

National Assessment Governing Board Executive Committee

May 16, 2013

AGENDA

| | | |
|---------|--|---|
| 4:30 pm | Welcome and Agenda Overview <i>David Driscoll, Chair</i> | |
| 4:35 pm | <p>Committee Issues and Challenges</p> <p>Ad Hoc Committee on Background Information: Reviewing the NAEP Background Information Framework <i>Terry Holliday, Ad Hoc Committee Chair</i></p> <p>ADC: Revisiting the NAEP Foreign Language Assessment <i>Alan Friedman, ADC Chair</i></p> <p>COSDAM: Setting Achievement Levels for TEL <i>Lou Fabrizio, COSDAM Chair</i></p> <p>Nominations: Expanding Outreach for the 2014 Nominations Cycle <i>Tonya Miles, Nominations Committee Chair</i></p> <p>Reporting & Dissemination: Examining the Policy on Reporting, Release and Dissemination of NAEP Results <i>Tom Luna, R & D Vice Chair</i></p> | |
| 4:55 pm | Draft Policy Statement on the Conduct and Reporting of NAEP <i>Alan Friedman and Shannon Garrison, Executive Committee</i> | Attachment A |
| 5:15 | Interpreting NAEP Proficient Using Preparedness Research Findings <i>Lou Fabrizio</i> | Attachment B (Will be sent separately) |
| 5:25 pm | <p>ACTION ITEM</p> <p>NAEP Schedule of Assessments (Tentative) <i>Ray Fields, Assistant Director for Policy and Research</i></p> | Attachment C |
| | <i>Tentative: CLOSED SESSION 5:30 – 6:30 p.m.</i> | |
| 5:30 pm | NAEP Contracts, Budget, and Schedule for 2013 and Beyond <i>Cornelia Orr, Executive Director</i> <i>Peggy Carr, Associate Commissioner, NCES</i> | |

Adopted: _____

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 4th 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 8th grade was selected because it is the typical transition point to high school. The 12th 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 12th 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 NCEES 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.

NCEES 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

For almost a decade, the Governing Board has been thoughtfully and deliberately working to determine the feasibility of NAEP reporting on the academic preparedness of 12th grade students for college and job training. Accordingly, the Governing Board is conducting a comprehensive program of preparedness research. The first phase of the research involved more than 30 studies in 5 areas: content alignment, statistical relationship, standard-setting, benchmarking, and a survey of higher education.

On the basis of the research results, Governing Board staff propose the following inferences for use in reporting NAEP 12th grade results:

12th grade students scoring at or above the Proficient achievement level on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college.

On the following pages is the draft validity argument in support of these inferences. With the approval of the Committee on Standards, Design and Methodology (COSDAM), the validity argument is based on a model described by Michael Kane, a renowned psychometrician widely recognized for his theoretical work in validity. In summary, the model begins with a score interpretation (i.e., the inferences above), a statement of the propositions or assumptions that underlie that score interpretation, and the presentation of evidence to evaluate those propositions or assumptions. Recognizing that validation is a continuing process, and that validity cannot be established absolutely, Kane's model provides that the criterion that must be met is the **plausibility** of the validity argument. Michael Kane serves as an advisor in the development of the validity argument and has reviewed this draft.

At the May 2013 Governing Board meeting, the draft validity argument will be discussed by the Executive Committee, COSDAM, and the full Board. The purpose is two-fold: to review the Phase I preparedness research and results, and to provide feedback to staff on the draft validity argument in relation to the requirements of the Kane model. Thus, the feedback should address the following questions—are the propositions/assumptions optimally framed, does the evidence appropriately address the propositions/assumptions, are there propositions/assumptions that should be added, and taken as a whole, does the argument seem to meet the plausibility criterion?

Following the May 2013 Board meeting, the draft validity argument will be revised per the feedback provided. The draft then will be subjected to independent external review by technical experts. Two noted psychometricians, Mark Reckase and Gregory Cizek have agreed to review the draft validity argument. The intention is to include their reviews as a part of the final validity argument.

The final validity argument will be presented at the August 2013 meeting for action by the full Board with respect to its use in reporting 12th grade reading and mathematics results.

Finally, please note that the appendices mentioned in the draft validity argument are not included here, but are available upon request.

Draft Validity Argument for NAEP Reporting on 12th Grade Academic Preparedness for College

Ray Fields – May 9, 2013

Introduction

Rationale for NAEP Reporting on 12th Grade Academic Preparedness

The National Assessment Governing Board is conducting a program of research to determine the feasibility of the National Assessment of Educational Progress (NAEP) reporting on the academic preparedness of U.S. 12th grade students, in reading and mathematics, for college and job training.

Since 1969, NAEP has reported to the public on the status and progress of student achievement in a wide range of key subjects at grades 4, 8, and 12. NAEP provides national and state-representative results, results for twenty-one urban districts, and results by subgroups of students (e.g., by race/ethnicity, gender, and for students with disabilities and English language learners). NAEP, by law, does not provide individual student results.

The Governing Board's initiative on 12th grade academic preparedness began in March 2004, with the report of a blue-ribbon panel.¹ The panel was composed of K-12 education leaders—the “producers” of high school graduates—and leaders in business, postsecondary education, and the military—the “consumers” of high school graduates.

Recognizing the importance of 12th grade as the gateway to postsecondary education and training, and viewing NAEP as a “truth teller” about student achievement, this distinguished panel of state and national leaders recommended unanimously that “NAEP should report 12th grade students’ readiness for college-credit coursework, training for employment, and entrance into the military.” (National Commission on NAEP 12th Grade Assessment and Reporting; p. 6.). They stated that “America needs to know how well prepared its high school seniors are...[only NAEP] can provide this information...and it is necessary for our nation’s well-being that it be provided.” (Ibid.; p. 2.).

The Governing Board approved this recommendation, with a minor modification. The term “readiness” was changed to “academic preparedness” and “entrance into the military” was subsumed by “job training.” “Readiness” was changed to “academic preparedness” because “readiness” is broadly understood to include academic preparedness and other characteristics needed for success in postsecondary education and training, such as habits of mind, time management, and persistence (Conley). NAEP does not measure such characteristics. Rather, NAEP is designed to measure academic knowledge and skills. “Entrance into the military” was subsumed by “job training” with the intention of identifying occupations with civilian and military counterparts and utilizing both the military’s experience as the world’s largest occupational training organization and its extensive research on the relationship between performance on the Armed Service Vocational Aptitude Battery (ASVAB) and job training outcomes.

¹ The blue-ribbon panel was known officially as the National Commission on NAEP 12th Grade Assessment and Reporting.

The Governing Board approved the 12th grade academic preparedness initiative because the academic preparation of high school students for postsecondary education and training is important to the nation's economic well-being, national security, and democratic foundations (see Governing Board resolution of May 21, 2005 at <http://www.nagb.org/content/nagb/assets/documents/policies/resolution-on-preparedness.pdf>).

The Governing Board is not alone in recognizing the importance of 12th grade academic preparedness for the nation. Since the acceptance of the blue-ribbon panel report in 2004, the focus on ensuring that 12th grade students graduate “college and career ready” has been widely embraced as a policy goal by state and national leaders. These include the National Governors Association (NGA), the Council of Chief State School Officers (CCSSO), the Business Roundtable (BRT), the U.S. Chamber of Commerce (the Chamber), and the Obama Administration. The impetus for this attention to academic preparedness for college and job training is well summarized by a statement of the Business Coalition for Student Achievement, an organization coordinated by BRT and the Chamber:

“Ensuring that all students graduate academically prepared for college, citizenship and the 21st century workplace...is necessary to provide a strong foundation for both U.S. competitiveness and for individuals to succeed in our rapidly changing world.”

Viewing the need for rigor in education achievement through the lens of national security, a similar conclusion was made in the report of the Independent Task Force on U.S. Education Reform and National Security of the Council on Foreign Relations, co-chaired by former New York City School Chancellor Joel Klein and Former Secretary of State Condoleezza Rice. The NGA and CCSSO collaborated to develop Common Core State Standards (CCSS) for mathematics and English language arts. These standards are aimed at fostering college and career readiness by the end of high school. The CCSS have been adopted formally by 45 states, several territories and the Department of Defense Education Activity (Fields and Parsad; pp. 3-4).

Twelfth grade is the end of mandatory schooling for most students, the transition point to adult postsecondary pursuits. If it is essential for students to graduate from high school academically prepared for college and job training, it is essential for the public and policymakers to know the degree to which this is occurring.

A trusted indicator is needed for reporting to the public and policymakers on the status of 12th grade academic preparedness in the U.S., but no such indicator exists. State tests at the high school level are typically administered at 10th and 11th grade. College admission tests, like the SAT and ACT, are administered before the 12th grade, generally to self-selected samples of students.

State tests and college admission tests do not provide a measure of what students know and can do at the very end of K-12 education. Using state tests and college admission tests for this purpose would be like performing final quality control on a product while it was still on the assembly line. Even if these state tests and college admission tests were administered at the 12th grade, they could not be combined to produce nationally representative results.

NAEP is the only source of national and state-representative student achievement data at the 12th grade. As such, NAEP is uniquely positioned to serve as an indicator of 12th grade academic preparedness.

Defining Academic Preparedness for College

In the United States in 2013, there is no single, agreed upon definition of “academic preparedness for college” used by colleges for admission and placement. Postsecondary education in the U.S. is a complex mix of institutions, public and private, that have different admission requirements and different procedures and criteria for placing individual students into education programs.

In this complex mix are 2-year institutions, 4-year public and private institutions with a wide range of selectivity, and proprietary schools. Institutions range from highly selective (i.e., with admission criteria including very high grade point averages, successful completion of rigorous high school coursework and very high SAT and/or ACT scores) to open admission (i.e., all applicants are admitted).

Even within institutions, requirements may vary across majors or programs of study. For example, the mathematics and science high school coursework and academic achievement needed for acceptance into an engineering program in a postsecondary institution may be more rigorous than the general requirements for admission to the institution, or for a degree in elementary education in the institution.

In order to design the NAEP 12th grade preparedness research, a working definition of preparedness was needed. The Governing Board’s Technical Panel on 12th Grade Preparedness Research recommended use of the following definition with respect to academic preparedness for college.

... the academic knowledge and skill levels in reading and mathematics necessary to be qualified for placement...into a credit-bearing entry-level general education course that fulfills requirements toward a two-year transfer degree or four-year undergraduate degree at a postsecondary institution [without the need for remedial coursework in those subjects]. (National Assessment Governing Board, 2009; p.3.)

This definition was intended to apply to the “typical” college, not to highly selective institutions, and thus, to the vast majority of prospective students, or about 80% of the college freshmen who enrolled in 2-year and 4-year institutions within 2 years following high school graduation (Ross, Kena, Rathbun, KewalRamani, Zhang, Kristapovich, and Manning, p 175). To make this clear, the definition is further elaborated as follows.

Academic preparedness for college refers to the reading and mathematics knowledge and skills needed to qualify for placement into entry-level, credit-bearing, non-remedial courses that meet general education degree requirements in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

This is consistent with the approach used by the College Board and ACT, Inc. in developing their respective college readiness benchmarks, which are used as external referents in the NAEP 12th grade preparedness research. The ACT benchmarks “represent predictive indicators of success for *typical* students at *typical* colleges (Allen and Scoring).” The SAT benchmarks are “an indication of college readiness at a typical college (College Board).”

The Central Issue: Validity

Having made the decision to determine the feasibility of NAEP reporting on 12th grade academic preparedness, the Governing Board recognized that the central concern would be establishing the validity of inferences about 12th grade academic preparedness for use in NAEP reports. The Governing Board would need to ensure that the content of NAEP 12th grade reading and mathematics assessments was appropriate for measuring academic preparedness and that research was conducted to collect evidence by which the validity of proposed inferences could be evaluated. Finally, a formal validity argument would need to be developed, specifying the proposed inference(s) for NAEP reporting, the underlying assumptions or propositions, and the evidence related to the assumptions or propositions.

Accordingly, the Governing Board

- revised the NAEP assessment frameworks for the 2009 12th grade reading and mathematics with the explicit purpose of measuring academic preparedness for college and job training,
- appointed a special panel of technical experts to recommend a program of research on 12th grade academic preparedness ((National Assessment Governing Board, 2009). approved and conducted a comprehensive set of preparedness research studies, and
- adopted the model for a validity argument described by Michael Kane (Kane).

