# National Assessment Governing Board
## Assessment Development Committee

**November 30, 2012**  
**9:30 am - 12:00 pm**

**AGENDA**

<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
<th>Speaker/Contact</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 9:40 am</td>
<td>Welcome, Introductions, and Agenda Overview</td>
<td><em>Alan Friedman, Chair</em></td>
<td></td>
</tr>
<tr>
<td>9:40 – 10:10 am</td>
<td>Plans for Reporting Information from the 2011 and 2012 Computer-Based NAEP Writing Administrations</td>
<td><em>Arnold Goldstein, NCES</em></td>
<td>Attachment A</td>
</tr>
<tr>
<td>10:10 – 10:40 am</td>
<td>Update on NAEP Technology and Engineering Literacy (TEL) Assessment</td>
<td><em>William Ward, NCES</em></td>
<td>Attachment B</td>
</tr>
</tbody>
</table>
| 10:40 am – 12:00 pm | **Closed Session**  
Review of Science Interactive Computer Tasks (ICTs) for the 2014 Pilot Test | *Andrew Latham, ETS*       | Attachment C                              
  |                  | Secure Materials Sent Under Separate Cover                          |                            |
| Information Item | NAEP Item Review Schedule                                           |                            | Attachment D                              |
NAEP Computer Based-Based Writing Assessment: Grade 4 Pilot

The 2011 NAEP Writing Framework specifies that,

“... In 2011 and beyond, the NAEP Writing Assessment at grades 8 and 12 will assess computer-based writing. For the purposes of the 2011 NAEP Writing Assessment, “computer-based writing” means that students compose and construct their responses using word processing software on a computer, with the option to use commonly available tools.

At grade 4, a computer-based assessment was previously impractical because of time constraints for computer instruction, the unequal availability of technology in elementary schools, and elementary school students’ current limited keyboarding proficiency. However, it is recommended that the assessment at grade 4 become computer-based during the tenure of the 2011 NAEP Writing Framework. ...”


Under this guidance, NCES conducted a computer-based grade 4 pilot writing assessment in the spring of 2012. Over 13,000 students from more than 500 schools participated in the pilot.

Data collected during the pilot will provide critical guidance and insight into 4th grade students’ ability to write on the computer, the development of computer based tasks and platforms, and the preparation, administration, and scoring of 4th grade writing computer-based assessments.

The presentation to the National Assessment Governing Board’s Assessment Development Committee will discuss plans for reporting results from the grade 4 writing pilot and lessons learned.
While NCES is still considering the most effective way of providing this information to the public, we are considering a series of short reports that would consider the following topics:

- Do 4th grade students have the skills and abilities required to respond to writing prompts on a computer-based assessment?

- Are there any limitations to students’ computer skills that have an effect on the quality of their responses, and can scorers be trained to score reliably, given these limitations?

- What have we learned in the design and development process that can inform future assessments?

- Are there opportunities to report more and different information about what students know and can do related to the data we can collect on students’ actions on the computer?
Update on the NAEP Technology and Engineering Literacy (TEL) Assessment

1. Information and Outreach Activities
A number of materials have been developed to provide schools, parents, and students with information about the TEL assessment. A description of how and when these contacts take place will be reviewed. Outreach material developed specifically for the TEL assessment will be presented, including the materials described below.

**Technology and Engineering Literacy Assessment Fact Sheet:** this one page fact sheet was sent to schools in August and again in January. It provides specific information about the TEL assessment and answers questions such as what is TEL and why is TEL important for today’s students. It also describes how TEL is assessed and what schools and students should expect.

**NAEP In Your School – Grade 8 TEL:** this one page document is sent to school principals in June. It gives a brief overview to the principal about what is involved in the TEL assessment and the main responsibilities of the NAEP state coordinator, principal, school coordinator, and NAEP field staff.

**NAEP In Your School – School Coordinator Responsibilities – TEL:** this one page document is sent in August to the school coordinator, who is the liaison for all NAEP activities at the school. This document provides detailed information specific to the responsibilities of the NAEP school coordinator. It also describes ways to promote NAEP with school staff and students.

**TEL Video:** NCES is currently developing a short, fully-animated video for the upcoming TEL assessment. It will aim to explain what Technology and Engineering Literacy actually means, what the NAEP TEL assessment is, what skills it will assess, and why it is important in today’s educational landscape. While it will initially be used by school personnel and NAEP state coordinators as a resource for schools and students selected for the upcoming grade 8 pilot, it will also be geared towards the general public and will be housed on the NCES website.

**TEL Tutorial:** the TEL tutorial is the student’s first experience with the TEL assessment and interactively shows the students the type of items they may encounter during the assessment. It also informs students how to use various tools in the program. The tutorial will be posted on the NAEP web site for the general public to view.

**TEL is coming in 2013!** this postcard is sent in the January preassessment packet and points schools to the TEL page on the NAEP web site. It lists the various resources on the web site, including information about how TEL was created, a link to the tutorial and TEL video, and links to the student survey questions.

