox="922 1350 1511 1606">The National Assessment program shall take affirmative steps to achieve statistically sound levels of school and student participation and optimal levels of student engagement, including steps to maximize the participation of students with disabilities and English language learners.</p> <p data-bbox="922 1644 1466 1753">The National Assessment program shall provide practical incentives for school and district participation at grades 4, 8, and 12.</p>

Current Policy:	Proposed Change
<p>Use innovations in measurement and reporting</p> <p>The National Assessment shall assess the merits of advances related to technology and the measurement and reporting of student achievement.</p> <p>Where warranted, the National Assessment shall implement such advances in order to reduce costs and/or improve test administration, measurement, and reporting.</p> <p>The next competition for National Assessment contracts, for assessments beginning around the year 2000, shall ask bidders to provide a plan for</p> <ol style="list-style-type: none"> (1) conducting testing by computer in at least one subject at one grade, and (2) making use of technology to improve test administration, measurement, and reporting. <p>OBJECTIVE 2: To develop, through a broadly inclusive process, sound assessments to measure what students know and can do as well as what students <u>should</u> know and be able to do.</p> <p>Keep test frameworks and specifications stable</p> <p>Test frameworks and test specifications developed for the National Assessment generally shall remain stable for at least ten years.</p>	<p>Note: see Objective 7 below: (“To continue to innovate in NAEP, and incorporate technology...”)</p> <p>OBJECTIVE 2: To develop, through a broadly inclusive process, technically sound, relevant assessments, with breadth and depth of coverage in each subject area, designed to measure what students know and can do, what students <u>should</u> know and be able to do, and trends over time.</p> <p>Develop assessment frameworks with input from the widest possible range of informed practitioners, researchers, and stakeholders</p> <p>Regularly evaluate the extent to which the set of tasks and items for an assessment meets the framework requirements, assessment specifications, and achievement level descriptions</p> <p>Keep assessment frameworks and specifications stable to report trends in achievement</p> <p>To ensure that trend results can be reported, assessment frameworks and specifications developed for the National Assessment generally shall remain stable for at least ten years</p>

Current Policy:	Proposed Change
<p>To ensure that trend results can be reported, the pool of test questions developed in each subject for the National Assessment shall provide a stable measure of student performance for at least ten years.</p> <p>In rare circumstances, such as where significant changes in curricula have occurred, the National Assessment Governing Board may consider making changes to test frameworks and specifications before ten years have elapsed.</p> <p>In developing new test frameworks and specifications, or in making major alterations to approved frameworks and specifications, the cost of the resulting assessment shall be estimated. The National Assessment Governing Board will consider the effect of that cost on the ability to test other subjects before approving a proposed test framework and/or specifications.</p> <p>Use an appropriate mix of multiple-choice and "performance" questions</p> <p>Both multiple-choice and performance items shall continue to be used in the National Assessment;</p> <p>In developing new test frameworks, specifications, and questions, decisions about the appropriate mix of multiple-choice and performance items shall take into account the nature of the subject, the range of skills to be assessed, and cost.</p>	<p>To ensure that trend results can be reported, the pool of items developed in each subject for the National Assessment shall provide a stable measure of student performance for at least ten years.</p> <p>In rare circumstances, such as where significant changes in curricula have occurred, the National Assessment Governing Board may consider making changes to assessment frameworks and specifications before ten years have elapsed.</p> <p>In deciding whether to develop new assessment frameworks and specifications, or to make major alterations to approved frameworks and specifications, the Governing Board shall consider the impact on reporting trends. Whenever feasible, technically defensible steps shall be taken to avoid breaking trend lines.</p> <p>In developing new assessment frameworks and specifications, or in making major alterations to approved frameworks and specifications, the cost of the resulting assessment shall be estimated. The National Assessment Governing Board will consider the effect of that cost on the ability to assess other subjects before approving a proposed assess framework and/or specifications.</p> <p>Develop NAEP assessments, using technology as appropriate, that comprehensively measure the subject domain as defined by the assessment framework and specifications.</p>

Current Policy:	Proposed Change
<p>OBJECTIVE 3: To help states and others link their assessments with the National Assessment and use National Assessment data to improve education performance.</p> <p>The National Assessment shall develop policies, practices, and procedures that assist states, school districts, and others who want to do so at their own cost to link their test results to the National Assessment.</p> <p>The National Assessment shall be designed so that others may access and use National Assessment test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data.</p> <p>The National Assessment shall employ safeguards to protect the integrity of the National Assessment program, prevent misuse of data, and ensure the privacy of individual test takers.</p>	<p>Note: See Objective 6 below.</p> <p>OBJECTIVE 3. To serve as a consistent external, independent measure by which state achievement results can be compared.</p> <p>OBJECTIVE 4. To bring attention to achievement gaps where they may exist between demographic subgroups and the urgency of closing those achievement gaps.</p> <p>OBJECTIVE 5. To continue to set and report achievement levels for NAEP results, with appropriate caveats, continuing validity studies, and in association with other relevant statistics, such as percentile scores across the full range of student performance on each assessment and average scores.</p>

Current Policy:	Proposed Change
<p>OBJECTIVE 3. To help states and others link their assessments with the National Assessment and use National Assessment data to improve education performance.</p> <p>The National Assessment shall be designed so that others may access and use National Assessment test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data.</p> <p>The National Assessment shall employ safeguards to protect the integrity of the National Assessment program, prevent misuse of data, and ensure the privacy of individual test takers.</p>	<p>OBJECTIVE 6. To help states and others use National Assessment data and resources to improve education performance and close achievement gaps.</p> <p>The National Assessment program shall take affirmative steps to facilitate access to and use of National Assessment frameworks, specifications, scoring guides, results, questions, achievement levels, and background data by states and other appropriate parties.</p> <p>The National Center for Education Statistics shall employ safeguards to protect the integrity of the National Assessment program, prevent misuse of data, and ensure the privacy of individual test takers.</p> <p>The National Center for Education Statistics shall develop procedures to facilitate the ability of states to link performance on NAEP with data in state longitudinal data bases.</p> <p>OBJECTIVE 7. To continue to innovate in NAEP, and incorporate technology, as appropriate, in framework development, assessment design, item development, assessment administration, data collection, test security, scoring, analysis and reporting.</p>



Adopted: August 2, 1996

National Assessment Governing Board

Redesigning the National Assessment of Educational Progress

Policy Statement

Foreword

This policy statement was adopted in 1996, at a time when Congress had codified National Education Goals targeted for accomplishment by the year 2000. It was the expectation that the National Assessment of Educational Progress would be a primary means for monitoring progress in achieving the goal addressing student achievement and this expectation is reflected in the policy below. The National Education Goals legislation is no longer in effect and has been superseded by other national policies, the No Child Left Behind Act of 2001 (NCLB) being the most germane. Therefore, the references to National Education Goals in this policy statement are no longer relevant.