The first phase of the Governing Board’s program of preparedness research is completed. The studies were conducted in connection with the 2009 NAEP 12th grade assessments in reading and mathematics. More than 30 studies of five distinct types have been conducted. Study results are available and the complete studies are posted at <http://www.nagb.org/what-we-do/preparedness-research.html>. The National Center for Education Statistics (NCES) has provide additional data drawn from analyses of the 2005 and 2009 High School Transcript Studies conducted in connection with the NAEP 12th grade assessments in those years.

From this research, Governing Board staff developed a set of proposed inferences related to 12th grade academic preparedness for college. Following below is the validity argument for these proposed inferences. The validity argument begins with a statement of the proposed inferences. This is followed by a discussion of the limitations on interpretation and other caveats. An outline is then presented of the propositions and assumptions on which the inferences are based and the evidence related to the propositions and assumptions. The outline is followed by the text of the validity argument.

Validity Argument

Proposed Inferences

12th grade students scoring at or above the Proficient achievement level on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college.

Limitations on Interpretation and Other Caveats

False Negatives

Some proportion of 12th grade students scoring below Proficient on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first-year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college,

but with a lower probability than those at or above Proficient. In mathematics, much more so than in reading, the research results suggest that the point on the NAEP scale indicating academic preparedness for college (i.e., “just academically prepared”) is below Proficient, somewhere in the middle of the range between the Basic and Proficient achievement level cut scores.

Not a Preparedness Standard

The proposed inferences are not intended to represent or be used as standards for minimal academic preparedness for college. The proposed inferences are intended solely to add meaning to interpretations of the 12th grade Proficient achievement levels in reading and mathematics as used in NAEP reports.

Academically Prepared for College

The proposed inferences are intended to apply to the typical degree-seeking entry-level college student at the typical college. Thus, “academically prepared for first year college courses” refers to the reading and mathematics knowledge and skills needed for placement into entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, the general policies for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

It is important to note the focus on “placement” rather than “admission.” This distinction is made because students who need remedial courses in reading, mathematics or writing may be admitted to college, but not placed into regular, credit-bearing courses. The criterion of importance is qualifying for regular credit-bearing courses, not admission.

The proposed inferences are not intended to reflect academic requirements for highly selective postsecondary institutions; to the additional academic requirements for specific majors or pre-professional programs, such as mathematics, engineering, or medicine; or to academic

requirements applicable to entry into certificate or diploma programs for job training or professional development in postsecondary institutions.

The proposed inferences are focused on the first year of college; they do not support conclusions about college persistence beyond the first year or completion of a degree.

GPA of B- or Better

The selection of “first-year GPA of B- or better” as a referent was made because of its use as a research-based criterion in defining college readiness benchmarks developed by an acknowledged leader in college testing programs—the College Board. The College Board had agreed to partner with the Governing Board in a study linking performance on 12th grade NAEP with the SAT. Another leader in college testing programs, ACT, Inc. has developed similar benchmarks for its college admission assessments using a similar criterion and similar methodology. Because they are based on credible research related to college outcomes, and because performance on the respective tests could be linked to performance on NAEP, the college readiness benchmarks used by these testing programs were embraced as relevant, useful points of reference for the NAEP preparedness research.

The College Board has set a score of 500 or better on the SAT Mathematics and Critical Reading tests as its college readiness benchmarks in those areas. Based on its research, the College Board has determined that the score of 500 or better predicts, with a probability of .65, attainment of a first-year overall GPA of B- or better. Similarly, the ACT college readiness benchmarks are based on research indicating a .50 probability of attaining first-year grades in relevant courses (e.g., college algebra and courses requiring college level reading) of B or better and .75 probability of C or better.

The proposed inferences are not intended to convey that a B- or any particular grade should be deemed a standard or goal for postsecondary student outcomes. This criterion was selected to foster comparability across the preparedness research studies, where applicable. However, it does seem self-evident that achieving a first-year GPA of B- or better, without enrollment in remedial/developmental courses, lends support to the likelihood of having possessed academic preparedness for first-year college courses upon entry to college.

Data Limitations

Although the preparedness research studies are comprehensive and the results consistent and mutually confirming, for reading they are limited to one year for data at the national level and to one state-based longitudinal study. For mathematics, there are two separate years of data at the national level and one state-based longitudinal study. Therefore, more evidence exists to support the plausibility of inferences related to mathematics than to reading.

Preparedness for Job Training

The completed research with respect to academic preparedness for job training does not support conclusions relative to the NAEP scale and will not be addressed at this time.

Validity Argument Outline

| Proposition/Assumption | Evidence |
|--|---|
| <p>1. The content, test questions, and scoring criteria of the NAEP 12th grade reading and mathematics assessments cover academic knowledge and skills needed for college freshmen to be placed into entry-level, credit bearing courses.</p> | <p>1a. The documentation of the content and changes to the content of the 12th grade NAEP reading and mathematics frameworks, as revised in response to recommendations by Achieve, based on their American Diploma Project research</p> <p>1b. Content alignment studies (ACT, SAT, ACCUPLACER)</p> |
| <p>2. The NAEP sampling, scaling and statistical procedures yield accurate estimates of the percentage of students scoring at or above a selected cut-score.</p> | <p>2. NAEP technical documentation of sampling, scaling and statistical procedures.</p> |
| <p>3. Scores on 12th grade NAEP reading and mathematics assessments provide accurate estimates of academic preparedness for entry level credit-bearing college courses.</p> | <p>Performance on the 12th grade NAEP reading and mathematics assessments is related to other indicators or criteria of academic preparedness for placement into entry-level credit-bearing college courses.</p> <p>3a. Linking studies with ACT and SAT</p> <p>3b. Cut-scores on SAT/ACT from higher education survey</p> <p>3c. College Readiness Standards/Benchmarks for the ACT and the SAT</p> <p>3d. Average NAEP scores of Florida students in/not in remedial and with GPA of B- or better</p> <p>Empirical indicators of student engagement do not support the assertion that NAEP 12th grade test-takers are not motivated.</p> <p>3e. Percentage of test items attempted, including constructed response test questions.</p> <p>3f. Correlations between performance on SAT and NAEP</p> |
| <p>4. The proposed test uses are appropriate and consequences are commensurate with intended uses.</p> | <p>5. Intended audience for the results is clearly stated; intended use is clearly described and disseminated to intended audience, along with caveats about potential over- or misinterpretation; the definition of preparedness is clearly defined and qualified; and materials are developed and disseminated consistent with the preceding requirements.</p> |

1. The content, test questions, and scoring criteria of the NAEP 12th grade reading and mathematics assessments cover academic knowledge and skills needed for college freshmen to be placed into entry-level, credit bearing courses.

NAEP Assessment Frameworks Were Revised to Measure Academic Preparedness

The National Assessment Governing Board intentionally revised the NAEP 12th grade reading and mathematics assessment frameworks with the purpose of measuring academic preparedness for college.

On March 5, 2004, the Governing Board accepted the report of the Commission on NAEP 12th Grade Assessment and Reporting. The Commission recommended that “NAEP should report 12th grade students’ [academic preparedness] for college-credit coursework, training for employment, and entrance into the military.”

For NAEP to report on 12th grade academic preparedness for college, it must measure relevant content at the 12th grade. The content of each assessment is determined by the NAEP assessment frameworks, which the Governing Board is responsible for developing and approving.

Accordingly, the Governing Board decided that the extant NAEP frameworks intended for the 2009 for reading and mathematics at the 12th grade would be reviewed. The review would identify changes needed to measure 12th grade academic preparedness for college.²

Assessments at the 12th grade in reading and mathematics are conducted at least once every 4 years. In 2004, when the Board decided to proceed with the 12th grade academic preparedness initiative, 2009 was the next assessment year in which the 12th grade reading and mathematics assessments could be affected by framework changes.

In September 2004, the Governing Board contracted with Achieve, Inc. (Achieve) to review the NAEP 12th grade reading and mathematics assessment frameworks and identify where changes, if any, would be needed. Achieve had established the American Diploma Project (ADP) “...to improve postsecondary preparation by aligning high school standards, graduation requirements and assessment and accountability systems with the demands of college and careers (see www.achieve.org/adp-network).” The ADP had conducted research to identify key competencies in English and mathematics needed for high school graduates who aspire to higher education. They refer to these as the “ADP benchmarks.”

The research and expertise of the American Diploma Project was widely accepted and was brought to bear in reviewing the NAEP frameworks for 12th grade reading and mathematics. Achieve convened a panel of nationally recognized experts in reading and a panel of nationally recognized experts in mathematics. The panels were comprised of individuals from the K-12, postsecondary, research, and policy spheres, knowledgeable about academic preparedness for college reading and college mathematics. The panels compared the 12th grade NAEP reading and mathematics frameworks and the ADP benchmarks.

² The review also addressed academic preparedness for job training, but that part of the NAEP preparedness initiative is not being addressed in this validity argument.

Reading

The Achieve reading panel found considerable similarity between NAEP and the ADP benchmarks for English, although not perfect agreement. This is displayed in the side-by-side chart on pages 30-40 of the Achieve Reading Report (Appendix A). The English benchmarks have eight major components and objectives under each component. Three of these major components were deemed “Not Applicable” to the reading domain: writing, research, and media.

For almost all of the applicable objectives under the five major components that were applicable to the reading domain, the Achieve reading panel found matches in the NAEP 2009 reading framework. Overall, the panel concluded that “...the 2009 NAEP Reading Framework...was aligned to the ambitious [ADP] benchmarks” (Achieve Reading Report, p. 2).

The reading panel also listed items in the NAEP framework that are not found in the ADP English benchmarks. For example, under Argumentation and Persuasive Text, figurative language and rhetorical structure, including parallel structure and repetition, was present in the NAEP reading framework at grade 12, but not in the ADP benchmarks. Under Poetry, tone, complex symbolism, and extended metaphor and analogy are present in the NAEP reading framework but not the ADP benchmarks. A complete listing of the items in the NAEP framework not present in the ADP benchmarks appears on page 41 of the Achieve Reading Report.

Although the Achieve reading panel concluded that the 12th grade NAEP reading framework for 2009 was aligned with the ADP benchmarks applicable to reading, the panel’s report does include six recommendations. The Governing Board approved these recommendations on February 14, 2005. For example, the Achieve reading panel recommended increasing the percentage of informational text passages from 60% to 70% and to feature additional items that ask students to compare texts. The changes were modest, sufficiently so to permit continuation of the 12th grade trend line from its initiation in 1992.

The NAEP reading framework used for the 2009, 2011, and 2013 assessments contains the following statement

In May 2005, the Governing Board adopted a policy statement regarding NAEP and 12th-grade preparedness. The policy states that NAEP will pursue assessment and reporting on 12th-grade student achievement as it relates to preparedness for post-secondary education and training. This policy resulted from recommendations of the Board’s National Commission on NAEP 12th Grade Assessment and Reporting in March 2004. Subsequent studies and deliberations by the Board took place during 2004 and 2005.

In reading, the Board adopted minor modifications to the 2009 NAEP Reading Framework at grade 12 based on a comprehensive analysis of the framework conducted by Achieve, Inc. The current version of the reading framework incorporates these modifications at grade 12 to enable NAEP to measure and report on preparedness for postsecondary endeavors (National Assessment Governing Board, 2008, *Reading Framework*, p. v).

Mathematics

The mathematics review began with the 2007 NAEP mathematics framework, which was the most current and included the changes approved for the 2005 12th grade mathematics assessment. The Achieve panel examined the NAEP mathematics at the 12th grade in relation to the ADP benchmarks for mathematics. The Achieve panel developed proposed revisions to the assessment objectives for grade 12. While acknowledging differences in language and purpose, the Achieve reading panel concluded that the “overall mathematics frameworks of ADP and [12th grade] NAEP are remarkably similar” (see Appendix B, Achieve Mathematics Report, p.9).