2. Update on TEL Development Work
A brief update will be presented to the ADC on November 30 on final TEL development work following the ADC's review of TEL tasks and items in August 2012.
To what extent can young people analyze the pros and cons of a proposal to develop a new source of energy? Can students use the Internet to find and summarize information in order to solve a problem? Do students understand how and why new technologies are developed to suit human needs and wants?

Technology and engineering have become critical components of 21st-century life. For generations students have been taught about technology and have been instructed on how to use various technological devices. However, there are currently no standardized, nationally representative assessments to provide evidence of what students know about technology and engineering; the role technology and engineering plays in our lives; and the extent to which students can use technologies and understand how engineers design and develop them.

The questions listed above are just a few examples of the types of questions the National Assessment of Educational Progress (NAEP) technology and engineering literacy (TEL) assessment will aim to answer. In 2013, a nationally representative sample of grade 8 students will participate in the TEL pilot. The results from this pilot will be used to prepare for the first-ever national TEL assessment of eighth-graders in 2014.

What is TEL?

The 2014 NAEP Technology and Engineering Literacy Framework broadly defines technology and engineering literacy as the capacity to use, understand, and evaluate technology as well as to understand technological principles and strategies needed to develop solutions and achieve goals.

This framework is the guide for the development of the TEL assessment and defines what students should know and be able to do with technology. The assessment is designed to assess three interconnected areas of technology and engineering literacy:

- Technology and Society
- Design and Systems
- Information and Communication Technology

For more information about NAEP, visit: http://nces.ed.gov/nationsreportcard
How is TEL assessed?

Allowing students to demonstrate the wide range of knowledge and skills detailed in the three TEL assessment areas will require a departure from the typical assessment designs used in other NAEP subjects. Students will be asked to perform a variety of computer-based tasks to solve problems within scenarios that reflect realistic situations. These scenario-based tasks are an innovative component of NAEP. In addition to scenario-based tasks, TEL will also rely on short answer and multiple-choice questions to measure students’ knowledge and skills.

Why is TEL important for today’s students?

The skills for technology and engineering literacy are increasingly being taught through a wide range of school coursework. This includes contemporary science, technology, engineering, and mathematics (STEM) education, as well as subjects such as social studies and language arts. These courses include instruction on the use of computers and information technology to complete school assignments, lessons that examine the role of technology in society, and information on engineering design. Information technologies are also essential tools in workplace and practical contexts.

Because of this growing importance of technology and engineering in the educational landscape, an assessment of technology and engineering literacy is an important addition to NAEP.

The TEL assessment opens the door to understanding what students know about technology and engineering, in the same way that NAEP already assesses their knowledge and capabilities in reading, mathematics, science, and other subjects.

What should schools expect?

The TEL pilot in 2013 will be administered by NAEP representatives who will work with school staff to organize the TEL assessment activities. NAEP representatives will bring necessary materials, including laptop computers and earbuds, to the school on assessment day. About 30 students in each school will be selected to participate, and the assessment will be administered in two sequential sessions of approximately 15 students each.

What should students expect?

Before the assessment begins, students will be shown a tutorial that will help them become familiar with the interface and how to use the program. It will take approximately 120 minutes for students to complete the assessment.
WHAT IS NAEP?

The National Assessment of Educational Progress (NAEP) is an essential measurement of student achievement in the United States.

- First administered in 1969, NAEP is the largest continuing and nationally representative assessment of what our nation’s students know and can do in core subjects such as mathematics, reading, science, and writing.
- The schools and students participating in NAEP represent other schools and students across the country.
- NAEP is considered the gold standard of assessments because of its high technical quality. From developing frameworks and questions to the reporting of results, NAEP represents the best thinking of assessment and content specialists, state education staff, and teachers from around the nation.
- NAEP monitors academic progress over time and reports on student achievement nationally.

NAEP will be administered to a sample of eighth-grade students in your school between January 28 and March 8, 2013. For the first time, eighth-grade students will participate in a computer-based technology and engineering literacy (TEL) pilot assessment. The results from this pilot will be used to inform the first-ever national TEL assessment of eighth-graders in 2014. This fact sheet will help you prepare your school and students for participation.

What is involved?

The TEL pilot will measure students’ capacity to use, understand, and evaluate technology, as well as to understand technological principles and strategies. Students will spend about 120 minutes completing the assessment, which includes problem-solving tasks based on interactive scenarios reflecting realistic situations, and multiple-choice and short-answer questions. Students will also be asked to complete a questionnaire that aims to get a better understanding of their opportunities to learn about technology and engineering both inside and outside the classroom.

Principals will also be asked to complete a questionnaire. This questionnaire is being piloted to prepare for the collection of contextual information for future assessments. Additional information will be collected about how selected students with disabilities and English language learners (SD/ELL) might participate in the assessment.

NAEP staff members will bring all necessary materials, including laptop computers and earbuds, to the school on assessment day. Schools will only need to provide rooms, desks or tables, and electrical outlets; schools do not have to provide Internet access.

How many students are assessed?

Nationally, approximately 15,000 students will be assessed in the grade 8 TEL pilot. Thirty students at each school will be selected to participate. The pilot assessment will be administered in two sequential sessions of 15 students each.
Who is responsible for coordinating and administering NAEP?