Under NCLB, state level participation in assessments in reading and mathematics in grades 4 and 8 became mandatory. Participation is required on a biennial basis, affecting costs and technical design. However, the overall intent and impact of the policy—to clarify purpose, define the audience, set forth limitations, maintain quality and integrity, and bring efficiencies to the design of the assessment—remain in effect and continue to guide the policy setting and operations of the National Assessment. (Foreword added August 2007.)

Comment [AU1]: How should we address the Common Core State Standards and Assessments Initiative in the policy?

Comment [AU2]: This NCLB requirement solidified a role for NAEP as an independent, external measure of achievement, comparable across all states.

The current policy does not convey this as a role of NAEP per se, but as an attribute that states use at their discretion (see p. 6 comment AU19).

Comment [AU3]: Should the overall intent of the policy statement be amended?

A Better Way to Measure Educational Progress in America

An effective democracy and a strong economy require well-educated citizens. A good education lays a foundation for getting a good job, leading a fulfilling life, and participating constructively in society.

But is the education provided in your state and in America good enough? How do our 12th graders compare with students in other nations in mathematics and science? Do our 8th grade students have an adequate understanding of the workings of our constitutional democracy? How well do our 4th grade students read, write, and compute? The National Assessment of Educational Progress is the only way for the public to know with accuracy how American students are achieving nationally and state-by-state.

Comment [AU4]: Should this statement be refined?

The National Assessment tests at grades 4, 8, and 12. By law, it covers ten subjects, including reading, writing, mathematics, and science. The National Assessment has performance standards that indicate whether student achievement is "good enough." The National Assessment is not a national exam taken by all students. In fact, only several thousand students are tested per grade, comprising carefully drawn samples that represent the nation and the participating states. Since its first test in 1969, the National Assessment has earned a trusted reputation for its quality and credibility. That reputation must be maintained.

The National Assessment is unique because of its national, state-by-state, and 12th grade results. State and local test results cannot be used to provide a national picture of student achievement. States and local schools use different tests that vary in many ways. The results cannot simply be "added up" to get a national score nor can state scores on their different tests be compared. The National Assessment Governing Board believes that twelfth grade achievement is important to monitor at the national level, because the 12th grade marks the end of elementary and secondary education, the transition point for most students from school to work, to college, or to technical training. The National Assessment is the only source of nationally representative data at the 12th grade. College entrance tests such as the ACT and the SAT are taken only by students planning on higher education; the results do not represent the achievement of the total 12th grade class. And to date, virtually no state-based assessment program tests 12th graders.

Comment [AU5]: The seeds of the Board's initiative on 12th grade academic preparedness were first sown here...how should we incorporate academic preparedness?

While there is much about the National Assessment that is working well, there is a problem. Under its current design, the National Assessment tests too few subjects, too infrequently, and reports achievement results too late—as much as 18 to 24 months after testing. Testing occurs every other year. During the 1990's, only reading and mathematics will be tested more than once using up-to-date tests and performance standards. Six subjects will be tested only once and two subjects not at all during the 1990's.

Comment [AU6]: Much of this is no longer true or out of date...testing is now scheduled for every year and NCES has set a performance standard for completing NAEP Report Cards in time for release by six months from the end of testing for 4th and 8th grade reading and math and twelve months for other report cards.

Why is the National Assessment testing so few subjects and fewer subjects now than years ago? Over the years, the National Assessment has become increasingly complex. Its quality and integrity have led to a multitude of demands and expectations beyond its central purpose. Meeting those expectations was done with good intentions and seemed right for the situation at the time. However, additions to the National Assessment have been "tacked on" without changing the basic design, driving up costs and reducing the number of subjects that can be tested.

For example, where a single 120 page mathematics report once sufficed, mathematics reporting in 1992 consisted of seven volumes totaling almost 1,800 pages, not including individual state reports. Also, there are now two separate testing programs for reading, writing, math, and science. One monitors trends using tests developed during the

1970's; the other reflects current views on instruction and uses performance standards to report whether achievement is good enough.

The current National Assessment design is overburdened, inefficient, and redundant. It is unable to provide the frequent, timely reports on student achievement the American public needs. The challenge is to supply more information, more quickly, with the funding available.

To meet this challenge, the National Assessment design must be changed, building on its strengths while making it more efficient. The design of the National Assessment must be simplified. The purpose of the National Assessment must be sharply focused and its principal audience clearly defined. Because the National Assessment cannot do all that some would have it do, trade-offs must be made among desirable activities. Useful but less important activities may have to be reduced, eliminated, or carried out by others. The National Assessment must "stick to its knitting" in order to be more cost-effective, reach more of the public, provide more information more promptly, and maintain its integrity.

Comment [AU7]: How well has NAEP adhered to the policy of shedding all but essential functions and "sticking to its knitting?"

National Assessment Redesign

To provide the American public with more frequent information in more subjects about the progress of student achievement, changes must be made in the way that the National Assessment is designed and the results are reported. These changes are described in this policy statement. Undergirding these changes is an explicit statement of the purposes, objectives, audiences, and limitations of the National Assessment.

While change is in order, many current policies should continue. For example, reliability, validity, and quality of data will remain hallmarks of the National Assessment. The sample of tested students will be as representative as possible, using policies and procedures that maximize the number of students included who are disabled or are of limited English proficiency. And reporting on trends over time will remain a central commitment of the National Assessment.

Comment [AU8]: Should this statement remain as is or be refined?

The intent of this policy statement is to guide current operations of the National Assessment, the development of new requests for proposals for contracts for conducting the National Assessment and the activities and structure of the National Assessment Governing Board. Contracts for current operations extend through assessments to be conducted in 1998. New contracts would cover assessments as early as 1999 and thereafter.

Comment [AU9]: Out of date.

Purpose and Objectives of the National Assessment of Educational Progress

The purpose of the National Assessment is stated in its legislation:

“...to provide a fair and accurate presentation of educational achievement in reading, writing, and the other subjects included in the third National Education Goal, regarding student achievement and citizenship.”

Comment [AU10]: Out of date—superseded by current purpose statement in the NAEP law: “The purpose of this section is to provide, in a timely manner, a fair and accurate measurement of student academic achievement and reporting of trends in such achievement in reading, mathematics, and other subject matter as specified in this section.”