The Governing Board convened a panel of mathematicians and mathematics educators to review and revise the objectives in relation to the objectives for grades 4 and 8. The panel conducted focus groups with various NAEP constituents, using repeated rounds of reviews. The Governing Board approved the final set of grade 12 objectives on August 5, 2006. The changes to the framework were sufficiently modest to permit the continuation of the 12th grade trend line begun with the 2005 12th grade mathematics assessment under the previous 12th grade framework. Like the reading framework, the 2009/2013 mathematics framework for grade 12 states the Board’s intention to measure 12th grade academic preparedness (National Assessment Governing Board, 2008, *Mathematics Framework*, pp. 2-3).

Examples of Objectives added to the 2009 Grade 12 Mathematics Framework

Number properties and operations

b) * Analyze or interpret a proof by mathematical induction of a simple numerical relationship.

Measurement

d) Interpret and use the identity $\sin^2\theta + \cos^2\theta = 1$ for angles θ between 0° and 90° ; recognize this identity as a special representation of the Pythagorean theorem.

e) * Determine the radian measure of an angle and explain how radian measurement is related to a circle of radius 1.

f) * Use trigonometric formulas such as addition and double angle formulas.

g) * Use the law of cosines and the law of sines to find unknown sides and angles of a triangle.

Geometry

e) * Use vectors to represent velocity and direction; multiply a vector by a scalar and add vectors both algebraically and graphically.

g) * Graph ellipses and hyperbolas whose axes are parallel to the coordinate axes and demonstrate understanding of the relationship between their standard algebraic form and their graphical characteristics.

h) * Represent situations and solve problems involving polar coordinates.

Data Analysis, Statistics, and Probability

c) * Draw inferences from samples, such as estimates of proportions in a population, estimates of population means, or decisions about differences in means for two “treatments”.

e) * Recognize the differences in design and in conclusions between randomized experiments and observational studies.

k) * Use the binomial theorem to solve problems.

e) * Recognize and explain the potential errors caused by extrapolating from data.

Algebra

e) Identify or analyze distinguishing properties of linear, quadratic, rational, exponential, or trigonometric functions from tables, graphs, or equations.

j) * Given a function, determine its inverse if it exists and explain the contextual meaning of the inverse for a given situation.

h) * Analyze properties of exponential, logarithmic, and rational functions.

g) * Determine the sum of finite and infinite arithmetic and geometric series.

Conclusion

The Governing Board, by official action, revised the NAEP 12th grade reading and mathematics frameworks for the explicit purpose of measuring 12th grade academic preparedness for college, beginning with the 2009 assessments. Setting forth the measurement purpose and making relevant revisions to the NAEP assessment frameworks are necessary elements of the validity argument; however, they are not sufficient. Evidence must be considered with respect to the alignment of the framework and the test questions administered to the measurement purpose. This will be addressed in the next section.

Content Alignment Studies Found Significant Overlap between NAEP and the ACT, SAT and ACCUPLACER

The Governing Board conducted studies to determine the degree of content similarity between NAEP 12th grade reading and mathematics assessments and relevant tests used for college admissions and placement.

The studies had two objectives. First, to determine the degree to which the content of 12th grade NAEP in reading and mathematics covers the reading and mathematics knowledge and skills needed for first year college work. The SAT, ACT, and ACCUPLACER are well-established tests that assess individual students' reading and mathematics proficiency in relation to college level expectations.

The ACT is developed with the purpose of "...[measuring] as directly as possible the degree to which each student has developed the academic skills and knowledge that are important for success in college..." (ACT Technical Manual, p. 62).

The SAT is developed "to ensure that the topics measured on the SAT...reflect what is being taught in the nation's high schools and what college professors consider to be required for college success." (Kim, Wiley, and Packman, p.1)

The ACCUPLACER has the purpose of "... [determining] which course placements are appropriate for [incoming college] students and whether or not remedial work is needed." (ACCUPLACER, p. A-2)

The SAT, ACT and ACCUPLACER in reading and mathematics are widely used for these purposes by admissions and placement professionals in postsecondary education institutions. These testing programs regularly conduct curriculum surveys, validity studies and other research to support their claims that the content measured is directly related to the reading and mathematics knowledge and skills needed to qualify for entry-level credit-bearing courses.

Therefore, with the assumption that the SAT, ACT, and ACCUPLACER do measure the content needed for college level work, significant content overlap between NAEP and these other assessments would support the conclusion that what NAEP measures covers the knowledge and skills needed by college freshmen to be placed into entry-level credit bearing courses. That is,

- If A (the college admissions and placement tests) = B (the knowledge and skills needed for placement into entry-level credit-bearing courses without remediation); and
- C (NAEP) = A;
- Then C = B).

The second reason for conducting the content alignment studies was to provide information for interpreting the results of planned statistical linking studies between NAEP and the other tests, which measure academic preparedness for college. The linking studies were designed to examine how performance on NAEP compares with performance on the other tests, with the purpose of supporting inferences about academic preparedness for college. For NAEP to support inferences about academic preparedness for college based on the linking studies, a sufficient content match would be needed, not just a statistical relationship.

The Content Alignment Studies: Overview

The Governing Board conducted content alignment studies in reading and mathematics comparing the 2009 12th grade NAEP and the ACT, SAT, and ACCUPLACER reading and mathematics tests. Overall, considerable overlap was found between the ACT and NAEP and the SAT and NAEP, with some differences. NAEP was found to measure much of what is measured on the ACCUPLACER, but the reading and mathematics domains measured by NAEP were much broader than ACCUPLACER. More details are provided in the summaries of the individual studies below.

The general design for the content alignment studies was to compare the 12th grade NAEP frameworks in reading and mathematics with the analogous document for the other test, and then to compare the test items from one test to the framework/analogous document of the other test. The reviews were performed by subject specific (i.e., mathematics, reading) panels, composed of experts in mathematics or reading and English instruction at the high school and college levels.

Alignment studies that compare an assessment to the content standards on which it is based are relatively common and have well-established methodologies. However, this is not true for the types of alignment studies the Governing Board planned to conduct: content alignment studies comparing different assessment programs. Different assessment programs have different purposes, different approaches to describing the domain being measured, and, possibly, different “grain size” in the level of detail in describing the domain. The Governing Board contracted with Norman Webb, a noted expert in content alignment studies, to prepare a design document for conducting the assessment to assessment alignment studies. The purpose was to put in place a methodology that considered the special challenges of assessment to assessment alignment studies and to foster comparability in the conduct of the studies and the reporting metrics across studies and contractors. The link to the Webb design document is at (<http://www.nagb.org/content/nagb/assets/documents/publications/design-document-final.pdf>). The Webb design was developed after the ACT alignment studies were completed. It was used in conducting the SAT and ACCUPLACER content alignment studies.

In the following sections are summaries of the content alignment study results, excerpted from the study reports. The results for the three content alignment studies in reading are presented first, followed by the three content alignment studies for mathematics, along with summary discussions for the reading and mathematics results.

The Content Alignment Studies: Reading Results

Reading: ACT

The Governing Board contracted with ACT, Inc. to conduct the content alignment study comparing the NAEP 12th grade reading assessment and the ACT reading test. The full report can be found at [http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACT-NAEP Math and Reading Content Comparison.pdf](http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACT-NAEP_Math_and_Reading_Content_Comparison.pdf).

The reading panel was composed of 7 members, with expertise in reading and/or English instruction at the high school and college levels. The panel was about evenly divided in terms of prior familiarity with either the ACT or NAEP reading domains.

The panel found considerable similarity in the content of the NAEP 12th grade reading assessment and the ACT. For example, the NAEP 12th grade reading framework was compared to the ACT reading domain and the ACT College Readiness Standards for reading. The ACT College Readiness Standards (CRS) are descriptions of the content (i.e., the knowledge and skills) measured by the ACT reading test in score bands along the ACT 1-36 point scale from 13-36 (see <http://www.act.org/standard/planaact/reading/>). The panel concluded that

“All of the skills highlighted in the ACT [reading] domain and in the [ACT] College Readiness Standards [for reading] were identified within the NAEP Reading framework. In performing the comparison in the other direction—NAEP to ACT—it was the sense of the panel that the ACT measured primarily those skills that NAEP identifies as *Locate/Recall* and *Integrate/Interpret* skills, those that pertain primarily to finding explicit information in text (what the ACT would call Referring skills) and to making inferences, drawing conclusions, and making generalizations from information within text (what the ACT would call Reasoning skills). The panel saw less evidence of the higher-level analytical and evaluative *Critique/Evaluate* skills in the ACT domain, and attributed that to the multiple-choice format of the ACT [whereas NAEP includes constructed response items as well as multiple choice]. Another difference is that NAEP includes items and texts measuring how well an examinee can apply reading skills across texts, whereas the paired passage format is not a feature of the ACT. So, while the NAEP Reading framework and the ACT Reading domain, test specifications, and College Readiness Standards share similarities, important differences in what and how the assessments measure suggest caution when drawing comparisons between the assessments.” (p.17)

The reading panel also conducted an item classification study, in which the NAEP 12th grade reading items were classified in relation to the ACT College Readiness Standards for Reading.

“A total of 152 Reading items (comprising 17 blocks) were classified in [the reading] study. Of these, 97 were multiple-choice (MC). Nine were dichotomously-scored (“incorrect” or “correct”) short constructed-response (DSCR) items. Thirty-three were polytomously-scored short constructed-response (PSCR) items, each scored using a three-point scoring rubric. Thirteen were extended constructed-response (ECR) items, each scored using a four-point rubric. Each DSCR had one creditable score category, each PSCR had two, and each ECR had three. Each Reading panelist, therefore, assigned a total of 211 classifications to the NAEP Reading items [and rubric scoring categories].” (p.54)

An item or score category was deemed “classified” if there was majority agreement (at least 4 of the 7 panel members) or supermajority agreement (5 or more panel members) about the score band to which an item (or creditable score category under an item rubric) was assigned.

Of the 211 determinations to be made, there was only one for which there was no majority agreement (the assignment of a PSCR rubric to a CRS score band). Of the remaining 210 determinations, 181 were unanimous.

The reading panel was able to classify 137 items or rubric categories (about two-thirds of the determinations to be made) to the CRS score bands. Of the 97 multiple choice items, 81 (or 84%) were classified. Of the 113 rubric score categories for items, 56 (or 50%) were classified. The reasons some multiple choice items and rubric score categories could not be classified were related to the differences in the ACT and NAEP reading domains described above. These reasons include the presence of constructed response items in NAEP but not the ACT, the presence of

items involving multiple texts in NAEP but not the ACT, and the greater presence of “Critique/Evaluate” type items in NAEP than the ACT.

Of the 137 classifications, 24 were in the score bands from 13-19; 113 of the classifications were in the score bands from 20-36. This is noted because the ACT College Readiness Benchmark for reading is 21. The ACT College Readiness Benchmark signifies the score at which a student has a 50% chance of attaining a grade of B or better in a relevant subject and a 75% change of a C or better. In addition, the Governing Board conducted a survey of postsecondary institutions’ use of tests in making entry-level decisions about placement into remedial or regular credit-bearing courses. With respect to the ACT, 18 was the mean reading score below which students were deemed to need remedial course work (Fields and Parsad, P. 19). While this provides a context for the study results, it must be kept in mind that in making their judgments about assessment content, the panelists did not have data about NAEP item difficulty or data on how performance on NAEP compares with performance on the ACT.

Finally, while the study results support the conclusion that the 12th grade NAEP reading assessment measures content directly related to academic preparedness for college, it is noted that the study was conducted by ACT, Inc., not an independent third party. Further, because a different methodology was used, the study results are not directly comparable to the results for the SAT and ACCUPLACER alignment studies in reading.

Reading: SAT

The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade reading assessment and the SAT critical reading test. WestEd conducted the content alignment study using the design developed for the Governing Board by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Reading_Content_Comparison.pdf

Overall, the study found similar content in the NAEP 12th grade reading assessment and the SAT critical reading test. Following below is an excerpt from the Executive Summary of the report (pp. iv-vi).

What is the correspondence between the reading content domain assessed by NAEP and that assessed by SAT?

The greatest commonality between the two tests is their shared emphasis on the broad skills of integrating and interpreting both informational and literary texts. This is evident in the majority of items from both tests aligned to NAEP Standard 2, Integrate/Interpret,” including many to Goal 2.1, “Make complex inferences within and across *both literary and informational texts.*”

Despite the difference in the degree of specificity of the two frameworks (most NAEP objectives are much more finely grained than the SAT objectives), there is also considerable overlap at the level of more specific skills.