Your NAEP State Coordinator, NAEP representatives, and school staff will work together to coordinate and administer the assessment.

A staff person in your school will need to be assigned to serve as the school coordinator, who will be the primary contact in the school for the assessment.

The NAEP State Coordinator works at your state Department of Education and will be responsible for:

- Working with schools to confirm the assessment date;
- Communicating with principals to convey the facts about NAEP and the importance of student participation;
- Providing schools with instructions for preparing a list of eligible students (if required) and information about notifying parents of participating students;
- Providing guidance for the inclusion of SD/ELL students; and
- Responding to questions from the school community throughout the assessment period.

NAEP representatives are employed by a U.S. Department of Education contractor to work directly with schools and will be responsible for:

- Selecting a random sample of students from the school list of eligible students;
- Sending preassessment materials to the school coordinator;
- Visiting the school coordinator shortly after the preassessment materials are received to finalize assessment arrangements;
- Bringing all assessment materials, including laptop computers, to the school on the scheduled day; and
- Conducting the assessment.

The school coordinator will be responsible for:

- Confirming with the NAEP State Coordinator that the suggested assessment date is convenient for the school;
- Providing NAEP with an electronic file of students in the selected grade, if requested;
- Registering for and using the MySchool website;
- Informing parents of the assessment (the NAEP State Coordinator will provide additional information about how this should be accomplished);
- Receiving the preassessment materials and making final preparations for the assessment;
- Meeting with the NAEP representative during the scheduled preassessment visit; and
- Ensuring that there is a plan with school staff to achieve high student participation in the assessment.

Detailed information about the school coordinator’s responsibilities will be sent at the beginning of the school year.

Each principal will be responsible for:

- Assigning a staff member to serve as school coordinator;
- Including the NAEP assessment date on the school calendar;
- Empowering the designated school coordinator to work with the NAEP representative and the NAEP State Coordinator to prepare for the assessment; and
- Informing school staff and students about NAEP and why student participation is critically important.
For the first time, eighth-grade students will participate in a computer-based technology and engineering literacy (TEL) pilot assessment. As the school coordinator, you are the liaison for all NAEP assessment activities in your school. Thank you in advance for your help preparing for this important assessment!

In the fall, you will be responsible for the following:

Registering for and using the MySchool website.
The MySchool website provides information for schools about what to expect throughout the NAEP assessment process. It is used to collect information about your school and provides documents that you can download and customize for the NAEP assessment. Multiple school staff may register to access the site. To register for MySchool, go to www.mynaep.com and complete the registration form using the MySchool registration ID provided by your NAEP State Coordinator.

Completing and submitting school information.
Go to the MySchool website and click on “Provide School Information” to enter and submit information about your school. Providing up-to-date information about your school ensures that materials can be prepared for the assessment.

Providing the NAEP State Coordinator with a list of eighth-grade students (if requested).
NAEP requires a complete list of eighth-grade students so that a random sample of students can be selected to participate in the assessment and demographic information about these students can be collected. This list is usually submitted electronically and may be prepared by the school, district, or state. Your NAEP State Coordinator will inform you if and when you need to provide this list. Student names will always be kept confidential, and individual student responses or scores on NAEP are never reported.

Before the assessment date, you will be responsible for the following:

Receiving the Preassessment Packet and beginning final preparations for the assessment.
In early January, you will receive a Preassessment Packet that contains the list of students selected to participate as well as instructions, forms, and resources to help you prepare for NAEP. After you receive the packet, your school’s NAEP representative will call you to confirm the date and time for the preassessment visit and to answer any questions about the materials. It is important that you review all the items and complete the necessary tasks described in the Preassessment Packet prior to the visit.
Promoting the importance of NAEP with school staff.
Teachers are essential for motivating students to do their best on NAEP. Here are some suggestions on how to gain teacher support:

- Show the 5-minute NAEP video, *Introducing NAEP to Teachers*, at a faculty meeting. This video can be accessed at [http://nces.ed.gov/nationsreportcard/about/schools.asp](http://nces.ed.gov/nationsreportcard/about/schools.asp).
- Place copies of NAEP brochures and *Measure Up* newsletters in areas commonly used by your teachers.
- Inform teachers that released NAEP test questions and responses, which they can use in their classroom, are posted at [http://nces.ed.gov/nationsreportcard/itmrlsx](http://nces.ed.gov/nationsreportcard/itmrlsx) and that NAEP frameworks are available online at [http://nagb.org/publications/frameworks.htm](http://nagb.org/publications/frameworks.htm).

Promoting the importance of NAEP with students.
Students who are selected for NAEP will represent hundreds of students across the nation, so it is vital that these eighth-graders participate and do their best. Here are some suggestions on how to encourage students to do their best:

- Show the 5-minute video, *Introducing NAEP to Students*, to students selected to participate in NAEP. This video can be accessed at [http://nces.ed.gov/nationsreportcard/students/](http://nces.ed.gov/nationsreportcard/students/).
- Speak with participating students prior to assessment day. Let them know why NAEP is important.
- Consider ways to thank students for their participation.