Thus, the central concern of the National Assessment is to inform the nation on the status of student achievement. The National Assessment Governing Board believes that this should be accomplished through the following objectives:

1. To measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations;

Comment [AU11]: Out of date

2. To develop, through a broadly inclusive process, sound assessments to measure what students know and can do as well what students should know and be able to do; and

3. To help states and others link their assessments with the National Assessment and use National Assessment data to improve education performance.

Comment [AU12]: How should these objectives be amended?

Should “conveying the urgency of improving achievement and closing achievement gaps” be incorporated?

The specific changes in the design of the National Assessment described below are discussed in relation to these objectives.

The Audience for the National Assessment

Comment [AU13]: Does the audience statement need revision?

The primary audience for National Assessment results is the American public, including the general public in states that receive their own results from the National Assessment. Reports should be written for this audience. Results should be released within 6 months of testing. Reports should be understandable, free of jargon, easy to use, and widely disseminated. Although more comprehensible, direct, and useful, the reports will not trade accuracy for simplicity. The tradition of high quality of National Assessment reports will be continued, with no erosion of validity and reliability. Assessment questions and samples of student work that illustrate performance standards are likely to receive heightened prominence in reports.

Principal users of National Assessment data are national and state policymakers and educators concerned with student achievement, curricula, testing, and standards. National Assessment data will be available to these users in forms that support their efforts to interpret results to the public, to improve education performance, and to perform secondary analysis.

Comment [AU14]: Have we done enough in these areas?

Limitations: What the National Assessment Is Not

Comment [AU15]: Does the limitations statement need revision?

The National Assessment is intended to describe how well students are performing, but not to explain why. The National Assessment only provides group results; it is not an individual student test. The National Assessment tests academic subjects and does not

collect information on individual students' personal values or attitudes. Each National Assessment test is developed through a national consensus process. This national consensus process takes into account education practices, the results of education research, and changes in the curricula. However, the National Assessment is independent of any particular curriculum and does not promote specific ideas, ideologies, or teaching techniques. Nor is the National Assessment an appropriate means, by itself, for improving instruction in individual classrooms, evaluating the effects of specific teaching practices, or determining whether particular approaches to curricula are working.

OBJECTIVE 1: To measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations.

Assess all subjects specified by Congress: reading, writing, mathematics, science, history, geography, civics, the arts, foreign language, and economics.

The gap must be closed between the number of subjects the National Assessment is required to assess and the number of subjects it can assess at the national level under the current design. By law, the National Assessment is required to assess ten subjects and report results and trends. In order to chart progress and report trends, subjects must be assessed more than once. However, during the 1990's only reading and mathematics will have been assessed more than once using up-to-date tests and performance standards to report how well students are doing.

Some have suggested that a solution is to combine into a single assessment several related subjects (e.g. reading and writing and/or history, geography, civics, and economics). Under such an approach, assessment data would be reported using both an overall score and sub scores for the respective disciplines. Although such an approach has the appeal of reducing the number of separate assessments, its feasibility, desirability, and costs are unknown. Also, such an approach has far-reaching implications for the test frameworks that guide the development of each assessment and for reporting results. These implications must be considered carefully. For the immediate future, subjects will continue to be assessed separately. However, the National Assessment Governing Board is committed to providing the public with more information as efficiently as possible. The Governing Board will consult with technical experts and education policymakers, in conjunction with the development of assessment frameworks, to determine the feasibility, desirability, and costs of combining several related subjects into a single assessment.

- The National Assessment shall be conducted annually, two or three subjects per year, in order to cover all required subjects at least twice a decade.
- The National Assessment shall assess all subjects listed in the third National Educational Goal—reading, writing, mathematics, science, history, geography, civics, the arts, foreign language and economics—

Comment [AU16]: What should we say about the subjects NAEP should assess and the associated values conveyed through the breadth vs. narrowness of the subjects on the Governing Board schedule of assessments?

Comment [AU17]: Current law provides flexibility to conduct subjects/grades other than 4th and 8th grade reading and math “as time and resources permit” and provides the Board flexibility to conduct assessments in subjects beyond the ten that are named...this is the source of authority for the Technology and Engineering Literacy assessments, for example.

Should assessments in other areas be considered?

according to a publicly released schedule adopted by the National Assessment Governing Board, covering eight to ten years, with reading, writing, mathematics, and science tested more frequently than the other subjects.

- The National Assessment Governing Board shall consult with technical experts and with education policymakers, in conjunction with the development of assessment frameworks, to determine the feasibility, desirability, and costs of combining several related subjects into a single assessment.

Provide National Assessment results for states

In 1988, testing at the state level was added to the National Assessment as a trial, with participation strictly voluntary, subjects and grades specified in law, and an independent evaluation required. Previously, the National Assessment had reported only national and regional results. For the first time, the information was relevant to individuals in states who make decisions about education funding, governance, and policy. As a result, states now are major users of National Assessment data.

Participation was strong in the first state-level assessment in 1990 and has grown to include even more states. In 1996, 44 states and 3 jurisdictions participated in the mathematics assessments at grade 4 and 8 and the science assessment at grade 8. The independent evaluation concluded that the trial state assessments produced valid and reliable data. The evaluation report recommended, and Congress agreed, that state-level assessments, with continued evaluations, be included in the 1994 reauthorization of the National Assessment.

Currently, the National Assessment draws a separate sample to obtain national results in addition to the samples drawn for individual state reports. Keeping the schools drawn for national samples completely partitioned from the state samples increases costs and creates additional burdens on states, particularly small states. Options should be identified for making the national and state samples more efficient and less burdensome. For example, it may be possible to reduce the current state sample size of 100 schools to a smaller number (e.g. 65-75) without a great loss in precision.

Comment [AU18]: This is no longer an accurate statement.

States participate in the National Assessment for many reasons, including to have an unbiased, external benchmark to help them make judgments about their own tests and standards. National Assessment data are used to make comparisons to other states, to help determine if curriculum and standards are rigorous enough, to develop questions about curricular strengths and weaknesses, to make state to international comparisons, and to provide a general indicator of achievement.

There is a strong interest among states to participate in the National Assessment to get state level information at grades 4 and 8 in reading, writing, mathematics, and science. The level of interest in participating in the National Assessment varies with respect to the other subjects (i.e., history, geography, civics, economics, the arts, and foreign language)

and at grade 12, where state officials say that obtaining cooperation from high schools and 12th grade students is difficult.