To what extent is the emphasis of reading content on NAEP proportionally equal to that on SAT?

Both tests had many of their item alignments to the same NAEP “Integrate/Interpret” objectives, often with similar percentages of alignments. Although there were some differences in emphasis, both tests also had notable percentages of alignments to SAT Objectives B.1.1–B.1.3 and B.1.5. Skills with overlap include inferring/analyzing the following:

- the “main idea” and “author’s purpose” (SAT Objective B.1.1 and NAEP Objectives 2.3.a and 2.1.f);
- the “tone and attitude” of an author or character (NAEP Objectives 2.2.a and 2.2.c and SAT Objective B.1.4);
- the use of “rhetorical strategies” (NAEP Objective 2.1.d and SAT Objective B.1.2); and
- connections between ideas, perspectives, or problems (NAEP Objective 2.1.b and SAT Objectives B.1.3 and B.1.5).

Additionally, in the area of greatest content overlap—items on both tests aligned to objectives for NAEP “Integrate/Interpret” and aligned to SAT “Passage-Based Reading” Objectives B.1.1– B.1.5—both tests met the typical threshold criteria for depth of knowledge consistency...

Despite these similarities, there are some notable differences in emphasis between the two assessments. Both tests assess vocabulary skills. However, NAEP addresses vocabulary exclusively in the context of passage comprehension, while the majority of SAT vocabulary items are in a sentence-completion format, in which context plays a more limited role. This difference reflects NAEP’s emphasis on the understanding of word meaning in context; the assessment is not intended to measure students’ prior knowledge of word definitions. The SAT sentence-completion items provide some context within the single sentence text, but in many cases, students’ success on the items almost certainly depends on their prior knowledge of word definitions.

In addition, panelists found considerably less emphasis in SAT than in NAEP on literal comprehension and critical evaluation, particularly the evaluation of the quality or effectiveness of an author’s writing, skills covered in the NAEP standards “Locate/Recall” (locating/recalling specific details and features of texts) and “Critique/Evaluate” (evaluating texts from a critical perspective), respectively. This difference suggests a greater emphasis on these skills in NAEP.

Even with the minimal coverage of NAEP “Locate/Recall” and “Critique/Evaluate” standards by SAT items, all NAEP items found a match in the SAT framework. However, the broad language of the SAT framework can encompass the range of the NAEP items. For example, SAT Goal B.2, “Literal Comprehension,” refers to items that “ask what is being said” in a “small but significant portion of a reading passage,” a description that can easily accommodate most NAEP “Locate/Recall” items and objectives. In fact, nearly all items on the NAEP short version that were coded to

“Locate/Recall” objectives in the NAEP framework were matched to SAT Goal B.2 in the SAT framework.

Similarly, SAT Objective B.1.3, to which approximately one-quarter of NAEP items aligned, includes “Evaluation,” the primary focus of NAEP “Critique/Evaluate.” The description in SAT Objective B.1.3 of items that “ask the test taker to evaluate ideas or assumptions in a passage” is compatible at a very general level with NAEP “Critique/Evaluate” objectives addressing the author’s point of view, logic, or use of evidence. SAT Objective B.1.2, “Rhetorical Strategies,” is also broad enough in its language to make it a reasonable match for some NAEP “Critique/Evaluate” items focused on “author’s craft” or use of “literary devices.” In the NAEP short version, all items that aligned to “Critique/Evaluate” objectives in the NAEP framework were aligned to either SAT Objectives B.1.2 or B.1.3, or both.

Are there systematic differences in content and complexity between NAEP and SAT assessments in their alignment to the NAEP framework and between NAEP and SAT assessments in their alignment to the SAT framework? Are these differences such that entire reading subdomains are missing or not aligned?

With regard to differences in content as described in the NAEP framework, SAT items had limited coverage of the knowledge and skills described by the NAEP standards “Locate/Recall” and “Critique/Evaluate.” This difference is also reflected in test format, with the use of longer reading passages and both constructed-response and multiple-choice items in NAEP. In comparison, all SAT items are multiple-choice. With regard to differences in content as described in the SAT framework, NAEP does not include sentence-completion items.

With regard to differences in complexity, NAEP items and objectives had a range of depth of knowledge including items at DOK Levels 1, 2, and 3, while SAT items and objectives were coded primarily at Levels 2 and 3.

Overall, the alignment results across the two sets of items and frameworks show a strong area of overlap in their coverage of SAT “Passage-Based Reading” objectives and NAEP “Integrate/Interpret” objectives, as well as some important differences.

Reading: ACCUPLACER

The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade reading assessment and the ACCUPLACER reading test. The ACCUPLACER is used specifically to determine whether entry-level students have the reading skills necessary for college level work or require remedial reading courses. WestEd conducted the content alignment study using the design developed for the Governing Board by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Reading_Content_Comparison.pdf.

Overall, the study found similar content in the NAEP 12th grade reading assessment and the ACCUPLACER reading test, although the content of NAEP is much broader and complex. Following below is an excerpt from the Executive Summary of the report (pp. iv-vi).

What is the correspondence between the reading content domain assessed by NAEP and that assessed by ACCUPLACER?

The greatest commonality between the two tests is in their shared emphasis on the broad skills of comprehending and interpreting informational text, primarily through inferential reasoning. This is evident in the majority of items on both tests (two-thirds to three-fourths) matched to the NAEP standard “Integrate/Interpret: Make complex inferences within and across texts.” On both tests, the majority of alignments to “Integrate/Interpret” were to objectives that apply to informational text only or across both informational and literary texts.

The shared emphasis on the comprehension and interpretation of informational text can also be seen in the alignments on both tests to the ACCUPLACER framework. Although the ACCUPLACER standards do not explicitly refer to text type, they focus almost exclusively on elements typical of informational text. A majority of both NAEP and ACCUPLACER items were matched to the ACCUPLACER standard “Inferences,” and both tests had notable percentages of alignments to “Direct statements and secondary ideas” and “Applications.” A smaller percentage of items on both tests were aligned to “Identifying main ideas.”

To what extent is the emphasis of reading content on NAEP proportionally equal to that on ACCUPLACER?

As previously discussed, the alignments both within and across frameworks show that both tests emphasize the comprehension and interpretation of informational text, particularly through the use of inference. Within this broad area of convergence, however, there are differences in emphasis revealed in the alignments to specific objectives within both frameworks. In relation to the NAEP framework, the NAEP short-version items showed a far greater emphasis on the comprehension of vocabulary in context (Objective 4.a) and on the analysis of an author’s use of language (Objective 1.d). In relation to the ACCUPLACER framework, NAEP items showed more emphasis on the use of inference to interpret text (“Inferences”). The higher percentage of NAEP items aligned to “Applications” also reflects the greater emphasis in NAEP on understanding authors’ use of language.

In relation to the ACCUPLACER framework, the ACCUPLACER items showed a greater emphasis than the NAEP items on the identification of main ideas. In relation to the NAEP framework, the ACCUPLACER items showed more emphasis on the recall of specific details, facts, and information (NAEP 1.1.a).

In general, in the cross-framework alignments, the matches found in each test to the other’s framework (NAEP to ACCUPLACER and ACCUPLACER to NAEP) tended to be for the most general objectives within that framework. For example, the great majority

of hits for ACCUPLACER items to NAEP objectives for “Integrate/Interpret” were to two of the most broadly stated NAEP objectives, “Draw conclusions” (2.3.b) and “Compare or connect ideas” (2.1.b). Many of the more specific NAEP objectives for “Integrate/Interpret,” such as “Find evidence in support of an argument” (2.2.c), received far fewer or no hits from ACCUPLACER items. Compared to ACCUPLACER, the NAEP items were more evenly distributed among NAEP objectives.

The majority of alignments for NAEP items to ACCUPLACER standards were also to the broadest of those standards—“Inferences” and “Applications,” both of which overlap in content with a number of NAEP objectives but at a higher level of generality. The more specific ACCUPLACER standard, “Identifying main ideas,” received far fewer alignments from NAEP items.

Are there systematic differences in content and complexity between the NAEP and ACCUPLACER assessments in their alignment to the NAEP framework and between the NAEP and ACCUPLACER assessments in their alignment to the ACCUPLACER framework? Are these differences such that entire reading subdomains are missing or not aligned?

In regard to differences in content, NAEP addresses reading skills related to both literary and informational text, while ACCUPLACER does not address reading skills specific to literary text. As expected, based on the framework-to-specifications [review]... ACCUPLACER items had minimal matches to NAEP objectives for literary text. The main area of alignment of ACCUPLACER items to the NAEP framework, NAEP objectives in “Locate/Recall” and “Integrate/Interpret,” applied to informational text only or to both informational and literary text.

The ACCUPLACER items also had minimal to no coverage of the NAEP standard “Critique/Evaluate.” ... overall, the language of the ACCUPLACER objectives (“understand,” “comprehend,” “recognize”) places more emphasis on comprehension and interpretation of text (“distinguish the main idea from supporting ideas” or “perceive connections between ideas made—implicitly—in the passage”) than on critical analysis or evaluation (“Evaluate the strength and quality of evidence used by the author to support his or her position” in NAEP Objective 3.3.b, or “Judge the author’s craft and technique” in NAEP Objective 3.1.a).

In regard to complexity, both assessments were found to meet the criteria for depth of knowledge consistency in relation to their own framework. In relation to the NAEP framework, however, only the NAEP items met the criteria for DOK consistency for all NAEP standards. The ACCUPLACER items met the criteria for depth of knowledge consistency only for NAEP “Locate/Recall.”

Although the majority of the ACCUPLACER item alignments were to objectives for NAEP “Integrate/Interpret,” over half of these items were found to have a DOK level below that of the standard. In addition, the use of very short reading passages and exclusively multiple-choice items in ACCUPLACER may be less conducive to the more

in-depth reasoning required by DOK Level 3. NAEP, by contrast, includes much longer reading passages and both multiple-choice and constructed-response items.

NAEP covers skills specific to the comprehension and analysis of literary text while ACCUPLACER does not. In addition, NAEP covers the skills of evaluating and critiquing text, skills not addressed by ACCUPLACER. Finally, NAEP has a wider range of cognitive complexity than ACCUPLACER, with a substantially higher percentage of items at DOK Level 3, requiring more in-depth analysis or evaluation. However, both tests show a similar emphasis on applying interpretive skills and inferential reasoning to the understanding of informational text.

Overall, the NAEP items covered a broader range of cognitive complexity than the ACCUPLACER items. This is also apparent in the frameworks. The three NAEP standards, defined in terms of three different “cognitive targets” (“Locate/Recall,” “Integrate/Interpret,” and “Critique/Evaluate”), cover a broader range of cognitive complexity supported by the use of longer reading passages and the inclusion of both short and extended constructed-response items. The language of the ACCUPLACER standards (“understand,” “comprehend,” “recognize”) places more emphasis on comprehension and interpretation of text (e.g., “distinguish the main idea from supporting ideas” in ACCUPLACER A, “Identifying main ideas,” or “perceive connections between ideas made—implicitly—in the passage” in ACCUPLACER C, “Inferences”) than on critical analysis or evaluation (e.g., “Evaluate the strength and quality of evidence” in NAEP 3.3.b, or “Judge the author’s craft” in NAEP 3.1.a). In addition, the use of very short reading passages and exclusively multiple-choice items in ACCUPLACER may be less conducive to the cognitive complexity typical of DOK Level 3 items. Although the NAEP items show a greater range of cognitive complexity and a greater emphasis on critical thinking, both tests show a similar emphasis on applying interpretive skills and inferential reasoning to the understanding of informational text.

The Content Alignment Studies: Summary Discussion for Reading

Three content alignment studies were conducted to examine the extent to which

- The content of the NAEP 12th grade reading assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing courses. and
- NAEP 12th grade reading test items and scoring criteria are appropriate for obtaining evidence of test takers’ possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing courses requiring college level reading.

For short-hand, this will be referred to as “academic preparedness for college.”

