

## **General Policy: Conducting and Reporting The National Assessment of Educational Progress**

### **Foreword**

*This policy is a guide for those responsible for the National Assessment of Educational Progress (NAEP)—the Nation’s Report Card. These are the members and staff of the National Assessment Governing Board (NAGB) that oversees NAEP; the Commissioner and staff of the National Center for Education Statistics (NCES) responsible for NAEP operations; and the staff of the contractors that carry out NAEP.*

*NAEP performs an exceptional public service. It provides trusted information on the performance and progress of the nation’s elementary and secondary schools and school children. Over the course of its history, a set of essential, enduring principles and values have become embodied in NAEP. These principles and values are set forth below.*

### **Introduction**

Thomas Jefferson said “If a nation expects to be ignorant and free in a state of civilization, it expects what never was and never will be.” Horace Mann, the advocate for the Common School, said “Education...beyond all other devices of human origin, is the great equalizer of the conditions of men—the balance-wheel of the social machinery.” John F. Kennedy, paraphrasing H.G. Wells, said “...the course of civilization is a race between catastrophe and education. In a democracy such as ours, we must make sure that education wins the race.”

The nation’s leaders have long recognized education as a foundation for democracy. Education fosters capable civic participation; supports individual human development; promotes national, state, and individual economic well-being; and advances national security. Providing for the education of its citizens and monitoring their levels of achievement are key functions of states and the nation. NAEP was established for the latter function—to monitor student achievement.

### **History and Evolution of the National Assessment of Educational Progress**

The first U.S. Department of Education was created by Congress in 1867. It was the early predecessor of NCES, established to “[collect] such statistics...as shall show the condition and progress of education in the several States and Territories, and [diffuse] such information...as shall aid the people of the United States in the establishment and maintenance of efficient school systems...” For more than 100 years, this Department and its successors provided information on the number of schools, school districts, student enrollment, revenues, expenditures, and the like, but collected no information on student achievement.

This began to change in 1963. U.S. Commissioner of Education Francis Keppel was testifying before the House Appropriations Committee on the FY 1964 budget for education. A committee member asked Keppel a simple question—“How well are U.S. students achieving?” Keppel was not able to answer the question because there was no source of information to answer it.

The question—important at any time—was raised in the context of the Cold War and concern about national security. What was then the Soviet Union had launched Sputnik by missile on October 4, 1957. This feat caused the nation’s leaders to fear that the United States lacked sufficient scientific and engineering capability to compete and keep the country safe.

Kepple recognized the threat inherent in failing to know the levels of U.S. student achievement. It set him on a path that led to the creation of the National Assessment of Educational Progress and the conduct of the first assessment—in science—in 1969. Kepple began by forming a committee to design a national assessment. The committee was established late in 1963. It was funded in large part by the Carnegie Corporation and led by Ralph Tyler, the preeminent education researcher of his day.

Some leaders in school administration, curriculum, and the teaching force opposed the idea of a federal assessment of student achievement. They were concerned that a federal test would lead to federal intrusion in school curriculum and accountability, responsibilities of state and local education officials. This is a recurring theme in the evolution of NAEP: finding the right way to serve the national interest as a monitor of student achievement while honoring state and local authority over schools.

The proposed design addressed the opponents’ concerns. The Education Commission of the States (ECS) would carry out the assessment with funding from the U.S. Office of Education. This put authority in a state-based organization and placed the federal role at arms-length from the assessment. Decisions about content and subjects to test would be made by ECS. There would be no student, school, district, or state-level results. Data would be reported for the nation and for regions of the country. Student samples would be age-based rather than grade-based. Together, these addressed concerns that the National Assessment would lead to a national curriculum and federal entanglement in school governance.

The education landscape has changed since the initial assessment in 1969. Accordingly, the National Assessment has evolved. Where there was some opposition in the beginning, NAEP has earned trust, is recognized for its quality, and is highly valued. Little known in the early years except by interested researchers, NAEP results have become widely used by education leaders and are featured by the news media. The original design was an innovation responsive to the times. Since then, many responsive innovations have been made in NAEP’s governance, the subjects assessed, item types, test procedures, and the use of information and communication technology. State level and grade-based reporting are now a regular part of NAEP.