Some states, however, would like to be able to use National Assessment tests in the other subjects and at grade 12. Such use of National Assessment tests would be conducted as a service, with the reporting of results and maintenance of data under the control of the state. States will be able to use National Assessment tests if they adhere to requirements to protect the integrity of the National Assessment program and pay the additional costs. At the present time, states that participate in the National Assessment to get state level information at grades 4 and 8 in reading, writing, mathematics, and science provide in-kind support to cover the cost of in-state coordination and test administration. The National Assessment program covers the majority of costs, including test development, sampling, analysis, and reporting. States that wish to use National Assessment tests in other subjects and at grade 12 would pay for much of these additional costs.

Comment [AU20]: This is no longer true under the current NAEP legislation...in-state coordination and test administration are now paid by the NAEP program,

States are active partners in the National Assessment program. States help develop National Assessment test frameworks, review test items, and assist in conducting the tests. The National Assessment program is effective, to a great degree, because of the involvement of the states.

Because it is useful to them, and because they invest time and resources in it, states want a dependable schedule for National Assessment testing. With a dependable schedule, states that want to will be better able to coordinate the National Assessment with their own state testing program and make better use of the National Assessment as an external reference point.

- **National Assessment state-level assessments shall be conducted on a reliable, predictable schedule according to an eight to ten year plan adopted by the National Assessment Governing Board.**
- **Reading, writing, mathematics, and science at grades 4 and 8 shall be given priority for National Assessment state-level assessments.**
- **States shall have the option to use National Assessment tests in other subjects and at grade 12 by assuming a larger share of the costs and adhering to requirements that protect the integrity of the National Assessment program. However, the National Assessment Governing Board shall seek ways to make such use of National Assessment tests attractive and financially feasible.**
- **Where possible, changes in national and state sampling procedures shall be made that will reduce burden on states, increase efficiency, and save costs.**

Comment [AU21]: This is still true, and the adoption of this policy in 1996 has led to the Board regularly adopting a long-term NAEP schedule of assessments that is the basis for NAEP operations and planning.

Comment [AU22]: What priorities should be set for assessments in the various subjects and grades?

Comment [AU23]: This option exists but has not been exercised.

Comment [AU24]: This is being implemented.

Vary the amount of detail in testing and in reporting results

More subjects can be assessed if different strategies are used. Currently, each time the National Assessment is conducted, it uses a similar approach, regardless of the nature of the subject or the number of times an assessment in a subject has been administered. This approach is locked-in through 1998 under current contracts. Under this approach, a larger

number of students is tested in order to provide not just overall results, but fine-grained details as well (e.g. the achievement scores of 4th grade students whose teachers that year had five hours or more of in-service training). The National Assessment also collects "background" information through questionnaires completed by students, teachers, and principals. The questionnaires ask about teaching practices, school policies, and television watching, to name a few. Data analyses are elaborate. Reports are detailed and exhaustive, involving as many as seven separate reports per subject. Although the National Assessment has been praised for this thoroughness, the cost of this thoroughness is that fewer subjects are assessed, assessments occur less frequently, and reports take longer to produce.

The different strategies needed might include several approaches to testing and reporting, all of which should be designed in ways that maintain the National Assessment's commitment to providing valid and reliable data of high quality. For example, these approaches could take the form of "standard report cards," "comprehensive reports," and special, focused assessments.

A standard report card would provide overall results in a subject with performance standards and average scores. Results for standard report cards could be reported by sex, race/ethnicity, socio-economic status, and for public and private schools, but would not be broken down further. This may reduce the number of students needed for testing and may reduce associated costs. Generally, subcategories within a subject (e.g. algebra, measurement, and geometry within mathematics) would not be reported. However, data from the National Assessment would continue to be available to state and local educators and policymakers for additional analysis.

Comprehensive reports, like the current approach, would be an in-depth look at a subject, perhaps using a newly adopted test framework, many students, many test questions, and ample background information. In addition to overall results using performance standards and average scores, subcategories within a subject could be reported. Results would be reported by sex, race/ethnicity, socio-economic status, and for public and private schools, and might be broken down further as well. In some cases, more than one report may be issued in a subject. Comprehensive reporting in a particular subject would occur infrequently, perhaps once in ten years, but under a planned schedule of assessments.

Special, focused assessments on timely topics also would be conducted. They would explore a particular question or issue and may be limited to particular grades. Generally, the cost would be less than the cost of a standard report card. Examples of these smaller-scale, focused assessments include: (1) assessing subjects using targeted approaches (e.g. 8th grade arts), (2) testing special populations (e.g. in-school 12th graders versus out-of-school youth), and (3) examining skills and knowledge across several subjects (e.g. readiness for work).

The use of background surveys also would be varied. The three kinds of background surveys—student, teacher and principal questionnaires—would not necessarily all be employed each time a subject is assessed. Instead, the use of such surveys would be limited and selective, with reports of results focused on a core of background questions addressing the most essential issues. Also, background surveys used for standard report

Comment [AU25]: The upcoming Mega-states report is an example, as are the reports on achievement gaps for black and Hispanic students.

cards in a particular year would be designed to complement, rather than duplicate, background surveys used for comprehensive reports in the same year.

- **National Assessment testing and reporting shall vary, using standard report cards most frequently, comprehensive reporting in selected subjects about once every ten years, and special, focused assessments.**
- **National Assessment results shall be timely, with the goal being to release results within 6 months of the completion of testing for standard report cards and within 9 months for comprehensive reports.**

Comment [AU26]: The Ad Hoc Committee on NAEP Background Information is being established to address issues such as these; the Committee's recommendations should be considered in the update of this policy.

Simplify the National Assessment design

The current design of the National Assessment is very complex and, in fact, has grown more complex over the years. Here are just three examples of this complexity. (1) No student takes the complete set of test questions in a subject and as many as twenty-six different test booklets are used within each grade. Scores are calculated using sophisticated statistical procedures. (2) Students, teachers, and principals complete separate background questionnaires and may submit them for scoring at different times. Data from the questionnaires are used in calculating results of the assessments. (3) Current requirements for data analysis demand that test scores be calculated for every background variable collected by the National Assessment before any report can be produced. This lengthens the time from data collection to reporting and adds significantly to cost.

The design became more complex, in part, because the National Assessment's purposes and audiences had proliferated and the amount of background information collected had expanded. Specifying the purposes, audiences, and limitations of the National Assessment, as well as providing for varied means for testing and reporting, will result in opportunities for simplifying the National Assessment design.

- **Options shall be identified to simplify the design of the National Assessment.**

Simplify the way the National Assessment reports trends in student achievement

From its beginning in 1969, monitoring achievement trends has been a central mission of the National Assessment of Educational Progress. Monitoring long-term trends in educational achievement, both for the population as a whole and for significant sub-groups, is a capacity unique to the National Assessment and should be continued as a central mission. However, as the National Assessment approaches its third decade, it must address the problem of how to assess trends in achievement when curricula continue to evolve and change. An assessment in a subject must be kept stable to monitor trends. However, stable assessments may not reflect important changes in curricula. Over time, there develops a

legitimate concern about the relevance of the content of the assessment versus the ability to track change in achievement.