The NAEP 12th grade reading framework, test questions, and, for constructed response items, the score category rubrics, were compared with the analogous domain descriptions and test questions for the ACT, SAT, and ACCUPLACER reading tests. These three tests are used for college admissions and placement. They are well established and have been used for these purposes for many years by professionals in postsecondary education. The test publishers regularly survey secondary and postsecondary educators about relevant content and have conducted research that supports the validity of the test content for the intended inferences and uses. The underlying

assumption is that if the content of the 12th grade NAEP reading assessment is similar to the content of these reading tests, then the NAEP content is directly related to “academic preparedness for college.”

The ACT study found that “All of the skills highlighted in the ACT [reading] domain and in the [ACT] College Readiness Standards [for reading] were identified within the NAEP Reading framework.” At the same time, there was content measured by NAEP that was not present in the ACT reading test. In assigning 211 NAEP 12th grade reading items and rubric score categories to the ACT College Readiness Standards for reading, there were 137 positive classifications, or about 65% of the possible classifications. The multiple choice items and rubric score categories that could not be classified were those that measured content not measured by the ACT reading test.

The SAT study found that “Overall, the alignment results across the two sets of items and frameworks show a strong area of overlap in their coverage of SAT “Passage-Based Reading” objectives and NAEP “Integrate/Interpret” objectives, as well as some important differences.” With respect to the differences, “...SAT items had limited coverage of the knowledge and skills described by the NAEP standards “Locate/Recall” and “Critique/Evaluate.” This difference is also reflected in test format, with the use of longer reading passages and both constructed-response and multiple-choice items in NAEP. In comparison, all SAT items are multiple-choice. With regard to differences in content as described in the SAT framework, NAEP does not include sentence-completion items.”

The ACCUPLACER study found that “The greatest commonality between the two tests is in their shared emphasis on the broad skills of comprehending and interpreting informational text, primarily through inferential reasoning. This is evident in the majority of items on both tests (two-thirds to three-fourths) matched to the NAEP standard ‘Integrate/Interpret: Make complex inferences within and across texts.’ On both tests, the majority of alignments to ‘Integrate/Interpret’ were to objectives that apply to informational text only or across both informational and literary texts... Overall, the NAEP [frameworks and] items covered a broader range of cognitive complexity than the ACCUPLACER items... The three NAEP standards, defined in terms of three different “cognitive targets” (“Locate/Recall,” “Integrate/Interpret,” and “Critique/Evaluate”), cover a broader range of cognitive complexity supported by the use of longer reading passages and the inclusion of both short and extended constructed-response items.”

The results across the three studies are consistent. In general, the content of the ACT, SAT, and ACCUPLACER reading tests are present in NAEP, but NAEP is generally broader. Alignment between NAEP and the other three respective assessments is substantial, but not perfect; perfect alignment is not expected. A component of the SAT critical reading assessment not present in NAEP is sentence completion, measuring vocabulary knowledge in a different way than NAEP does.

These results support the conclusion that

- The content of the NAEP 12th grade reading assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing courses. and
- NAEP 12th grade reading test items and scoring criteria are appropriate for obtaining evidence of test takers' possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing courses requiring college level reading.

The Content Alignment Studies: Mathematics Results

Mathematics: ACT

The Governing Board contracted with ACT, Inc. to conduct the content alignment study comparing the NAEP 12th grade mathematics assessment and the ACT mathematics test. The full report can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACT-NAEP_Math_and_Reading_Content_Comparison.pdf.

The mathematics panel was composed of 7 members, with expertise in mathematics instruction at the high school and college levels. The panel was about evenly divided in terms of prior familiarity with either the ACT or NAEP mathematics domains.

The panel found considerable similarity in the content of the NAEP 12th grade mathematics assessment and the ACT. For example, the NAEP 12th grade mathematics framework was compared to the ACT mathematics domain and the ACT College Readiness Standards for mathematics. The ACT College Readiness Standards (CRS) are descriptions of the content (i.e., the knowledge and skills) measured by the ACT mathematics test in score bands along the ACT 1-36 point scale from 13-36 (see <http://www.act.org/standard/planact/math/index.html>). The panel concluded that

“... the two assessments have much of their content domains in common. However, in the NAEP-to-ACT comparison, the difference in specificity with which the domains are articulated in the assessment documents left the panel uncertain as to whether a number of NAEP content topics—those pertaining to transformations, probability, statistics, and data analysis—are assessed by the ACT. In addition, there was some uncertainty within the panel on the degree to which higher-order analytic skills were assessed, and it was the sense of the panel that the ACT Mathematics Test contained few items involving high mathematical complexity, at least as the NAEP defines it. With regard to the ACT to-NAEP comparison, the Mathematics panel found nearly all of the ACT Mathematics domain and College Readiness Standards reflected in the NAEP Mathematics domain, but determined that a number of the lower-level topics in the ACT Pre-Algebra subdomain were more consistent with Grade 8 NAEP topics. All of these points suggest that while there may be substantial overlap in what the two assessments measure and how they measure it, there are areas of difference, as well. (p. 17)

The mathematics panel also conducted an item classification study, in which the NAEP 12th grade mathematics items were classified in relation to the ACT College Readiness Standards for Mathematics.

An item or score category was deemed “classified” if there was majority agreement (at least 4 of the 7 panel members) or supermajority agreement (5 or more panel members) about the score band to which an item (or creditable score category under an item rubric) was assigned.

Of the 229 determinations to be made, panel members believed that every item or rubric category could be classified to some CRS score range. However, there were 39 for which there was no majority agreement (17 multiple choice items and 22 rubric categories) on what the classification should be; therefore those items were not considered assigned to a CRS score band. Of the remaining 190 determinations, 24 were unanimous, 142 involved classifications to adjacent score ranges and 24 involved classifications to non-adjacent score ranges.

Of the 108 multiple choice items, 91 (or 84%) were classified. Of the 121 rubric score categories for items, 99 (or 82%) were classified.

Of the 190 classifications, 10 were in the score bands from 13-19; 180 of the classifications were in the score bands from 20-36. This is noted because the ACT College Readiness Benchmark for mathematics is 22. The ACT College Readiness Benchmark signifies the score at which a student has a 50% chance of attaining a grade of B or better in a relevant subject and a 75% change of a C or better. In addition, the Governing Board conducted a survey of postsecondary institutions’ use of tests in making entry-level decisions about placement into remedial or regular credit-bearing courses. With respect to the ACT, 19 was the mean mathematics score below which students were deemed to need remedial course work in mathematics (Fields and Parsad, p. 13). While this provides a context for the study results, it must be kept in mind that in making their judgments about content, the panelists did not have data about NAEP item difficulty or data on how performance on NAEP compares with performance on the ACT.

Finally, while the study results support the conclusion that the 12th grade NAEP mathematics assessment measures content directly related to academic preparedness for college, it is noted that the study was conducted by ACT, Inc., not an independent third party. Further, because a different methodology was used, the study results are not directly comparable to the results for the SAT and ACCUPLACER alignment studies in mathematics.

Mathematics: SAT

The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade mathematics assessment and the SAT mathematics test. WestEd conducted the content alignment study using the design developed for the Governing Board by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Math_Content_Comparison.pdf.

Overall, the study found similar content in the NAEP 12th grade mathematics assessment and the SAT mathematics test. Following below is an excerpt from the Executive Summary of the report (pp. iv-vi).

“What is the correspondence between the mathematics content domain assessed by NAEP and that assessed by SAT?”

At the standard level, the wording of the standards in the two frameworks is very similar. Both the NAEP and SAT frameworks include virtually the same five broad content categories, with SAT combining geometry and measurement into one standard. Each framework contains both general and specific objectives, although the SAT objectives, which are presented as content topics without indication of the cognitive level at which that content would be assessed, may be interpreted as more general than the NAEP objectives.

Although the structures of the two frameworks differ greatly beyond the standard level (including the NAEP framework having three levels while SAT has two), the mathematics areas typically expected of grade 12 students—number and operations, geometry and measurement, data analysis and probability, and algebra—are addressed in somewhat similar proportions.

To what extent is the emphasis of mathematics content on NAEP proportionally equal to that on SAT?”

The greatest commonality between the two tests is their emphasis at the standard level. This is evident in the distribution of percentages of total hits from both assessments matched to each set of standards. Although there are some differences of emphasis, such as the full NAEP item pool’s greater proportion of alignment to SAT “Data analysis, statistics, and probability,” and the SAT short-version’s greater proportion of alignment to SAT “Geometry and measurement,” the proportions of alignments to “Algebra and functions” and “Number and operations” are comparable. There is also considerable overlap among some specific skills, with both assessments addressing many of the same NAEP “Number properties and operations” objectives and SAT objectives...

Despite the difference in the degree of specificity of the two frameworks (most NAEP objectives are much more finely grained than the SAT objectives), it is clear that both assessments emphasize a number of the same or closely related skills. These include properties, equivalence, and operations on rational numbers (included in NAEP Goals 1.1 and 1.3 and included in SAT Objective N.2) and properties of two-dimensional shapes (included in NAEP Goals 3.1 and 3.3 and included in SAT Objective G.6).

Are there systematic differences in content and complexity between NAEP and SAT assessments in their alignment to the NAEP framework and between NAEP and SAT assessments in their alignment to the SAT framework? Are these differences such that entire mathematics subdomains are missing or not aligned?”

While there is considerable overlap between the two assessments, primarily in the intersection of the NAEP “Algebra” and SAT “Algebra and functions” standards, there

are notable differences as well. The SAT items had a somewhat limited range of coverage of the NAEP standards “Measurement,” “Geometry,” and “Data analysis, statistics, and probability,” with several goals receiving few item alignments. Even given the minimal coverage of some of the goals within each NAEP standard by SAT items, however, almost all NAEP items found a match in the SAT framework. The language of the objectives in the SAT framework is sufficiently broad to encompass the range of the NAEP items. For example, SAT Objective A.10, “Basic concepts of algebraic functions,” may accommodate most of the items aligning to the seven objectives within NAEP Goal 5.1, “Patterns, relations, and functions.” Finally, some NAEP items were found to be uncodable to the SAT objectives. These items assessed skills not present in the SAT framework.

The two tests are also similar in the average DOK [Depth of Knowledge] levels of items. However, while most items in both tests were found to be at DOK Level 2, NAEP items had a wider range of DOK than did SAT items, with more NAEP items coded to Levels 1 and 3. The Level 3 NAEP items often involved application of concepts through short or extended constructed-response items. Both tests also met depth-of-knowledge consistency overall (with each not meeting this criterion for only one standard as rated by one panel).

Overall, despite differences in alignment at the detailed specific objective level, differences in emphasis at the standard level, and a small difference in ranges of depth of knowledge, there is considerable overlap of content and complexity between [the NAEP 12th grade mathematics assessment and the SAT mathematics test].”

Mathematics: ACCUPLACER

The Governing Board contracted with WestEd, an independent third party, to conduct the content alignment study comparing the NAEP 12th grade mathematics assessment and the ACCUPLACER mathematics test. The ACCUPLACER is used specifically to determine whether entry-level students have the mathematic knowledge and skills necessary for college level work or require remedial mathematics courses.

WestEd conducted the content alignment study using the design developed for the Governing Board by Norman Webb. The full report of the content alignment study can be found at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Math_Content_Comparison.pdf.

Overall, the study found similar content in the NAEP 12th grade reading assessment and the ACCUPLACER reading test, although the content of NAEP is much broader and complex. Following below is an excerpt from the Executive Summary of the report (pp. iv-vi).

“What is the correspondence between the mathematics content domain assessed by NAEP and that assessed by ACCUPLACER?”

The NAEP and ACCUPLACER assessments both cover certain content traditionally expected of grade 12 students, namely the two content subdomains of number or number operations and algebra (included in NAEP’s “Number properties and operations” and “Algebra” standards and in ACCUPLACER’s “Arithmetic,” “Elementary algebra,” and “College level math” standards), although their respective degrees of alignment and focus in these subdomains vary. Whereas the NAEP items focus primarily on number or number operations and algebra content at the grade 12 level, with an emphasis on problem solving and application of concepts at that grade level, the ACCUPLACER items span a wider developmental and grade-level range (from basic to more advanced). This difference in focus is consistent with the purposes of the two assessments and their frameworks. The NAEP objectives are written to describe assessable content for grade 12 mathematics; thus, the 130 objectives tend to address the skills and concepts specific to that grade. The purpose of ACCUPLACER is to help determine appropriate placement for an individual student, and so the 87 ACCUPLACER objectives are spread more broadly across grade levels and are intended to be more general.