Change in the education environment continues. Change and innovation in response to the needs of the time are hallmarks of NAEP. These are balanced against the imperative to maintain NAEP’s independence as a stable measure for reporting achievement trends. Balancing competing goals is a continual challenge to NAEP, a tension that is the source of its continual creative evolution to better serve the American public.

## Policy Statement

### **Purpose and Characteristics of the National Assessment of Educational Progress**

The National Assessment of Educational Progress is as an independent monitor of student academic achievement in the United States at the elementary and secondary levels. It reports on achievement at specific points in time and trends in achievement over time.

Congressionally authorized and funded, NAEP is uniquely positioned to serve as an independent monitor of student achievement. As the Nation's Report Card, NAEP is uniquely obliged to maintain the public trust. This is achieved through a governance structure and assessment procedures that are transparent, involve stakeholders, and are subject to scrutiny by technical experts, policymakers, and the public. These mechanisms ensure the accuracy, timeliness, integrity and credibility of NAEP results. They provide for the validity of inferences made about the results. They keep NAEP free of ideology, inappropriate influences and special interests.

Each NAEP assessment is a complex project, with a five-to-six-year life cycle for new assessments. This includes about 18 months for developing a new framework, about one year for test development, one year for pilot testing, one year for the conduct of the assessment and scoring and analysis, and one year for achievement-level setting and reporting. Each step is conducted in a thoughtful, deliberate manner with input from hundreds of stakeholders and experts, requiring careful coordination among NAGB, NCES, and the many NAEP contractors and participants.

NAEP is a representative sample survey, using statistically sound means for drawing its samples. NAEP results are presented in a manner that assures fairness in comparisons of achievement and trends over time for all subgroups reported; for geographic units, such as the nation, states, and school districts; and for public and private schools.

NAEP covers a wide range of important subjects or topics. This includes reading, mathematics, science, writing, U.S. history, civics, geography, economics, foreign language, the arts, and technology and engineering literacy. NAEP uses matrix sampling to ensure breadth and depth of subject coverage while minimizing testing time for students.

Assessments are conducted at grades 4, 8, and 12. The 4<sup>th</sup> grade was selected as the point at which the foundations for further learning are expected to be in place (e.g., when "learning to read" becomes "reading to learn"). The 8<sup>th</sup> grade was selected because it is the typical transition point to high school. The 12<sup>th</sup> grade was selected because it is the end of the K-12 experience, the transition point for most students to postsecondary education, training, the military, and other adult pursuits. NAEP is unique as the only source of 12<sup>th</sup> grade results at the national and state levels. Assessments are also administered at ages 9, 13, and 17, in connection with the reading and mathematics assessments conducted at NAEP's beginning (referred to as the long-term trend assessments), and when appropriate for comparisons with international assessments.

NAEP reports results by gender, race/ethnicity, and income level, for students with disabilities and for students who are English language learners. NAEP was a pioneer in reporting data on education achievement disaggregated by student demographic subgroups. The Nation's Report Card brings public attention to gaps in achievement between subgroups, where they exist, and to trends over time in the size of these gaps.

**Limitations: What the National Assessment of Educational Progress Is Not**

NAEP only provides group results; it does not produce results for individual students. NAEP is intended to describe how well students are performing, but not to explain why. While NAEP collects background information on student demographics and other characteristics, it does not collect information that is intrusive to individual students or families.

NAEP's background information includes factors that may affect student achievement, such as educational policy, instructional activities and teacher preparation. However, data from the background information NAEP collects do not, by themselves, support conclusions about the effectiveness of these factors on student outcomes nor about ways to improve education practice.

Each NAEP assessment is developed through a national consensus process. This process takes into account education practices, the results of education research, and changes in curricula. However, NAEP is independent of any particular curriculum and does not promote specific ideas, ideologies, or teaching techniques.