As a solution to this problem, since 1990, the National Assessment has reported achievement trends using two unconnected assessment programs. The tests, criteria for selecting students, and reporting are all different. The first program, "the main National Assessment," tests at grades 4, 8, and 12 and covers ten subjects. The assessments are based on a national consensus representing current views of each subject. Performance standards are used to report whether student achievement on the National Assessment is "good enough." The schedule of subjects to be assessed in the main National Assessment is unrelated to the schedule of subjects under the second testing program.

The second assessment program reports long-term trends that go as far back as 1970. Only four subjects are covered: reading, writing, mathematics, and science. The assessments are based on views of the curricula prevalent during the 1970's and have not been changed. Testing is at ages 9, 13, and 17 except for writing, which tests at grades 4, 8, and 11. Trends are reported by average score; performance standards are not used. The long-term trend program has been valuable for documenting declines and increases in student achievement over time and a decrease in the achievement gap between minority and non-minority students.

It may be impractical and unnecessary to operate two separate assessment programs. However, it also is likely that curricula will continue to change and that current test frameworks may be less relevant in the future. The tension between the need for stable measures of student achievement and changing curricula should be recognized as a continuing policy matter for the National Assessment, requiring efficient and balanced design solutions. Among the factors to consider are: (1) setting a standard period of time for a long-term trend (e.g. 15-20 years) using a particular "metric" in a subject; (2) providing for overlapping administrations of old and new assessments and "bridge" studies to determine whether the new can be linked to the old assessment; and (3) periodic administration of older assessments (e.g. once every ten years once a new trend-line has been established so that it would be possible to compare performance in 2010 with that in 1970 on the old trend line and with that in 1990 on a new trend line).

- **A carefully planned transition shall be developed to enable "the main National Assessment," to become the primary way to measure trends in reading, writing, mathematics, and science in the National Assessment program.**

Use performance standards to report whether student achievement is "good enough"

In reporting on "educational progress," the National Assessment has, until recently, only considered current student performance compared to student achievement in previous years. Under this approach, the only standard was how well students had done previously, not how well they should be doing on what is measured by the National Assessment.

Comment [AU27]: At the time of the adoption of the policy in 1996, the trend data for "main NAEP" were only a few years old; today we are about to begin the third decade of trend data for 4th and 8th grade reading and math.

The current long-term trend NAEP in reading and math goes back to the early 1970's.

Is the time right to begin the "carefully planned transition" for main NAEP?

Comment [AU28]: This section, assuming it should be retained, should be moved under objective 2.

Although this approach has been useful, it began to change in 1988 from a sole focus on "where we have been" to include "where we want to be" as well.

In 1988, Congress created a non-partisan citizen's group—the National Assessment Governing Board—and authorized it to set explicit performance standards, called achievement levels, for reporting National Assessment results.

The achievement levels describe "how good is good enough" on the various tests that make up the National Assessment. Previously, it might have been reported that the average mathematics score of 4th graders went up (or down) four points on a five-hundred-point scale. There was no way of knowing whether the previous score represented strong or weak performance and whether the amount of change should give cause for concern or celebration. In contrast, the National Assessment now also reports the percentage of students who are performing at or above "basic," "proficient," and "advanced" levels of achievement. Proficient, the central level, represents "competency over challenging subject matter," as demonstrated by how well students perform on the questions on each National Assessment test. Basic denotes partial mastery and advanced signifies superior performance on the National Assessment. Using achievement levels to report results and track changes allows readers to make judgments about whether performance is adequate, whether "progress" is sufficient, and how the National Assessment standards and results compare to those of other tests, such as state and local tests.

First employed in 1990, the achievement levels have been the subject of several independent evaluations and some controversy. Information from these evaluations, as well as from other experts, has been used over the last six years to improve and refine the procedures by which achievement levels are set. Although the current procedures may be among the most comprehensive and sophisticated standard-setting procedures used in education, the Governing Board remains committed to improving the process and to the continuing conduct of validity studies.

- **The National Assessment shall continue to report student achievement results based on performance standards.**

Comment [AU29]: What should we say about achievement levels and NAEP as an indicator of 12th grade academic preparedness?

Use international comparisons

Comment [AU30]: Should there be a stronger emphasis on global competitiveness as the impetus?

Looking at student performance and curriculum expectations in other nations is yet another way to consider the adequacy of U.S. student performance. The National Assessment is, and should be, a domestic assessment. However, decisions on the content of National Assessment tests, the achievement standards, and the interpretation of test results, where feasible, should be informed, in part, by the expectations for education set by other countries, such as Japan, Germany, and England. Although there are technical hurdles to overcome, consideration of such qualitative information can be used to good effect. In addition, the National Assessment should promote "linking" studies with international assessments, as has been done with the Third International Mathematics and Science Study, so that states that participate in the National Assessment can have state, national, and international comparisons. This, in turn, should take into account problems in making

international comparisons truly comparable, such as differences in the samples of students tested, differences in the curricula, and differences in the translated test questions.

- **National Assessment test frameworks, test specifications, achievement levels, and data interpretations shall take into account, where feasible, curricula, standards, and student performance in other nations.**
- **The National Assessment shall promote "linking" studies with international assessments.**

Emphasize reporting for grades 4, 8, and 12

An aspect of the National Assessment design that needs reconsideration is age versus grade-based reporting. At its inception, the National Assessment tested only by age. Current law requires testing both by age (ages 9, 13, and 17) and by grade (grades 4, 8, and 12). Grade-based results are generally more useful than age-based results. Schools and curricula are organized by grade, not by age. Grades 4, 8, and 12 mark key transition points in American education. Grade 12 performance is particularly important as an "exit" measure from the K-12 education system. Grades 4, 8, and 12 are specified for monitoring in National Education Goal 3. Age-based samples may be more appropriate with respect to international comparisons and, given high school dropout rates, would be more inclusive for age 17 than for grade 12 samples, which are limited to youth enrolled in school. However, assessing the knowledge and skills of out-of-school youth may properly fall under the purpose of another program, such as the National Adult Literacy Survey.

Although grade-based reporting is generally preferable, there is a problem about the accuracy of grade 12 National Assessment results. At grade 12, a smaller percentage of schools and students that are invited actually participate in testing than is the case with 4th and 8th graders. Also, more 12th graders fail to complete their tests than do 4th and 8th graders. In addition, when asked, "How hard did you try on this test?" and "How important is doing well on this test?" many more 12th graders than 4th or 8th graders say that they didn't try hard and that the test wasn't important. Low participation rates, low completion rates, and indicators of low motivation suggest that the National Assessment may be underestimating what 12th graders know and can do.