On the assessment date, you will be responsible for the following:

Ensuring that students attend the session.
Prior to the assessment start time, you will need to be available to ensure that students attend the sessions. You and/or teachers of the selected students are encouraged to remain in the room during the assessment. NAEP staff members will bring all necessary materials, including laptop computers and earbuds, to the school on assessment day. Schools will only need to provide rooms, desks or tables, and access to electrical outlets; schools do not have to provide Internet access. It is very important that attendance rates be as high as possible to avoid the need for makeup sessions. If attendance of sampled students is less than 90 percent, a makeup session will be needed, and the NAEP representative will schedule another date to administer the assessment to the students who were absent.

Informing parents/guardians of student participation.
By law, parents/guardians of students selected to participate in NAEP must be informed of their child’s selection prior to the administration of the assessment. Parents must be informed that their child may be excused from participation for any reason, is not required to finish the assessment, and is not required to answer all test questions. Your NAEP State Coordinator will provide a sample Parent/Guardian Notification Letter and additional information about how this requirement should be fulfilled. These details, as well as electronic copies of the Parent/Guardian Notification Letter, will also be provided through the MySchool website. Parent notification should be completed prior to the preassessment visit. Parents may also visit [http://nces.ed.gov/nationsreportcard/pdf/parents/2012469.pdf](http://nces.ed.gov/nationsreportcard/pdf/parents/2012469.pdf) to download a brochure about NAEP.

Distributing and collecting worksheets and questionnaires from school staff.
You will be responsible for distributing, collecting, and reviewing a set of worksheets and questionnaires.

- Worksheets completed by school staff provide important information about how to assess SD and ELL students. Review the instructions for distributing and completing these worksheets and then share them with the staff person(s) most knowledgeable about how these students are tested on your state assessment. Collect and review the completed worksheets prior to the preassessment visit to ensure they are accurate.
- A school questionnaire for the principal will be provided during the visit and should be completed prior to the assessment.

Meeting with the NAEP representative during the scheduled preassessment visit.
In late January, you will meet with the NAEP representative to review the Preassessment Packet contents and go over logistics for the NAEP administration. During the meeting, you will review the list of selected students to verify that their demographic information is accurate and complete. Details for the assessment day will be finalized, including the time and location(s) of the assessment, how students and teachers will be notified, and which students will require accommodations. The NAEP representative will also verify that parents have been notified and will collect a copy of the parent notification letter during this visit.
The National Assessment of Educational Progress (NAEP), otherwise known as The Nation’s Report Card, informs the public about the academic achievement of elementary and secondary students in the United States. Report cards communicate the findings of NAEP, a continuing and nationally representative measure of achievement in various subjects over time. For more than 35 years, NAEP has assessed achievement by testing samples of students most often in the fourth, eighth, and 12th grades. The results have become an important source of information on what U.S. students know and are able to do in a range of subject areas.

To create the new assessment, the National Assessment Governing Board sought a framework of technological literacy knowledge and skills that identifies the understandings and applications of technology principles that are important for all students. The framework defines “literacy” as the level of knowledge and competencies needed by all students and citizens. More than testing students for their ability to “do” engineering or produce technology, then, the assessment is designed to gauge how well students can apply their understanding of technology principles to real-life situations. At grade 4, for example, all students are expected to identify types of technologies in their world, design and test a simple model, explain how technologies can result in positive and negative effects, and use common technologies to achieve goals in school and in everyday life. By grade 12, students are expected to select and use a variety of tools and media to conduct research, evaluate how well a solution meets specified criteria, and develop a plan to address a complex global issue. To learn more, see a video clip (“Ecosystems”) in the interactive framework of a sample scenario for grade 8 showing a student investigation of how organisms in an ecosystem are affected by a pollutant.

Technological literacy at grades 4, 8, and 12 is a pathway promoting further study and occupational pursuits. The Governing Board assembled a broad array of individuals and organizations to create a test of students’ abilities to grasp and apply technology principles. The resulting framework is the culmination of a long, complex process that drew on the contributions of thousands of individuals and organizations including technology experts, engineers, teachers, researchers, business leaders, testing experts, and policymakers.

The 2014 NAEP Technology and Engineering Literacy Assessment will provide important results and information that can be used to determine whether our nation’s students have the essential knowledge and skills needed in the technology and engineering areas. Policymakers, educators, and the public can use data from the initial assessments as tools for monitoring certain aspects of student achievement in technology and engineering literacy over time.
Recognizing that it is not possible to assess every aspect of technology and engineering literacy, the TEL assessment framework targets the nature, processes, and uses of technology and engineering that are essential for 21st century citizens.

The assessment objectives are organized into three major areas: Technology and Society; Design and Systems; and Information and Communication Technology (ICT). Each broad category is further broken down into discrete areas to be assessed.

Definitions of Technology, Engineering, and Technology and Engineering Literacy

Any assessment of students’ technology and engineering literacy must start with a clear idea of exactly what technology and engineering literacy means. That in turn requires clear definitions of technology and engineering.