**The Audiences for the National Assessment of Educational Progress**

The primary audience for NAEP results is the American public and their congressional representatives, including especially those in states and districts that receive their own NAEP results. With this audience as the target, NAEP reports are written to be understandable, free of jargon, easy to use, and easy to access. Assessment questions and samples of student work are included in NAEP reports when appropriate to illustrate the meaning of NAEP achievement levels and NAEP scores. Although written for a lay audience, NAEP reports do not trade accuracy for simplicity.

Another audience is made up of those who use NAEP data—the national and state policymakers and educators concerned with student achievement, curricula, testing, and standards. NAEP data and related information (e.g., assessment frameworks and items) are available to these users in forms designed to support their efforts to interpret NAEP results to the public, to improve education performance, and to perform secondary analysis.

To be relevant and useful to these audiences, NAEP results must be timely. Therefore, NAGB has set the goal of releasing NAEP results within six months of testing.

**Objectives and Activities for Conducting the National Assessment of Educational Progress**

**Objective 1: To serve as a consistent external, independent measure of student achievement by which results across education systems can be compared at points in time and over time.**

National, state, and local education leaders and policymakers—public and private—rely on NAEP data as an independent monitor of student achievement and as a way to compare performance across education systems. For NAEP to serve in this role, NAGB, in consultation with NCES and stakeholders, periodically establishes a dependable, publicly announced assessment schedule of at least ten years in scope. The schedule specifies the subject or topic

(e.g., High School Transcript Study), grades, ages, assessment year, and sampling levels (e.g., national, state) for each assessment.

The NAEP schedule of assessments is the foundation for states' planning for participation in the assessments. It is the basis for NCES operational planning, annual budget requests, and contract statements of work. In making decisions about the NAEP schedule of assessments, NAGB includes the wide range of important subjects and topics to which students are exposed. NAGB also considers opportunities to conduct studies linking NAEP with international assessments.

As the NAEP authorizing legislation provides, assessments are conducted in reading and mathematics, and, as time and resources allow, in subjects such as science, writing, history, civics, geography, the arts, foreign language, economics, technology and engineering literacy and other areas, as determined by NAGB. The goal for the frequency of each subject area assessment is at least twice in ten years, to provide for reporting achievement trends.

In order to compare results across geographic jurisdictions, the samples drawn must be representative. For each assessment, the National Assessment program takes affirmative steps to achieve statistically sound levels of school and student participation and optimal levels of student engagement in the assessment, including steps to maximize the participation of students with disabilities and students who are English language learners.

NCES employs safeguards to protect the integrity of the National Assessment program, prevent misuse of data, and ensure the privacy of individual test takers. NAEP results are accompanied by clear statements about school and student participation rates; student engagement in the assessment, when feasible; and cautions, where appropriate, about interpreting achievement results.

**OBJECTIVE 2: To develop technically sound, relevant assessments designed to measure what students know and can do.**

NAEP assessment frameworks spell out how each subject area assessment will be put together. The frameworks are the foundation for what NAEP will assess and report. Assessment frameworks describe the knowledge and skills most important for NAEP to assess at each grade. They provide for the item types and appropriate mix that best represent such knowledge and skills (e.g., multiple-choice, constructed response, hands-on task, information and communication technology-based task or simulation, etc.). Test specifications provide detailed instructions to the test writers about the specific content to be tested at each grade, the item type for each test question, and how items will be scored.

The National Assessment Governing Board is responsible for developing assessment frameworks and specifications for NAEP. NAGB does this through a comprehensive, broadly inclusive process lasting about 18 months. It involves hundreds of teachers, curriculum experts, state and local testing officials, administrators, policymakers, practitioners in the content area (e.g., chemists for science, demographers for geography, etc.) and members of the public.

The framework development process helps determine what is important for NAEP to assess and how it should be measured. The frameworks also include preliminary achievement level descriptions (see Objective 3). The framework development process considers both current classroom teaching practices and important advances in each subject area. Where applicable, the curricula, performance standards, and student achievement in other nations are also considered.

NCES is responsible for developing items for each assessment that comprehensively measure the subject domain as defined by the assessment framework and specifications. NAGB is responsible for approving all items, including those for background information, before use in an assessment.