To what extent is the emphasis of mathematics content on NAEP proportionally equal to that on ACCUPLACER?

Regarding alignment to the NAEP framework, within the “Number properties and operations” and “Algebra” standards, NAEP items had broader overall coverage of the NAEP objectives than did ACCUPLACER. The 42 NAEP items (the short version used for within-framework alignment) aligned to 72 NAEP objectives, whereas the 105 ACCUPLACER items (one complete form of each of the three ACCUPLACER Mathematics Core tests) aligned to only 56 NAEP objectives, with 44% of the ACCUPLACER item alignments aligning to only three NAEP objectives (all in “Number properties and operations” and “Algebra”). These differences in breadth and emphasis between the two assessments were evident across all NAEP standards. For example, in each assessment, items were aligned to four NAEP “Algebra” objectives for which the other assessment had no alignments, reflecting differences in emphasis within that standard.

Regarding alignment to the ACCUPLACER framework, ACCUPLACER items in the short version of 45 items covered all three standards—“Arithmetic,” “Elementary algebra,” and “College level math”—with a relatively even distribution, although “College level math” had the lowest percentage of item alignments. NAEP items in the full pool of 164 items also covered “Arithmetic,” “Elementary algebra,” and “College level math,” with a fairly even distribution of approximately one-third of NAEP codable items aligned to each standard, although “Elementary algebra” received somewhat fewer item alignments. Despite these differences in emphasis, however, considering only codable items, the percentages of alignments to each ACCUPLACER standard were relatively evenly distributed in both assessments and similar in distribution across assessments. At the objective level, the distribution of item alignments to objectives was relatively even on both tests, although each assessment was aligned to some objectives to which the other was not.

In summarizing cross-framework alignment, there was somewhat less even distribution of items than observed in within-framework alignment. The majority of items on each test were found to align to objectives on the other test. However, the 105 ACCUPLACER items aligned primarily (90%) to a total of seven out of 24 NAEP goals: three of the six goals from “Number properties and operations” in the NAEP framework, and four of the five goals in “Algebra.” Conversely, the NAEP items from the full pool of 164 items that aligned to the ACCUPLACER framework were distributed fairly evenly across the three ACCUPLACER standards and found to align to 75 ACCUPLACER objectives.

Are there systematic differences in content and complexity between NAEP and ACCUPLACER assessments in their alignment to the NAEP framework and between NAEP and ACCUPLACER assessments in their alignment to the ACCUPLACER framework? Are these differences such that entire mathematics subdomains are missing or not aligned?

Regarding differences in alignment of content, ACCUPLACER items had very limited coverage of measurement, geometry, and data analysis, content that is not included in the ACCUPLACER framework but that is included in the NAEP framework. Many NAEP items assessing these subdomains were found to be uncodable to the ACCUPLACER objectives (20 were rated uncodable by the majority of panelists in each panel). For other NAEP items that were aligned to an ACCUPLACER objective, there were often parts of those items not addressed by the objective. These items were coded as aligned, since they do assess an ACCUPLACER objective, but parts of the items also cover other skills not included in the ACCUPLACER framework.

Regarding differences in alignment of complexity, the items from both tests that aligned to the NAEP standards met the typical depth-of-knowledge (DOK) consistency threshold; that is, the items assessed the objectives at or above the DOK level of the objective. The items from both tests that aligned to the ACCUPLACER standards had somewhat different ranges of DOK. The ACCUPLACER short-version items were divided fairly evenly between Level 1 and Level 2. The NAEP items aligned to the ACCUPLACER framework had a wider range of DOK, with items at Level 1, 2, and 3, and a greater emphasis on Level 2 than was in the ACCUPLACER items.”

The Content Alignment Studies: Summary Discussion for Mathematics

Three content alignment studies were conducted to examine the extent to which

- The content of the NAEP 12th grade mathematics assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing mathematics courses. and
- NAEP 12th grade mathematics test items and scoring criteria are appropriate for obtaining evidence of test takers’ possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing mathematics courses.

For short-hand, this will be referred to as “academic preparedness for college.”

The NAEP 12th grade mathematics framework, test questions, and, for constructed response items, the score category rubrics, were compared with the analogous domain descriptions and test questions for the ACT, SAT, and ACCUPLACER mathematics tests. These three tests are used for college admissions and placement. They are well established and have been used for these purposes for many years by professionals in postsecondary education. The test publishers regularly survey secondary and postsecondary educators about relevant content and have conducted research that supports the validity of the test content for the intended inferences and uses. The underlying assumption is that if the content of the 12th grade NAEP mathematics assessment is similar to the content of these mathematics tests, then the NAEP content is directly related to “academic preparedness for college.”

The ACT study found that “With regard to the ACT to-NAEP comparison...nearly all of the ACT Mathematics domain and College Readiness Standards [are] reflected in the NAEP Mathematics domain, but...a number of the lower-level topics in the ACT Pre-Algebra subdomain were more consistent with Grade 8 NAEP topics.” In the NAEP-to ACT comparison, there was uncertainty about “...whether a number of NAEP content topics—those pertaining to transformations, probability, statistics, and data analysis—are assessed by the ACT...and the degree to which higher-order analytic skills were assessed...and it was the sense of the panel that the ACT Mathematics Test contained few items involving high mathematical complexity, at least as the NAEP defines it.”

The SAT study found similar content in the NAEP 12th grade mathematics assessment and the SAT mathematics test. “At the standard level, the wording of the standards in the two frameworks is very similar. Both the NAEP and SAT frameworks include virtually the same five broad content categories, with SAT combining geometry and measurement into one standard... Although the structures of the two frameworks differ greatly beyond the standard level (including the NAEP framework having three levels while SAT has two), the mathematics areas typically expected of grade 12 students—number and operations, geometry and measurement, data analysis and probability, and algebra—are addressed in somewhat similar proportions... While there is considerable overlap between the two assessments, primarily in the intersection of the NAEP “Algebra” and SAT “Algebra and functions” standards, there are notable differences as well. The SAT items had a somewhat limited range of coverage of the NAEP standards “Measurement,” “Geometry,” and “Data analysis, statistics, and probability,” with several goals receiving few item alignments. Even given the minimal coverage of some of the goals within each NAEP standard by SAT items, however, almost all NAEP items found a match in the SAT framework

The ACCUPLACER study found that “The NAEP and ACCUPLACER assessments both cover certain content traditionally expected of grade 12 students, namely the two content subdomains of number or number operations and algebra...although their respective degrees of alignment and focus in these subdomains vary... the 105 ACCUPLACER items aligned primarily (90%) to a total of seven out of 24 NAEP goals: three of the six goals from “Number properties and operations” in the NAEP framework, and four of the five goals in “Algebra.” Conversely, the NAEP items from the full pool of 164 items that aligned to the ACCUPLACER framework were distributed fairly evenly across the three ACCUPLACER standards and found to align to 75

ACCUPLACER objectives...Regarding differences in alignment of content, ACCUPLACER items had very limited coverage of measurement, geometry, and data analysis, content that is not included in the ACCUPLACER framework but that is included in the NAEP framework. Many NAEP items assessing these subdomains were found to be uncodable to the ACCUPLACER objectives...”

The results across the three studies are consistent. In general, the content of the ACT, SAT, and ACCUPLACER mathematics tests are present in NAEP, but NAEP is generally broader. Alignment between NAEP and the other three respective assessments is substantial, but not perfect; perfect alignment is not expected.

These results support the conclusion that

- The content of the NAEP 12th grade mathematics assessment covers the knowledge and skills needed for college freshmen to be placed into entry-level credit bearing mathematics courses. and
- NAEP 12th grade mathematics test items and scoring criteria are appropriate for obtaining evidence of test takers’ possession of knowledge and skills needed for college freshmen to be placed into entry-level credit-bearing mathematics courses.

2. The NAEP sampling, scaling and statistical procedures yield accurate estimates of the percentage of students scoring at or above a selected cut-score.

The NAEP sampling, scaling, and statistical procedures are widely accepted, well documented (for example, see National Center for Education Statistics, pp. 70-71) and have been periodically evaluated over two decades (for example, see complete list of research conducted by the NAEP Validity Studies Panel at

http://www.air.org/reports-products/index.cfm?fa=viewContent&content_id=890 and

“Evaluation of the National Assessment of Educational Progress: Study Reports” at

<http://www2.ed.gov/rschstat/eval/other/naep/naep-complete.pdf>). Other than issues relating to the comparability among the state-level NAEP samples of inclusion rates of students with disabilities and students who are English language learners (about which the Governing Board and NAEP have taken and continue to take significant action), there is little dispute about the appropriateness of the NAEP sampling, scaling and statistical procedures for estimating the percentage of students scoring at or above a selected cut-score.

This is relevant because the proposed inferences that are the subject of this validity argument are interpretations to add meaning to the Proficient achievement levels for NAEP 12th grade reading and mathematics. The percentages of students at or above each of the NAEP achievement levels (Basic, Proficient, and Advanced) have been estimated and reported regularly, beginning with assessments in 1992. The added meaning being given to the Proficient achievement levels will not affect in any way the accuracy of the estimates of the percentages of students scoring at or above the Proficient cut-score.

3. Scores on 12th grade NAEP reading and mathematics assessments provide accurate estimates of academic preparedness for entry level credit-bearing college courses.

- **Performance on the 12th grade NAEP reading and mathematics assessments is related to other indicators or criteria of academic preparedness for placement into entry-level credit-bearing college courses.**

In addition to examining the overlap in test content between NAEP and the tests for college admission and placement, the Governing Board determined that it would be relevant and important to examine how performance on NAEP relates to performance on the SAT and ACT, including the college readiness benchmarks associated with these testing programs. There are several data sources for the analyses: the NAEP/SAT linking studies (see report at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/SAT-NAEP_Linking_Study.pdf), the Florida longitudinal study (see report at http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/Florida_Statistical_Study.pdf), the 2005 and 2009 NAEP High School Transcript Studies, and the Governing Board's survey of postsecondary education institutions' use of tests and the cut-scores on those tests for determining whether incoming students need remedial instruction in reading and mathematics (Fields and Parsad).

Indicators: College Board and ACT College Readiness Benchmarks

The College Board and ACT, Inc. have established college readiness benchmarks for the SAT and the ACT in a number of subjects tested, including reading and mathematics. The SAT College Readiness Benchmark for critical reading and mathematics is a score of 500 on the respective tests. According to the College Board's research, a score of 500 predicts, with a .65 probability, a first-year GPA of B- or better. The ACT College Readiness Benchmark for reading is a score of 21. According to ACT's research, a score of 21 predicts, with a .50 probability, a grade of B or better (or .75 probability of a C or better) in first year courses requiring college reading, such as history and the social sciences. A score of 22 on the ACT mathematics tests predicts a .50 probability of a grade of B or better in a first-year mathematics course, or a .75 probability of a grade of C or better. The College Board and ACT research is based on the first-year outcomes of their respective test takers.

Indicators: First Year GPA of B- or Better and Remedial/non-Remedial Placement

The Governing Board has a partnership with the state of Florida as a part of the Board's program of preparedness research. Florida was one of 11 states that volunteered to provide state-representative samples of 12th grade students for the 2009 NAEP reading and mathematics assessments. Under the partnership, the Florida 12th grade sample is being followed through the postsecondary years via the highly developed Florida longitudinal education data system. For comparability with the SAT College Readiness Benchmarks, the Governing Board analyzed the Florida data to determine the average score and interquartile range for the NAEP test takers with a first year GPA of B- or better. In addition, the Governing Board analyzed the Florida data to determine the average score and interquartile range for the NAEP test takers who were and who were not placed into remedial reading or remedial mathematics in their first year of college.

Analysis of Results for Mathematics

The Governing Board's program of preparedness research included a statistical linking study between the NAEP 12th grade mathematics assessment and the SAT mathematics test. Through a partnership with the College Board, the mathematics SAT scores of students who took the 12th grade NAEP mathematics assessment in 2009 were obtained and analyzed.