**Technology** is any modification of the natural world done to fulfill human needs or desires.

This definition sees technology as encompassing the entire human-made world, from paper to the Internet. Technology also includes the entire infrastructure needed to design, manufacture, operate, and repair technological artifacts, from corporate headquarters and engineering schools to manufacturing plants and media outlets.

**Engineering** is a systematic and often iterative approach to designing objects, processes, and systems to meet human needs and wants.

This framework defines technology and engineering literacy in a broad fashion:

**Technology and engineering literacy** is the capacity to use, understand, and evaluate technology as well as to understand technological principles and strategies needed to develop solutions and achieve goals.

Thus—as with scientific, mathematical, and language literacy—technology and engineering literacy involves the mastery of a set of tools needed to participate intelligently and thoughtfully in society.

The interconnected relationship among these three major assessment areas can be illustrated as a three-sided pyramid in which each side supports the other two. For example, in order to address an issue related to technology and society, such as clean water, energy needs, or information research, a person who is literate in technology and engineering must understand technological systems and the engineering design process and be able to use various information and communication technologies to research the problem and develop possible solutions.
Area 1. Technology and Society deals with the effects that technology has on society and on the natural world and with the sorts of ethical questions that arise from those effects.

The four sub-areas in which students are assessed include:

A. Interaction of Technology and Humans concerns the ways in which society drives the improvement and creation of new technologies and how technologies serve society as well as change it. Fourth-graders are expected to know that people’s needs and desires determine which technologies are developed or improved. For example, cell phones were invented, produced, and sold because people found it useful to be able to communicate with others wherever they were. Eighth-graders are expected to understand how technologies and societies co-evolve over significant periods of time. For example, the need to move goods and people across distances prompted the development of a long series of transportation systems from horses and wagons to cars and airplanes. By 12th grade, students are expected to realize that the interplay between culture and technology is dynamic, with some changes happening slowly and others very rapidly. They should be able to use various principles of technology design—such as the concepts of trade-offs and unintended consequences—to analyze complex issues at the interface of technology and society and to consider the implications of alternative solutions.

B. Effects of Technology on the Natural World is about the positive and negative ways that technologies affect the natural world. Fourth-graders are expected to know that sometimes technology can cause environmental harm. For example, litter from food packages and plastic forks and spoons discarded on city streets can travel through storm drains to rivers and oceans where they can harm or kill wildlife. Eighth-graders are expected to recognize that technology and engineering decisions often involve weighing competing priorities, so that there are no perfect solutions. For example, dams built to control floods and produce electricity have left wilderness areas under water and affected the ability of certain fish to spawn. By 12th grade, students should have had a variety of experiences in which technologies were used to reduce the environmental impacts of other technologies, such as the use of environmental monitoring equipment.

C. Effects of Technology on the World of Information and Knowledge focuses on the rapidly expanding and changing ways that information and communication technologies enable data to be stored, organized, and accessed and on how those changes bring about benefits and challenges for society. Fourth-graders should know that information technology provides access to vast amounts of information, that it can also be used to modify and display data, and that communication technologies make it possible to communicate across great distances using writing, voice, and images. Eighth-graders should be aware of the rapid progress in development of ICT, should know how information technologies can be used to analyze, display, and communicate data, and should be able to collaborate with other students to develop and modify a knowledge product. By 12th grade, students should have a full grasp of the types of data, expertise, and knowledge available online and should be aware of intelligent information technologies and the uses of simulation and modeling.

D. Ethics, Equity, and Responsibility concerns the profound effects that technologies have on people, how those effects can widen or narrow disparities, and the responsibility that people have for the societal consequences of their technological decisions. Fourth-graders should recognize that tools and machines can be helpful or harmful. For example, cars are very helpful for going from one place to another quickly, but their use can lead to accidents in which people are seriously injured. Eighth-graders should be able to recognize that the potential for misuse of technologies always exists and that the possible consequences of such misuse must be taken into account when making decisions. By 12th grade, students should be able to take into account both intended and unintended consequences in making technological decisions.
Area 2. Design and Systems covers the nature of technology, the engineering design process by which technologies are developed, and basic principles of dealing with everyday technologies, including maintenance and troubleshooting.

The four sub-areas in which students are assessed include:

A. Nature of Technology offers a broad definition of technology as consisting of all the products, processes, and systems created by people to meet human needs and desires. Fourth-graders are expected to distinguish natural and human-made materials, to be familiar with simple tools, and to recognize the vast array of technologies around them. Eighth-graders should know how technologies are created through invention and innovation, should recognize that sometimes a technology developed for one purpose is later adapted to other purposes, and should understand that technologies are constrained by natural laws. By 12th grade, students should have an in-depth understanding of the ways in which technology coevolves with science, mathematics, and other fields; should be able to apply the concept of trade-offs to resolve competing values, and should be able to identify the most important resources needed to carry out a task.

B. Engineering Design is a systematic approach to creating solutions to technological problems and finding ways to meet people's needs and desires. Fourth-graders should know that engineering design is a purposeful method of solving problems and achieving results. Eighth-graders should be able to carry out a full engineering design process to solve a problem of moderate difficulty. By 12th grade, students should be able to meet a complex challenge, weigh alternative solutions, and use the concept of trade-offs to balance competing values.