One possible reason for low response and low motivation is that schools and students receive very little in return for their participation in the National Assessment beyond the knowledge that they are performing a public service. They do not receive test scores nor do they receive other information from the National Assessment that teachers and principals might wish to use as a part of the instructional program. This should be changed. The National Assessment design should use meaningful, practical incentives that will give school principals and teachers a greater reason to participate and students more of a reason to try harder. The underlying idea is clear: if principals and teachers see direct benefits, they are more likely to agree to participate in the National Assessment. Students may be more likely to take the assessment seriously if they see that their teachers and principals are enthusiastic about participating. Without practical incentives, even at grades 4 and 8, the willingness of district and school administrators and staff to participate in the National Assessment may diminish over time.

Comment [AU31]: Are these still the right grades?

Should these be augmented?

Is the argument for grade versus age-based testing still relevant?

Much has been done about 12th grade participation and motivation over the years since this report—should this be incorporated?

- **The National Assessment shall continue to test in and report results for grades 4, 8, and 12; however, in selected subjects, one or more of these grades may not be tested.**
- **Age-based testing and reporting shall be permitted when deemed appropriate and when necessary for international comparisons and for long-term trends, should the National Assessment Governing Board decide to continue long-term trends in their current form.**
- **Grade 12 results shall be accompanied by clear, highlighted statements about school and student participation, student motivation, and cautions, where appropriate, about interpreting 12th grade achievement results.**
- **The National Assessment design shall seek to improve school and student participation rates and student motivation at grade 12.**
- **The National Assessment shall provide practical incentives for school and district participation at grades 4, 8, and 12.**

Use innovations in measurement and reporting

The National Assessment has a record of innovations in large-scale testing. These include the early use of performance items, sampling both students and test questions, using standards describing what students should know and be able to do, and employing computers for such things as inventory control, scoring, data analysis, and reporting. The National Assessment should continue to incorporate promising innovative approaches to test administration and improved methods for measuring and reporting student achievement.

Technology can help improve National Assessment reporting and testing. For example, reports could be put on computer disc, transmitted electronically, and made available on the World Wide Web. Test questions could be catalogued and made available on-line for use by state assessment personnel and classroom teachers. Also, the National Assessment could be administered by computer, eliminating the need for costly test booklet systems and reducing steps related to data entry of student responses. Students could answer "performance items" in cost-effective, computerized formats. The increasing use of computers in schools may make it feasible to administer some parts of the National Assessment by computer under the next contract for the National Assessment, beginning around the year 2000.

Other examples of promising methods for measuring and reporting student achievement include adaptive testing and domain-score reporting. In adaptive testing, each student is given a short "pre-test" to estimate that student's level of achievement. Students are then administered test exercises that are in the range of difficulty indicated by the pre-test. Since the test is "adapted" to the individual, it is more precise and can be markedly more efficient than regular test administration. In domain-score reporting, a subject (or "domain") is well defined, a goodly number of test questions are developed that encompass the subject, and student results are reported as a percentage of the "domain" that students "know and can do." This is in contrast to reporting results using an arbitrary scale, such as the 0-500 scale used in the National Assessment.

Comment [AU32]: Inter-active computer-based testing and hands-on science tasks, as well as the advances evident in the tasks for the technology and engineering literacy assessment are but the most recent examples of NAEP embracing cutting-edge technology.

Should this section be updated accordingly?

- The National Assessment shall assess the merits of advances related to technology and the measurement and reporting of student achievement.
- Where warranted, the National Assessment shall implement such advances in order to reduce costs and/or improve test administration, measurement, and reporting.
- The next competition for National Assessment contracts, for assessments beginning around the year 2000, shall ask bidders to provide a plan for
 - (1) conducting testing by computer in at least one subject at one grade, and
 - (2) making use of technology to improve test administration, measurement, and reporting.

OBJECTIVE 2: To develop, through a broadly inclusive process, sound assessments to measure what students know and can do as well as what students should know and be able to do.

Keep test frameworks and specifications stable

Comment [AU33]: Is this section about right?

Test frameworks spell out in general terms how an assessment will be put together. The frameworks also determine what will be reported and influence how expensive an assessment will be. Should 8th grade mathematics include algebra questions? Should there be both multiple-choice questions and questions in which students show their work? What is the best mix of such types of questions for each grade? Which grades are appropriate for assessment in a subject area? Test specifications provide detailed instructions to the test writers about the specific content to be tested at each grade, how test questions will be scored, and the format for each test question (e.g. multiple choice, essay, etc.).

Since 1989, the National Assessment Governing Board has been responsible for developing test frameworks and specifications for NAEP. The Governing Board has done this through a broadly inclusive process, involving hundreds of teachers, curriculum experts, directors of state and local testing programs, administrators, policymakers, practitioners in the content area (e.g., chemists for science, demographers for geography, etc.) and members of the public. This process helps determine what is important for the National Assessment to test, how it should be measured, and how much of what is measured by the National Assessment students should know and be able to do in each subject.

The process of developing frameworks and specifications involves consideration of both current classroom teaching practices and important developments in each subject area for inclusion in the National Assessment. In order to ensure that National Assessment data fairly represent student achievement, the test frameworks and specifications are subjected to wide public review before adoption and test questions developed for the National Assessment are reviewed for relevance and quality by representatives from participating states.

An important role of the National Assessment is to report on trends in student achievement over time. For the National Assessment to be able to measure trends, the frameworks (and hence the tests) must remain stable. However, as new knowledge is gained in subject areas and as teaching practices change and evolve, pressures arise to change the test frameworks and tests to keep them current. But, if frameworks, specifications, and tests change too frequently, trends may be lost, costs go up, and reporting time may increase.

- **Test frameworks and test specifications developed for the National Assessment generally shall remain stable for at least ten years.**
- **To ensure that trend results can be reported, the pool of test questions developed in each subject for the National Assessment shall provide a stable measure of student performance for at least ten years.**
- **In rare circumstances, such as where significant changes in curricula have occurred, the National Assessment Governing Board may consider making changes to test frameworks and specifications before ten years have elapsed.**
- **In developing new test frameworks and specifications, or in making major alterations to approved frameworks and specifications, the cost of the resulting assessment shall be estimated. The National Assessment Governing Board will consider the effect of that cost on the ability to test other subjects before approving a proposed test framework and/or specifications.**

Use an appropriate mix of multiple-choice and "performance" questions

Comment [AU34]: Should this section be retained?