A correlation of .91 was found for performance on the NAEP 12th grade mathematics assessment and the SAT mathematics test. This high correlation, together with the substantial overlap in test content found in the content alignment studies between the NAEP and SAT mathematics tests, supports inferences about NAEP performance in relation to SAT performance. Of particular interest, is how performance on NAEP relates to the SAT College Readiness Benchmark for mathematics (i.e., a score on the SAT mathematics test of 500 or better). The SAT benchmark provides "an indication of college readiness at a typical college (College Board)." This is consistent with the Governing Board's definition of academic preparedness.

Academic preparedness for college refers to the reading and mathematics knowledge and skills needed to qualify for placement into entry-level, credit-bearing, non-remedial courses that meet general education degree requirements in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

The analysis of the mathematics indicators is displayed in Figure 1. A consistent pattern is evident across studies and across time. This consistent pattern supports the inferences that 12th grade students scoring at or above Proficient on the 12th grade NAEP Mathematics Assessment are

- likely to be academically prepared for first year college mathematics courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in mathematics in college.

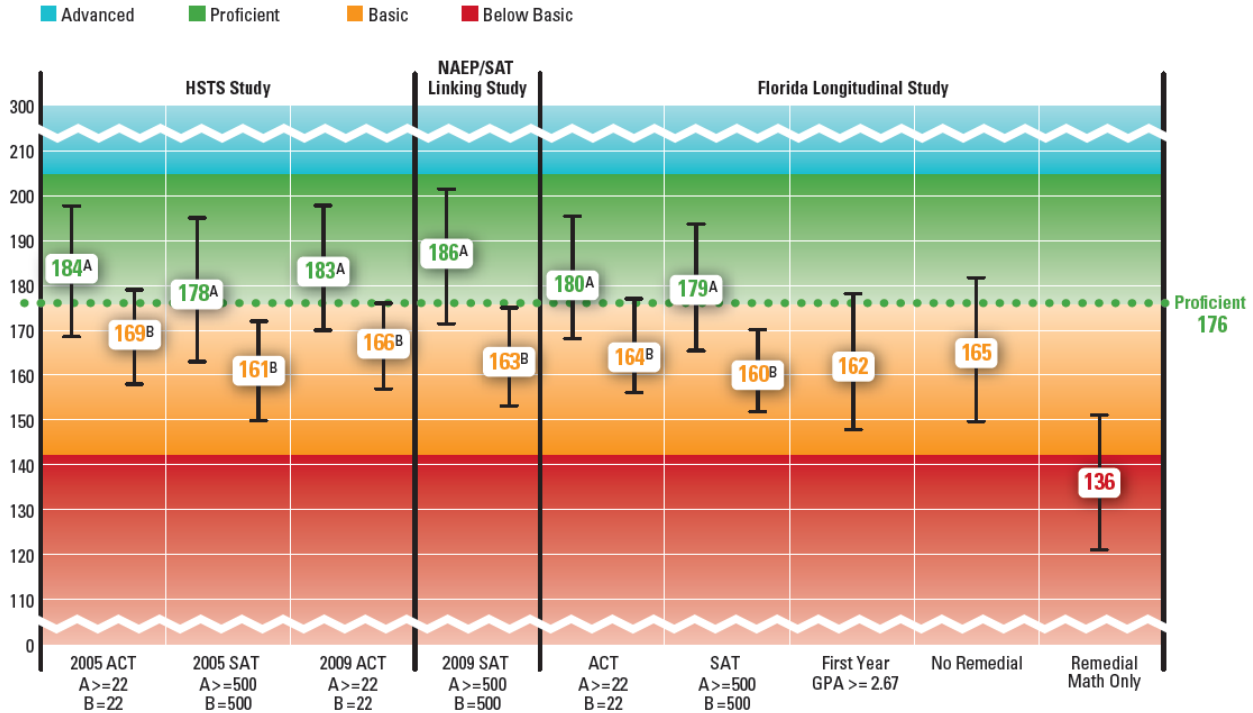
The average NAEP mathematics scores for 12th grade students scoring **at** the SAT College Readiness Benchmark for mathematics are compared first for the two national studies: the 2005 High School Transcript Study (HSTS) and the 2009 NAEP/SAT Linking Study (NSLS). The average scores are 161 and 163 respectively. These scores are somewhat below the cut-score for Proficient, which is 176 on the NAEP 12th grade mathematics scale. If the interpretation is made that students scoring at 163 on the NAEP 12th grade mathematics scale have a .65 chance of attaining a first-year GPA of B- or better, and this score is below Proficient, then it follows that students scoring at or above Proficient have increasingly higher probabilities of attaining a first-year GPA of B- or better.

It also means that many students who score in the mid-range and above of the Basic achievement level on the 12th-grade NAEP mathematics assessment may be academically prepared for college. For example, considering the NSLS results, students with 2009 SAT scores of 500 have an average NAEP score of 163, with an interquartile range of 153 to 175. A substantial percentage of the NAEP scores for these students are in this range. Similar results are observed for other measures in Table 1.

Figure 1

NAEP 12th Grade Preparedness Research: Mathematics

Average Scores and Inter-quartile Ranges For Selected Variables, 2005 High School Transcript Study, 2009 High School Transcript Study, 2009 NAEP/SAT Linking Study, 2009 Florida Longitudinal Study



Thus, while getting a NAEP score in the Proficient category provides a very strong indication that a student is academically prepared for college, students in the upper region of the Basic range are also likely to be academically prepared for college, but with a lower probability.

These results are confirmed by the Florida longitudinal study results (FLS). The average NAEP mathematics score for the 12th grade Florida NAEP test takers who scored at the SAT College Readiness Benchmark of 500 was 160, somewhat below the Proficient cut score, like the 2009 NSLS results and the 2005 and HSTS results.

Another analysis examines the average scores and interquartile ranges for students scoring **at or above** the SAT College Readiness Benchmark for mathematics from the 2005 HSTS, 2009 NSLS, and 2009 FLS. In all three cases, the interquartile ranges fall around Proficient and overlap to a high degree.

As discussed previously, the ACT College Readiness Benchmark for mathematics is defined differently than the SAT College Readiness Benchmark for mathematics. However, it is noteworthy that even with this different definition, the results from the 2005 HSTS, 2009 HSTS, and 2009 FLS analyses are very similar to the results for the SAT.

Taken together, these results support the inference that students scoring at or above Proficient on the NAEP 12th grade mathematics scale are likely to be academically prepared for entry-level credit-bearing mathematics courses and to attain a first-year GPA of B- or better.

To answer the question, what is the relationship between performance on NAEP and student outcomes, we look to the Florida longitudinal study results. First we examine the average NAEP mathematics score for the 12th grade Florida NAEP test takers who actually attained a first-year GPA of B- or better. The average NAEP score for these students was 162, somewhat below the Proficient cut point. This is consistent with the SAT College Readiness Benchmark analyses and further supports the inference that students at or above Proficient are likely to be academically prepared and attain a first-year GPA of B- or better. It follows, of course, that students who are academically prepared will not require remedial courses.

Thus, another outcome of interest is placement of entry-level students into remedial college courses versus non-remedial credit-bearing courses. Here again, we look to the FLS as a data source. The average NAEP mathematics score for the Florida NAEP test-takers not placed into remedial courses was 165, somewhat below the NAEP Proficient cut-score of 176. The average score for Florida students placed into remedial mathematics was 136, which is in the range below Basic. These results lend support, together with the SAT and ACT analyses, to the inference that students scoring at or above Proficient are not likely to need remedial courses in mathematics.

Analysis of Results for Reading

The Governing Board's program of preparedness research included a statistical linking study between the NAEP 12th grade reading assessment and the SAT critical reading test. Through a partnership with the College Board, the SAT critical reading scores of students who took the 12th grade NAEP reading assessment in 2009 were obtained and analyzed.

A correlation of .74 was found for performance on the NAEP 12th grade reading assessment and the SAT critical reading test. This is a substantial correlation. While it may not be high enough to predict the performance of individual students from one test to another, it is sufficient to support the group-level inferences reported by NAEP. This, together with the substantial overlap in test content found in the content alignment studies between the NAEP and SAT reading tests, supports inferences about NAEP performance in relation to SAT performance. Of particular interest, is how performance on NAEP relates to the SAT College Readiness Benchmark for reading (i.e., a score on the SAT mathematics test of 500 or better). The SAT benchmark provides "an indication of college readiness at a typical college (College Board)." This is consistent with the Governing Board's definition of academic preparedness.

Academic preparedness for college refers to the reading and mathematics knowledge and skills needed to qualify for placement into entry-level, credit-bearing, non-remedial courses that meet general education degree requirements in broad access 4-year institutions and, for 2-year institutions, for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

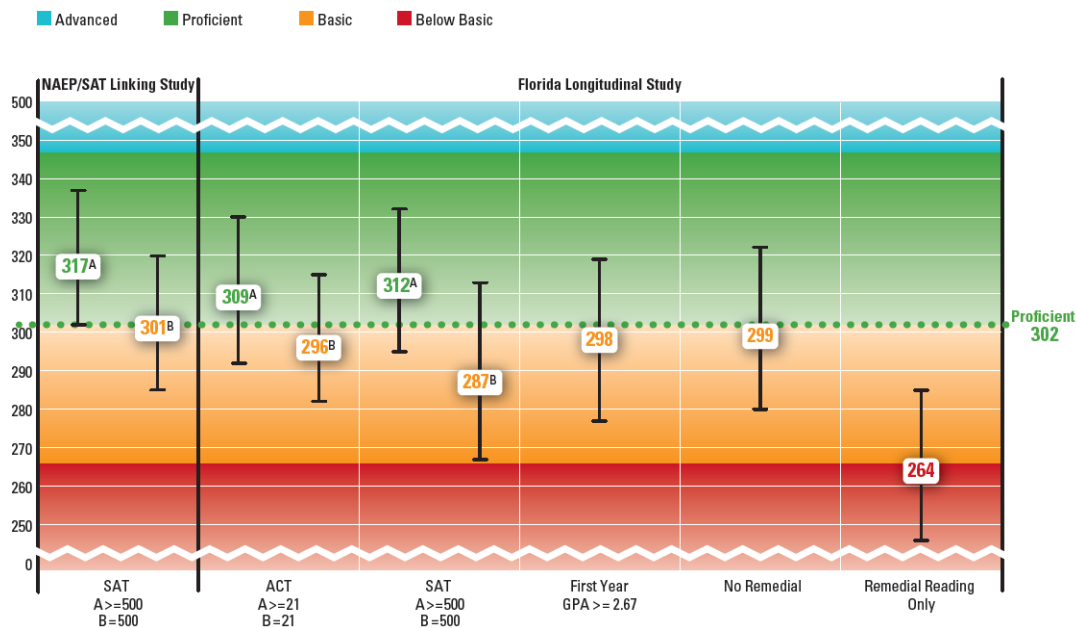
The analysis of the reading indicators is displayed in Figure 2. A consistent pattern is evident across studies. This consistent pattern supports the inferences that 12th grade students scoring at or above Proficient on the 12th grade NAEP Reading Assessment are

- likely to be academically prepared for first year courses requiring college level reading,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading in college.

Figure 2

NAEP 12th Grade Preparedness Research: Reading

Average Scores and Inter-quartile Ranges For Selected Variables for the 2009 NAEP SAT Linking Study and 2009 Florida Longitudinal Study



The average NAEP reading score for 12th grade students scoring at the SAT College Readiness Benchmark for reading is examined first for the national 2009 NSLS. The average score is 301, just below the cut score for Proficient, which is 302 on the NAEP 12th grade reading scale. If the interpretation is made that students scoring at 301 on the NAEP 12th grade reading scale have a .65 probability of attaining a first-year GPA of B- or better, and this score is below Proficient, then it follows that students scoring at or above Proficient have increasingly higher probabilities of attaining a first-year GPA of B- or better.

These results are confirmed by the Florida longitudinal study results (FLS). The average NAEP reading score for the 12th grade Florida NAEP test takers who scored at the SAT College Readiness Benchmark of 500 was 287, somewhat below the Proficient cut score, like the 2009 NSLS results.

Another analysis examines the average scores and interquartile ranges for students scoring at or above the SAT College Readiness Benchmark for reading from the 2009 NSLS and 2009 FLS. In both cases, the interquartile ranges fall around Proficient and overlap to a high degree.