C. Systems Thinking is a way of thinking about devices and situations so as to better understand interactions among components, root causes of problems, and the consequences of various solutions. Fourth-graders should know that a system is a collection of interacting parts that make up a whole, that systems require energy, and that systems can be either living or non-living. Eighth-graders should be able to analyze a technological system in terms of goals, inputs, processes, outputs, feedback, and control, and they should be able to trace the life cycle of a product from raw materials to eventual disposal. By 12th grade, students should be aware that technological systems are the product of goal-directed designs and that the building blocks of any technology consist of systems that are embedded within larger technological, social, and environmental systems. They should also be aware that the stability of a system is influenced by all of its components, especially those in a feedback loop.

D. Maintenance and Troubleshooting is the set of methods used to prevent technological devices and systems from breaking down and to diagnose and fix them when they fail. Fourth-graders should know that it is important to care for tools and machines so they can be used when they are needed. Students should also know that if something does not work as expected, it is possible to find out what the problem is in order to decide if the item should be replaced or how to fix it. Eighth-graders should be familiar with the concept of maintenance and should understand that failure to maintain a device can lead to a malfunction. They should also be able to carry out troubleshooting, at least in simple situations. By 12th grade, students should know that many devices are designed to operate with high efficiency only if they are checked periodically and properly maintained. They should also have developed the capability to troubleshoot devices and systems, including those that they may have little experience with.

Area 3. Information and Communication Technology includes computers and software learning tools, networking systems and protocols, hand-held digital devices, and other technologies for accessing, creating, and communicating information and for facilitating creative expression.

The five sub-areas in which students are assessed include:

A. Construction and Exchange of Ideas and Solutions concerns an essential set of skills needed for using ICT and media to communicate ideas and collaborate with others. Fourth-graders should understand what is expected from members working as part of a team and should realize that teams are better than individuals at solving many kinds of problems. Eighth-graders should know that communicating always involves understanding the audience—the people for whom the message is intended. They should also be able to use feedback from others, and provide constructive criticism. By 12th grade, students are expected to have developed a number of effective strategies for collaborating with others and improving their teamwork. They should be able to synthesize information from different sources and communicate with multiple audiences.

B. Information Research includes the capability to employ technologies and media to find, evaluate, analyze, organize, and synthesize information from different sources. Fourth-graders should be aware of a number of digital and network tools that can be used for finding information, and they should be able to use these tools to collect, organize, and display data in response to specific questions and to help solve problems. Eighth-graders should be aware of digital and network tools and be able to use them efficiently. They should be aware that some of the information they retrieve may be distorted, exaggerated, or otherwise misrepresented, and they should be able to identify cases where the information is suspect. By 12th grade, students should be able to use advanced search methods and select the best digital tools and resources for various purposes. They should also be able to evaluate information for timeliness and accuracy.

C. Investigation of Problems concerns the use of information and communication technology to define and solve problems in core school subjects and in practical situations. Fourth-graders should be able to use a variety of information and communication technologies to investigate a local or otherwise familiar issue and to generate, present, and advocate
for possible solutions. **Eighth-graders** should be able to use digital tools to identify and research a global issue and to identify and compare different possible solutions. By **12th grade**, students should be able to use digital tools to research global issues and to fully investigate the pros and cons of different approaches. They should be able to design and conduct complex investigations in various subject areas using a variety of digital tools to collect, analyze, and display information and be able to explain the rationale for the approaches they used in designing the investigation as well as the implications of the results.

**D. Acknowledgement of Ideas and Information** involves respect for the intellectual property of others and knowledge of how to credit others’ contributions appropriately, paying special attention to the misuse of information enabled by rapid technological advances. **Fourth-graders** should understand that it is permissible to use others’ ideas as long as appropriate credit is given. They should also know that copyrighted materials cannot be shared freely. **Eighth-graders** should be aware of general principles concerning the use of other people’s ideas and know that these principles are the basis for such things as school rules and federal laws governing such use. They should know about the limits of fair use of verbatim quotes and how to cite sources. By **12th grade**, students should understand the fundamental reasons for intellectual property laws and should know acceptable practices for citing sources when incorporating ideas, quotes, and images into their own work.

**E. Selection and Use of Digital Tools** includes both knowledge and skills for choosing appropriate tools and using a wide variety of electronic devices, including networked computing and communication technology and media. **Fourth-graders** should know that different digital tools have different purposes and they should also be able to use a variety of digital tools that are appropriate for their age level. **Eighth-graders** should be familiar with different types of digital tools and be able to move easily from one type of tool to another—for example, creating a document or image with one tool and then using a second tool to communicate the result to someone at a distant location. By **12th grade**, students should be competent in the use of a broad variety of digital tools and be able to explain why some tools are more effective than others that were designed to serve the same purpose, based on the features of the individual tools.