To provide information about "what students know and can do," the National Assessment uses both multiple-choice questions and questions in which students are asked to produce their own answers, such as writing a response to an essay question or explaining how they solved a math problem. Questions of the latter type are sometimes called "performance items." Both types of questions can vary in difficulty and the richness of information they provide, and may require students to demonstrate different kinds of skills and knowledge.

Performance items are desired because they provide direct evidence of what students can do. They range in length of test taking time from a short-answer or fill-in-the-blank format requiring about a minute of response time, to items requiring about 5 minutes of response time, to writing exercises that may allow 15 to 50 minutes response time. Although they may be desirable, performance items are more expensive than multiple-choice to develop, administer, and score. In addition, much larger proportions of students fail to respond to performance items, particularly as the amount of required response time increases.

Multiple-choice questions can be challenging and are desired because they are efficient in collecting information about student knowledge. However, multiple-choice questions are more subject to guessing than are performance items.

Currently, all students tested by the National Assessment are given both types of questions. Generally, about half the testing time is devoted to each type of question, but the amount of time for each differs based on the skills and knowledge to be assessed, as established in the National Assessment test frameworks. For example, in a writing assessment, all students are asked to write their responses to specific exercises. In other subjects, the mix of multiple-choice and performance items varies. The appropriate mix of items for each subject should be determined by the nature of the subject, the range of skills to be assessed, and cost.

- **Both multiple-choice and performance items shall continue to be used in the National Assessment;**
- **In developing new test frameworks, specifications, and questions, decisions about the appropriate mix of multiple-choice and performance items shall take into account the nature of the subject, the range of skills to be assessed, and cost.**

OBJECTIVE 3: To help states and others link their assessments with the National Assessment and use National Assessment data to improve education performance.

The primary job of the National Assessment is to report frequently and promptly to the American public on student achievement. The resources of the National Assessment must be focused on this central purpose if it is to be achieved. However, the products of the National Assessment—test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data—are widely regarded as being of high quality. They are developed with public funds and, therefore, should be available for public use as long as such uses do not threaten the integrity of the National Assessment or its ability to report regularly on student achievement.

The National Assessment should be designed in a way that permits its use by others, while protecting the privacy of students, teachers, and principals who have participated in the National Assessment. This should include making National Assessment test questions and data easy to access and use, and providing related technical assistance upon request. Generally, the costs of a project should be borne by the individual or group making the proposal, not by the National Assessment.

Examples of areas in which particular interest has been expressed for using the National Assessment include linking state and local tests with the National Assessment and performing in-depth analysis on National Assessment data. States that link their tests to the National Assessment would have an unbiased external benchmark to help make judgments about their own tests and standards and also would have a means for comparing their tests and standards with those of other states.

Comment [AU35]: With CCSSI, linking state tests with NAEP may not be relevant—

Should this be changed to:

To help states and others use National Assessment data and resources to improve student achievement and close achievement gaps?

The National Assessment shall develop policies, practices, and procedures that assist states, school districts, and others who want to do so at their own cost to link their test results to the National Assessment.

- **The National Assessment shall be designed so that others may access and use National Assessment test frameworks, specifications, scoring guides, results, questions, achievement levels, and background data.**
- **The National Assessment shall employ safeguards to protect the integrity of the National Assessment program, prevent misuse of data, and ensure the privacy of individual test takers.**

Setting Priorities in an Uncertain Budget Environment: The NAEP Schedule of Assessments

Background

One of the Governing Board's congressionally mandated responsibilities is to select the subject areas to be assessed by NAEP. The Governing Board's practice is to develop a long-range schedule with at least a 10-year outlook. The Executive Committee is responsible for developing proposals for the NAEP schedule of assessments for Board consideration.

The purpose of having a long-range schedule of assessments is to aid states in planning for participation in NAEP, and for planning NAEP's budget requests, operations, and contracts. The current approved NAEP schedule (adopted in December 2011), which covers 2005-2017, appears on the fourth page of this attachment.

On the last page of this attachment is a staff proposal, developed in consultation with NCES, for the NAEP schedule of assessments through 2022. This proposal was presented previously to the Executive Committee, as a vehicle for discussing priorities for assessments, taking a very long-range view.

The proposal also had the purpose of providing NCES with necessary guidance for developing the scope of work for the NAEP contract competitions for 2013-2017 (n.b., proposals have been received and are being reviewed), which will include some test development and field testing for assessments in 2018 and beyond.

The NAEP schedule is the primary driver of the NAEP budget. While the NAEP annual appropriation over the last decade has been generally stable with intermittent increases, future NAEP funding levels will be affected by actions taken by the Administration and Congress to address the fiscal strains on the federal budget generally. The FY 2013 appropriation has not been passed—the Senate mark for NAEP is a reduction of \$5 million while the House mark is level funding at the FY 2012 level of \$129.1 million. The Continuing Resolution through March 27, 2013 keeps NAEP essentially at the FY 2012 level. If sequestration takes effect on March 1, 2013, there could be a reduction of about 5-8%, or about \$7 to \$10 million, not just in 2013, but annually.

Given the prospect of reduced funding, it is prudent for Board discussion to begin on what it values and the trade-offs regarding subjects and grades to assess and their frequency.

To help prompt discussion, the two organizing principles for the staff proposal for assessments through 2022 are provided below, followed by a series of questions to sort out values and trade-offs.

Staff Proposal “Principles”

The first organizing principle for years 2018-2022 is to continue current policy for the schedule:

- reading and mathematics (national and state) conducted biennially in odd-numbered years
- science and writing (national and state) once every four years in alternating odd-numbered years
- high school transcript study once every four years in the same year as mathematics and science
- U.S. history, civics and geography (national) once every four years in even-numbered years
- long-term trend reading and mathematics assessments once every four years in even-numbered years
- other subjects—arts, economics, foreign language, world history—in even-numbered years as time and resources permit

A second principle for the schedule is that NAEP will evolve incrementally to fully computer-based administration by 2022.

This principle assumes that NAEP administrations can be conducted using school-based informational technology (IT). It also assumes that state IT systems will be enhanced during the coming years to accommodate the Common Core State Standards assessments as a consequence of the Common Core Assessment Consortia initiatives.

Other additions to the schedule include:

- economics at grade 12 in 2016
- Technology and Engineering Literacy is scheduled once every four years starting in 2014 at grade 8, and expanding in 2018 to grades 8 and 12, and 2022 at grades 4, 8, and 12
- the Trial Urban District Assessments are not displayed on the schedule. Governing Board policy is for this trial to continue with sufficient resources to support at least 21 districts—the number participating in 2011. In general, assessments would continue in the same years and subjects as state-level assessments.
- as the schedule indicates, Governing Board policy is to continue to offer state level assessments at grade 12 with participation on a voluntary basis, and with sufficient resources to support at least 13 states—the number participating in 2013.