NCES regularly evaluates the extent to which the set of items for each assessment meets the framework requirements, assessment specifications, and achievement level descriptions.

To ensure that NAEP data fairly represent what students know and can do, the frameworks and specifications are subjected to wide public review before adoption, and the items developed are reviewed for relevance and quality by representatives from participating states.

For NAEP to measure trends in achievement accurately, the frameworks (and hence the assessments) must remain sufficiently stable. However, as new knowledge is gained in subject areas, the information and communication technology for testing advances, and curricula and teaching practices evolve, it is appropriate for NAGB to consider changing the assessment frameworks and items to ensure that they support valid inferences about student achievement. But if frameworks, specifications, and items change too abruptly or frequently, the ability to continue trend lines may be lost prematurely, costs go up, and reporting time may increase.

For these reasons, NAGB generally maintains the stability of NAEP assessment frameworks and specifications for at least ten years. NCES assures that the pool of items developed for each subject provides a stable measure of achievement for at least the same ten year period. In deciding to develop new assessment frameworks and specifications, or to make major alterations to approved frameworks and specifications, NAGB considers the impact on reporting trends. Whenever feasible, technically defensible steps are taken to avoid breaking trend lines. In rare circumstances, such as where significant changes in curricula have occurred, NAGB may consider making changes to assessment frameworks and specifications before ten years have elapsed.

In developing new assessment frameworks and specifications, or in making major alterations to approved frameworks and specifications, NAGB, in consultation with NCES, estimates the cost of the resulting assessment. NAGB considers the effect of that cost on the overall priorities for the NAEP schedule of assessments.

### **OBJECTIVE 3. To continue to set and report achievement levels for NAEP results.**

In the 1988 re-authorization of NAEP, Congress made three major innovations. It provided for the first ever state-level assessments, created NAGB to oversee and set policy for NAEP, and

authorized NAGB to set explicit performance standards, called achievement levels, for reporting NAEP results.

Previously, NAEP reporting focused primarily on average scores and whether they had changed since prior assessments. The average mathematics score of 4th graders may have gone up (or down) four points on a five-hundred-point scale. But there was no way of knowing whether the current and previous scores represented strong or weak performance and whether the amount of change should give cause for concern or celebration.

There had been attempts to give meaning to the NAEP scales through what were referred to as “performance levels.” Starting at 250—the midpoint of the 0-500 scale—points were selected for reporting at 50-point intervals above and below. The cluster of skills that differentiated each major level were identified by the items that students were more likely to answer correctly at one level than students at lower levels. Descriptions of what students know and can do at each performance level were developed from the content of the respective item clusters. However, the performance levels still did not address whether achievement was “good enough.”

NAGB approved the first policy statement on the use of achievement levels in May 1990. The policy called for the NAEP achievement levels to be denoted as “Basic,” “Proficient,” and “Advanced.” Proficient, the central level, represents “competency over challenging subject matter,” as demonstrated by how well students perform on NAEP. Basic denotes partial mastery and Advanced signifies superior performance on NAEP. Using achievement levels to report results and track change over time adds meaning to the score scale. Reporting by achievement levels helps readers judge whether performance is adequate and progress over time sufficient.

The NAEP achievement levels are developed through a thorough procedure with comprehensive technical documentation, involving expert judgment. For each achievement level-setting project, an explicit design document is developed. The design document describes the qualifications for the individuals who will serve on the achievement level-setting panels and the specific process that will be conducted, including evaluation procedures and validity research. The panels’ recommendations are subject to technical and public comment. Ultimately, while considering the panels’ recommendations, the achievement levels are set by NAGB.

NAEP achievement levels are widely used by national, state, and local education leaders and policymakers. They contribute to NAEP’s role as an independent external monitor of student achievement. The achievement levels provide a common reference by which state and local performance standards and results can be compared.