Although these elements are central to the design of the NAEP Technology and Engineering Literacy Assessment, they are not sufficient to describe the kinds of reasoning to be expected from students, the context or subject matter that will be used to construct test items, or the overall shape of the entire assessment. The assessment targets and the sub-areas within each describing what students should be able to do foreshadow the cross-cutting practices—ways of thinking and reasoning—for which the TEL is designed.

**Practices and Contexts for Technology and Engineering Literacy**

In all three areas of technology and engineering literacy, students are expected to be able to apply particular ways of thinking and reasoning when approaching a problem, and they are expected to do so in various contexts.

The practices can be grouped into three broad categories: **Understanding Technological Principles; Developing Solutions and Achieving Goals; and Communicating and Collaborating.**

**Understanding Technological Principles** focuses on students’ knowledge and understanding of technology and their capability to think and reason with that knowledge.

**Developing Solutions and Achieving Goals** refers to students’ systematic application of technological knowledge, tools, and skills to address problems and achieve goals presented in societal, design, curriculum, and realistic contexts.

**Communicating and Collaborating** centers on students’ capabilities to use contemporary technologies to communicate for a variety of purposes and in a variety of ways, working individually or in teams.

These practices are applied across all three major assessment areas. For example, communicating effectively and collaborating with others are necessary skills for understanding the effects of technology on the natural world, designing an engineering solution to a technological problem, and achieving a goal using information and communication technologies.
As crucial to the assessment as the practices are the **contexts**—the situations and types of problems in which assessment tasks and items will be set.

Below are examples of the types of tasks and items that result when these three elements are combined. The table shows how the three practices—Understanding Technological Principles, Developing Solutions and Achieving Goals, and Communicating and Collaborating—can be used to classify the general types of thinking and reasoning intended by the assessment targets in the three major assessment areas of Technology and Society, Design and Systems, and Information and Communication Technology.

### Elements of the NAEP Technology and Engineering Literacy Assessment

The practices expected of students are general, cross-cutting reasoning processes that students must use in order to show that they understand and can use their technological knowledge and skills. The contexts in which technology and engineering literacy tasks and items appear will include typical issues, problems, and goals that students might encounter in school or practical situations. Together, the assessment targets, practices, and contexts provide a structure for the generation of tasks and items.

### Classification of types of assessment targets in the three major assessment areas according to the practices for technology and engineering literacy

<table>
<thead>
<tr>
<th>Technology and Society</th>
<th>Design and Systems</th>
<th>Information and Communication Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Understanding Technological Principles</strong></td>
<td><strong>Design and Systems</strong></td>
<td><strong>Information and Communication Technology</strong></td>
</tr>
<tr>
<td>Analyze advantages and disadvantages of an existing technology</td>
<td>Describe features of a system or process</td>
<td>Describe features and functions of ICT tools</td>
</tr>
<tr>
<td>Explain costs and benefits</td>
<td>Identify examples of a system or process</td>
<td>Explain how parts of a whole interact</td>
</tr>
<tr>
<td>Compare effects of two technologies on individuals</td>
<td>Explain the properties of different materials that determine which is suitable</td>
<td>Analyze a process or outcome</td>
</tr>
<tr>
<td>Propose solutions and alternatives</td>
<td>Evaluate a need</td>
<td>Evaluate examples of effective resolution of opposing points of view</td>
</tr>
<tr>
<td>Predict consequences of a technology</td>
<td>Classify the elements of a system</td>
<td>Justify tool choice for a given purpose</td>
</tr>
<tr>
<td>Select appropriate technology to solve a societal problem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop a plan to investigate an issue</td>
<td>Design and Build a product using appropriate processes and materials</td>
<td>Select and Use appropriate tools to achieve a goal</td>
</tr>
<tr>
<td>Gather and Organize data and information</td>
<td>Develop forecasting techniques</td>
<td>Search media and digital resources</td>
</tr>
<tr>
<td>Analyze and Compare advantages and disadvantages of a proposed solution</td>
<td>Construct and Test a model or prototype</td>
<td>Evaluate credibility and solutions</td>
</tr>
<tr>
<td>Investigate environmental and economic impacts of a proposed solution</td>
<td>Produce an alternative design or product</td>
<td>Propose and Implement strategies</td>
</tr>
<tr>
<td>Evaluate trade-offs and impacts of a proposed solution</td>
<td>Evaluate trade-offs</td>
<td>Predict outcomes of a proposed approach</td>
</tr>
<tr>
<td><strong>Developing Solutions and Achieving Goals</strong></td>
<td><strong>Plan research and presentations</strong></td>
<td><strong>Evaluate effectiveness of design teams</strong></td>
</tr>
<tr>
<td>Select appropriate technology to solve a problem</td>
<td>Organize and present data and information</td>
<td>Critique a need</td>
</tr>
<tr>
<td>Develop a plan to investigate an issue</td>
<td>Transform from one representational form to another</td>
<td>Critique presentations</td>
</tr>
<tr>
<td>Design and Build a product using appropriate processes and materials</td>
<td>Conduct experiments using digital tools and simulations</td>
<td>Express historical issues in a multimedia presentation</td>
</tr>
<tr>
<td>Develop forecasting techniques</td>
<td>Construct and Test a model or prototype</td>
<td>Argue from an opposing point of view</td>
</tr>
<tr>
<td>Analyze and Compare advantages and disadvantages of a proposed solution</td>
<td>Plan and implement strategies</td>
<td>Explain to a specified audience how something works</td>
</tr>
<tr>
<td>Investigate environmental and economic impacts of a proposed solution</td>
<td>Evaluate trade-offs</td>
<td>Address multiple audiences</td>
</tr>
<tr>
<td>Evaluate trade-offs and impacts of a proposed solution</td>
<td>Evaluate trade-offs</td>
<td>Synthesize data and points of view</td>
</tr>
</tbody>
</table>