As discussed previously, the ACT College Readiness Benchmark for mathematics is defined differently than the SAT College Readiness Benchmark for mathematics. However, it is noteworthy that even with this different definition, the results from the 2009 FLS analysis is very similar to the results for the SAT.

Taken together, these results support the inference that students scoring at or above Proficient on the NAEP 12th grade reading scale are likely to be academically prepared for entry-level credit-bearing courses requiring college level reading and to attain a first-year GPA of B- or better.

To answer the question, what is the relationship between performance on NAEP and student outcomes, we look to the Florida longitudinal study results. First we examine the average NAEP reading score for the 12th grade Florida NAEP test takers who actually attained a first-year GPA of B- or better. The average NAEP score for these students was 298, just below the Proficient cut point. This is consistent with the SAT College Readiness Benchmark analysis and further supports the inference that students at or above Proficient are likely to be academically prepared and attain a first-year GPA of B- or better. It follows, of course, that students who are academically prepared will not require remedial courses.

Thus, another outcome of interest is placement of entry-level students into remedial versus non-remedial credit-bearing courses. Here again, we look to the FLS as a data source. The average NAEP reading score for the Florida NAEP test-takers not placed into remedial courses was 299, again, just below the NAEP Proficient cut-score of 302. This lends support, together with the SAT and ACT analyses, to the inference that students scoring at or above Proficient are not likely to need remedial courses in reading.

➤ **Empirical indicators of student engagement do not support the assertion that NAEP 12th grade test-takers are not motivated.**

A recurring question about NAEP in general is whether student achievement is underestimated because the test-takers receive no test results back and bear no consequences for their performance. It is a relevant and legitimate question potentially affecting the accuracy of NAEP estimates.

The question is asked with special skepticism about 12th grade NAEP test-takers: will “test-wise” high school seniors in the last semester of their K-12 experience, knowing that the results will not affect their grades or future opportunities, apply the same effort that they would to tests that do come with high stakes for them? Will they be “motivated” when they sit for NAEP?

Associated with this question is the assertion that 12th grade NAEP test takers are not motivated. This assertion has been supported by anecdote, the logic of the apparent incentives inherent in the NAEP 12th grade testing situation, or common wisdom, but it has not been supported by empirical evidence. Research on this topic has been inconclusive. Similarly, information from NAEP background questions has been inconclusive. For example, NAEP background questions asking 12th graders whether they tried hard when taking NAEP are consistently associated with

higher average scores for the students who say they didn't try hard and lower average scores for students who say they did try hard. Perhaps student proficiency and effort are being conflated in the analysis of the responses, but the data provide no evidence that achievement is underestimated by NAEP because of a lack of student motivation.

In 2009, the 12th grade students who took the NAEP reading and mathematics assessments answered 95% of the test questions, including the constructed response items that require students to do much more than merely fill in a bubble on a multiple choice answer sheet. With respect to the multiple choice questions, there was little if any evidence of "Christmas tree" or random responses to the questions, which would have been a sign that students were not seriously engaged in the test-taking task.

The correlations between performance on the high stakes SAT and the low stakes NAEP are additional evidence to consider. The correlation was .91 in comparing mathematics performance and .74 for reading. While these substantial correlations do not prove that the 12th grade NAEP test takers were motivated, they do not support the assertion that they are not motivated. Although it is logical to assume that the 12th grade students sitting for both tests may not have taken low stakes NAEP as seriously as the high stakes SAT to some degree, it is not possible to determine if this is true. And of course, a decrease of all of the NAEP scores by any consistent number of points would still yield the same correlations. However, the correlations do suggest that any diminution in motivation that might be present is not diminishing the effectiveness of the NAEP scores as indicators of academic preparedness for college.

4. The proposed test uses are appropriate and consequences are commensurate with intended uses.

The National Assessment of Educational Progress is an independent monitor of student academic achievement in the United States. It reports on achievement at specific points in time and trends in achievement over time. NAEP reports to the public, national and state policymakers, and education leaders. It assesses student achievement at grades 4, 8, and 12 in important subjects. NAEP is used to compare performance across states and for 21 urban school districts. NAEP results are reported by gender, race/ethnicity, poverty status, and for students with disabilities and students who are English language learners.

The audiences and the uses of NAEP are well established. They will not change as a result of the added meaning to the NAEP 12th grade Proficient achievement levels for reading and mathematics afforded by the inferences proposed in this validity argument. However, providing familiar external referents for performance on 12th grade NAEP in relation to Proficient performance will greatly enhance the understanding of NAEP results by its audiences.

Currently, there are either no or very low stakes consequences associated with the use of NAEP results. NAEP is not used as a basis for evaluating or diagnosing individual students, classroom or school performance, the effectiveness of individual teachers or administrators, or for any other accountability purpose. This will not change with the added meaning to the NAEP 12th grade

Proficient achievement levels for reading and mathematics afforded by the inferences proposed in this validity argument.

While the uses and consequences of NAEP will not change, the added meaning to NAEP Proficient at the 12th grade brings with it the potential for misinterpretation. These were discussed in detail on pages 5-6 above, and will be summarized here. NAEP reports should include text explaining the limitations on interpretation and other caveats that follow.

False Negatives

Some proportion of 12th grade students scoring below Proficient on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first-year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college,

but with a lower probability than those at or above Proficient.

Not a Preparedness Standard

The proposed inferences are not intended to represent or be used as standards for minimal academic preparedness for college.

Academically Prepared for College

The proposed inferences are intended to apply to placement policies affecting the typical degree-seeking entry-level college student at the typical college, not the admission policies. Thus, “academically prepared for first year college courses” refers to the reading and mathematics knowledge and skills needed for placement into entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, the general policies for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institutions.

The proposed inferences are not intended to reflect academic requirements for highly selective postsecondary institutions; to the additional academic requirements for specific majors or pre-professional programs, such as mathematics, engineering, or medicine; or to academic requirements applicable to entry into certificate or diploma programs for job training or professional development in postsecondary institutions.

Data Limitations

Although the preparedness research studies are comprehensive and the results consistent and mutually confirming, for reading they are limited to one year for data at the national level and to one state-based longitudinal study. For mathematics, there are two separate years of data at the national level and one state-based longitudinal study. Therefore, more evidence exists to support the plausibility of inferences related to mathematics than to reading.

Preparedness for Job Training

The completed research with respect to academic preparedness for job training does not support conclusions relative to the NAEP scale and will not be addressed at this time.

Summary and Conclusion

The National Assessment Governing Board decided to determine the feasibility of transforming NAEP into a measure of academic preparedness for college. Consequently, the Governing Board made changes to the NAEP 12th grade reading and mathematics frameworks with the explicit purpose of measuring academic preparedness for college. The Governing Board conducted research that established a high degree of overlap between the content of the NAEP 12th grade reading and mathematics assessments and the content of widely used college admissions and placement tests.

Through a partnership with the College Board, performance on 12th grade NAEP was compared with performance on the SAT mathematics and critical reading assessments, with correlations of .91 and .74 respectively. Analyses of these data examined the average NAEP scores and interquartile ranges for students scoring “at” and “at or above” the College Board College Readiness Benchmarks for reading and mathematics. Similar analyses were conducted using data from the 2005 and 2009 NAEP High School Transcript Studies, using the college readiness benchmarks developed by ACT and by the College Board. A longitudinal study was conducted in partnership with the Florida Department of Education, following the 12th grade students in the state NAEP sample into postsecondary employing Florida’s longitudinal data base. The average NAEP scores and interquartile ranges were calculated for the Florida students in relation to the ACT or SAT college readiness benchmarks, whether they achieved a first-year GPA of B- or better, and whether they were placed into a remedial course in their first year of college. The results of these analyses were consistent across studies and across years. In addition, indicators of the engagement in the NAEP test taking of 12th grade students in 2009 do not lend support to the assertion that NAEP 12th grade results are underestimates due to a lack of student motivation.

That the NAEP sampling, scaling and statistical procedures yield accurate estimates of the percentage of students scoring at or above a selected cut-score (i.e., NAEP achievement level) is well established as a result of numerous validity studies and evaluations.

Thus, the NAEP 12th grade preparedness research results support the inferences that students scoring at or above the Proficient achievement level on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first year college courses
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading in college.

A substantial percentage of students whose scores are in the range between Basic and Proficient are likely to be academically prepared for college, but with a lower probability.

Including these inferences in NAEP 12th grade reports will add meaning to the interpretation of the NAEP achievement levels. However, steps must be taken to avoid potential misinterpretation. NAEP reports using these inferences must also include the limitations on interpretation and caveats described previously in this validity argument. In addition, the reports should explain the rationale for NAEP reporting on academic preparedness and describe appropriate and inappropriate uses of the results.

References

ACT Technical Manual; http://www.act.org/aap/pdf/ACT_Technical_Manual.pdf

ACCUPLACER on-line technical manual;

http://isp.southtexascollege.edu/ras/research/pdf/ACCUPLACER_OnLine_Technical_Manual.pdf

AERA/APA/NCME *Standards for Educational and Psychological Testing* (1999)

Allen, Jeff, Sconing, Jim (2005). *Using ACT Assessment Scores to Set Benchmarks for College Readiness* (ACT Research Series 2005-3). Iowa City, IA: ACT, Inc,

Bozick, R., and Lauff, E. (2007). *Education Longitudinal Study of 2002 (ELS:2002): A First Look at the Initial Postsecondary Experiences of the Sophomore Class of 2002* (NCES 2008-308). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

College Board. *The SAT® College and Career Readiness Benchmark User Guidelines*; http://media.collegeboard.com/digitalServices/pdf/sat/12b_6661_SAT_Benchmarks_PR_120914.pdf

Conley, D.T. (2007). *Toward a More Comprehensive Conception of College Readiness*. Eugene, OR: Educational Policy Improvement Center.

Fields, R. & Parsad, B. (2012). *Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011*. Washington, DC: National Assessment Governing Board.

Kane, Michael T. (2013, Spring). *Validating the Interpretations and Uses of Test Scores*. Journal of Educational Measurement, Vol. 50, No. 1, pp. 1-73.

Kim, Y.K., Wiley, A., and Packman, S., *National Curriculum Survey on English and Mathematics*, College Board Research Report 2011-13 (New York: The College Board, 2011) <http://professionals.collegeboard.com/data-reports-research/cb/RR2011-13>

National Assessment Governing Board. (2008, September). *Mathematics Framework for the 2009 National Assessment of Educational Progress*. Washington, DC: Author.

National Assessment Governing Board. (2008, September). *Reading Framework for the 2009 National Assessment of Educational Progress*. Washington, DC: Author

National Assessment Governing Board. (2009, June). *Making New Links: 12th Grade and Beyond*. Washington, DC: Technical Panel on 12th Grade Preparedness Research.

National Center for Education Statistics (2010). *The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results* (NCES 2011-455). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Commission on NAEP 12th Grade Assessment and Reporting. (2004, March 5). *12th Grade Student Achievement in America: A New Vision for NAEP*. Washington, DC: Author.

Ross, T., Kena, G., Rathbun, A., KewalRamani, A., Zhang, J., Kristapovich, P., and Manning, E. (2012). *Higher Education: Gaps in Access and Persistence Study* (NCES 2012-046). U.S. Department of Education, National Center for Education Statistics. Washington, DC: Government Printing Office.

WestEd ACCUPLACER Mathematics Report;

http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Mathematics_Content_Comparison.pdf.

WestEd ACCUPLACER Reading Report;

http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/ACCUPLACER-NAEP_Reading_Content_Comparison.pdf.

WestEd SAT Mathematics Report; http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Mathematics_Content_Comparison.pdf.

WestEd SAT Reading Report; http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/content-alignment/SAT-NAEP_Reading_Content_Comparison.pdf.

Wyatt, Jeffrey, Kobrin, Jennifer, Wiley, Andrew, Camara, Wayne J., and Proestler, Nina (2011). *SAT Benchmarks: Development of a College Readiness Benchmark and its Relationship to Secondary and Postsecondary School Performance (Research Report 2011-5)*. New York, NY: College Board.

<http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-5-sat-college-readiness-benchmark-secondary-performance.pdf>

| 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.**