The NAEP achievement levels have been the subject of several independent evaluations. NAGB uses information from these evaluations, as well as from other experts, to improve and refine the procedures by which achievement levels are set. Although NAGB’s standard-setting procedures may be among the most comprehensive and sophisticated used in education, NAGB continually improves the achievement level-setting and reporting process.

NAGB conducts continuing research to support the validity of inferences made in relation to NAEP achievement levels. Where the research indicates that there are limitations on the inferences that can be made in relation to NAEP achievement levels, these limitations are included in NAEP reports. Average scores, percentiles, and other relevant statistics are reported along with NAEP achievement levels to provide context and avoid misinterpretations.

**OBJECTIVE 4. To bring attention to achievement gaps where they exist among demographic subgroups and the urgency of closing those achievement gaps.**

Because education is the cornerstone of a nation's strength, the existence of persistent achievement gaps between demographic subgroups in the U.S. is a threat ignored at our peril. The nation's founding documents and Constitution provide for equal opportunity and equal justice under law for all. Supreme Court decisions and federal legislation undergird these civil protections against discrimination, especially in the arena of public education.

For these reasons, NAEP monitors student achievement by gender, race/ethnicity, and income level, and for students with disabilities and who are English language learners. In order to address achievement gaps, it is necessary first to identify them.

NAEP reports highlight achievement gaps among the student demographic subgroups so that the public is made aware and officials with responsibility have information on which to take action. The members of NAGB, individually and collectively, carry out initiatives to convey the urgency of closing achievement gaps to the public. These initiatives include preparation of special NAEP reports focused on achievement gaps, presentations, symposia, and public statements made in connection with the release of NAEP results.

**OBJECTIVE 5. To disseminate timely NAEP reports and to make NAEP data and information useful and easily accessible to various audiences, including educators, policymakers and the public**

Given the importance of NAEP results, their timely release is critical to their impact. The goal is to release NAEP assessment results within six months of the completion of testing.

The information available from the National Assessment program is rich and varied. It includes:

- NAEP reports;
- assessment frameworks and specifications for the broad array of subjects included in NAEP;
- hundreds of released assessment items, including student data, exemplar student responses, and scoring guides;
- assessment results;
- achievement level results and descriptions; and
- background information collected from students, teachers, and school administrators.



This information is available on-line at no charge. Providing electronic versions of these materials makes them easily accessible and minimizes the need for printed copies.

NAGB and NCES continually evaluate audience needs and employ innovations in information and communication technology to improve access, usability, and usefulness of NAEP data and related resources. The aim is to optimize the potential of NAEP information to help states and others improve education achievement and close achievement gaps.

This includes procedures developed by NCES to facilitate the ability of states to link performance on NAEP with data in state longitudinal data bases. It also includes the option for states to use NAEP assessments planned for administration at the national level only. States can do this by assuming the costs and adhering to requirements that protect the integrity of the NAEP program. NAGB and NCES ensure that state decision makers receive timely notice of this option and that the cost to states is minimized.

**OBJECTIVE 6. To continue to innovate in NAEP framework development, item development, data collection, test security, scoring, analysis and reporting.**

Innovation is at the heart of NAEP and has been since its inception. NAEP is recognized for its advances in large-scale assessment administration, item formats, data collection, test security, scoring, analysis, quality assurance, and reporting. NAEP has embraced information and communication technology as subject matter (e.g., in the Technology and Engineering Literacy Assessment), as a tool for conducting assessments (e.g., the Writing Assessment and the Science Assessment interactive computer tasks), and as a channel to disseminate NAEP information (e.g., the on-line data tools). NAEP continually seeks innovations in national and state sampling procedures to reduce burden on schools and students, increase efficiency, and minimize costs. Innovation is built into the NAEP *modus operandi* and this will continue into the future.

Executive Committee Subcommittee on Updating NAEP Policy:  
Lou Fabrizio, Alan Friedman, and Shannon Garrison

Prepared by: Ray Fields, April 24, 2013

### **Interpreting NAEP Proficient Using Preparedness Research Findings**

The background document for this discussion item will be uploaded on the Governing Board Member secure website prior to the May 2013 Governing Board meeting.

The background document will be entitled: “Draft Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College.”