Some questions to consider:

- Should 4th and 8th grade reading and mathematics assessments at the state level continue every two years?
- Should state level writing and science continue once every four years?
- Should the frequency of some state-level subjects be reduced so that other state-level subjects can be added?
- Should 12th grade state assessments be given more/less frequently?
- Should the number of 12th grade states be expanded?
- Should the frequency of TUDA be changed?
- Should the frequency of Long-term Trend assessments in reading and mathematics be changed? Should Long-term trend assessments be discontinued? Should Long-term Trend assessments be incorporated into Main NAEP?
- Are there subjects that should be added/dropped or be administered more/less frequently?
- Should the frequency of the High School Transcript Study be changed?

NAEP Schedule of Assessments – Approved December 3, 2011		
Year	National	State
2005	Reading MATHEMATICS Science High School Transcript Study	Reading (4, 8) MATH (4, 8) Science (4, 8)
2006	U.S. History Civics ECONOMICS (12)	
2007	Reading (4, 8) Mathematics (4, 8) Writing (8, 12)	Reading (4, 8) Math (4, 8) Writing (8)
2008	Arts (8) Long-term trend	
2009	READING Mathematics* SCIENCE** High School Transcript Study	READING (4, 8, 12) Math (4, 8, 12) SCIENCE (4, 8)
2010	U.S. History Civics Geography	
2011	Reading (4, 8) Mathematics (4, 8) Science (8)** WRITING (8, 12)**	Reading (4, 8) Math (4, 8) Science (8)
2012	Economics (12) Long-term trend	
2013	Reading Mathematics	Reading (4, 8, 12) Math (4, 8, 12)
2014	U.S. History Civics Geography TECHNOLOGY AND ENGINEERING LITERACY (8) **	
2015	Reading Mathematics Science** High School Transcript Study	Reading (4, 8, 12) Math (4, 8, 12) Science (4, 8, 12)
2016	Arts (8) Long-term trend	
2017	Reading Mathematics Writing**	Reading (4, 8, 12) Math (4, 8, 12) Writing (4, 8, 12)

*New framework for grade 12 only.

**Assessments involving test administration by computer.

NOTES:

(1) Grades tested are 4, 8, and 12 unless otherwise indicated, except that long-term trend assessments sample students at ages 9, 13, and 17 and are conducted in reading and mathematics.

(2) Subjects in **BOLD ALL CAPS** indicate the year in which a new framework is implemented or assessment year for which the Board will decide whether a new or updated framework is needed.

(3) In 2009, 12th grade assessments in reading and mathematics at the state level were conducted as a pilot in 11 volunteering states (AR, CT, FL, IA, ID, IL, MA, NH, NJ, SD, WV). For 2013, 13 states agreed to participate (with MI and TN added).

(4) The Governing Board intends to conduct assessments at the 12th grade in World History and Foreign Language during the assessment period 2018-2022.

NAEP Schedule of Assessments – DISCUSSION DRAFT		
Year	National	State
2010	U.S. History Civics Geography	
2011	Reading (4, 8) Mathematics (4, 8) Science (8)** WRITING (8, 12)**	Reading (4, 8) Math (4, 8) Science (8)
2012	Economics (12) Long-term trend	
2013	Reading Mathematics	Reading (4, 8, 12) Math (4, 8, 12)
2014	U.S. History Civics Geography TECHNOLOGY AND ENGINEERING LITERACY (8) **	
2015	Reading Mathematics Science** High School Transcript Study	Reading (4, 8, 12) Math (4, 8, 12) Science (4, 8, 12)
2016	Arts (8) Add: Economics (12) Long-term trend	
2017	Reading Mathematics MATHEMATICS** (nat'l and state) Writing**	Reading (4, 8, 12) Math (4, 8, 12)** Writing (4, 8, 12)**
2018	U.S. History Civics CIVICS** Geography Technology and Engineering Literacy (8, 12) **	
2019	Reading READING ** (nat'l and state) Mathematics** Science** High School Transcript Study	Reading (4, 8, 12)** Math (4, 8, 12)** Science (4, 8, 12)**
2020	Long-term trend NOTE: administer by computer? Economics (12)** FOREIGN LANGUAGE (12) **	
2021	Reading ** NOTE: PIRLS is expected this year Mathematics** Writing**	Reading (4, 8, 12)** Math (4, 8, 12)** Writing (4, 8, 12)**
2022	U.S. HISTORY** Civics** GEOGRAPHY** WORLD HISTORY (12) ** Technology And Engineering Literacy (4, 8, 12) **	

**Assessments involving test administration by computer.

NOTES:

- (1) Grades tested are 4, 8, and 12 unless otherwise indicated, except that long-term trend assessments sample students at ages 9, 13, and 17 and are conducted in reading and mathematics.
- (2) Subjects in **BOLD ALL CAPS** indicate the year in which a new framework is implemented or assessment year for which the Board will decide whether a new or updated framework is needed.
- (3) In 2009, 12th grade assessments in reading and mathematics at the state level will be conducted as a pilot in 11 volunteering states (AR, CT, FL, IA, ID, IL, MA, NH, NJ, SD, WV). For 2013, 13 states agreed to participate (with MI and TN added).

(4) The Governing Board intends to conduct assessments at the 12th grade in World History and Foreign Language during the assessment period 2018-2022.

Action Item: Delegation of Authority

Draft Resolution

Whereas, the FY 2013 funding for the National Assessment of Educational Progress (NAEP) and the National Assessment Governing Board is under a continuing resolution through March 27, 2013; and

Whereas, absent action by Congress, the Budget Control Act of 2011 would, as of March 1, 2013, result in cuts estimated at approximately 7 to 10 percent of the NAEP and Governing Board appropriations in FY 2013; and

Whereas, the National Center for Education Statistics (NCES) is reviewing contract bids for NAEP operations for 2013-2017 that are scheduled to be awarded by the end of March 2013; and

Whereas, the status of the FY 2013 appropriation and contract awards may have an impact on the NAEP schedule of assessments for 2014 and beyond that would require action by the Governing Board prior to the May 2013 Governing Board meeting;

Therefore, the National Assessment Governing Board hereby approves a delegation of authority to the Executive Committee to act on behalf of the Governing Board, prior to the May 2013 Board meeting, to make decisions on changes to the NAEP schedule of assessments in 2014 and beyond, as necessary, in consultation with NCES.