| **Communicating and Collaborating** | **Provide and Integrate feedback from virtual peers and experts to make changes in a presentation** | **Evaluate effectiveness of design teams** |
| Present innovative, sustainable solutions | Request input from virtual experts and peers | Evaluate effectiveness of design teams |
| Represent alternative analyses and solutions | | |
| Display positive and negative consequences using data and media | | |
| Compose a multimedia presentation | | |
| Produce an accurate timeline of a technological development | | |
| Delegate team assignments | | |
| Exchange data and information with virtual peers and experts | | |
| **Technology (ICT)** | **Communications** | **Information and** |
| Technology and Society | Information and Communication Technology |
| Design and Systems | Technology |
| **ASSESSMENT AREAS** | **Communication** | **Technology** |
| Technology and Society | | |
| Design and Systems | | |
| Information and Communication Technology (ICT) | | |
Content and Design

To identify what students know and can do with regard to technology and engineering, the NAEP TEL framework calls for the assessment to be totally computer-based. In 2014 the NAEP TEL assessment will be conducted at grade 8 with a national sample of students in public and private schools. The assessment will include tasks and items sampled from the domain of technology and engineering literacy achievement identified by the intersection of the three major areas of technology and engineering literacy and the cross-cutting practices at grades 4, 8, and 12—grades that will participate in the TEL assessment in future years.

Allowing students to demonstrate the wide range of knowledge and skills detailed in the NAEP Technology and Engineering Literacy Assessment targets will require a departure from the typical assessment designs used in other NAEP content areas. Thus students will be asked to perform a variety of actions using a diverse set of tools in the process of solving problems and meeting goals within rich, complex scenarios that reflect realistic situations. Consequently, this assessment will rely primarily on scenario-based assessment sets that test students through their interaction with multimedia tasks that include conventional item types, such as selected response items, and also monitor student actions as they manipulate components of the systems and models that are presented as part of the task.

Because of their capability to replicate authentic situations examinees may encounter in their lives, scenarios have the potential to provide a level of authenticity other types of assessment tasks cannot provide. At the same time, the choice to use these complex tasks reduces the number of measures that can be included in any one test and causes many of the measures to be interdependent because they are related to the same scenario. To counteract this interdependency and ensure reliability, the NAEP assessment of technology and engineering literacy will also include sets of discrete items that produce independent measures.

Scenario-Based Assessment Sets

There will be two types of scenario-based assessment sets, one long and one short. The long scenarios will take students approximately 25 minutes. The short scenarios will take students about 12 to 15 minutes to respond. The two types of scenarios have common characteristics, but they differ in the complexity of the scenario and the number of embedded assessment tasks and items to which a student is asked to respond.

A set of sample video clips demonstrates the types of interactivity and functionality of tools that students might be expected to use as they respond to short and long scenarios that will be developed for the Technology and Engineering Literacy Assessment.

Background Variables

Background data on students, teachers, and schools are needed to fulfill the statutory requirement that NAEP include information, whenever feasible, for various subgroups of students at the national level including gender, race/ethnicity, eligibility for free or reduced-price lunch, English language learners, and students with disabilities. Therefore, students, teachers, and school administrators participating in NAEP are asked to respond to questionnaires designed to gather demographic information.

Information is also gathered from non-NAEP sources, such as state, district, or school records. For the 2014 NAEP Technology and Engineering Literacy Assessment, only student and school information will be collected as many students will not have taken a separate course in technology and engineering literacy taught by a specific teacher.

In addition to demographic information, background questionnaires include questions about variables related to opportunities to learn and achievement in technology and engineering literacy. The variables are selected to be of topical interest, to be timely, and to be directly related to academic achievement and current trends and issues in technology and engineering literacy. Questions do not solicit information about personal topics or information irrelevant to the collection of data on technology and engineering literacy achievement.

Discrete Item Sets

Discrete item sets will include conventional selected response items and short constructed response items. The discrete item sets will comprise approximately 10-15 stand-alone items in either selected or constructed response format to be completed within a 25-minute block. Each discrete item would provide a stimulus that presents enough information to answer the particular question posed in the stem of the item. Items in discrete sets will be selected response items (e.g., multiple choice) or short constructed response items in which a student writes a text-based response.
Achievement Levels

The Governing Board uses student achievement levels of Basic, Proficient, and Advanced to report results of NAEP assessments. The achievement levels represent an informed judgment of “how good is good enough” in the various subjects that are assessed. Technology and Engineering Literacy achievement levels specific to the 2014 NAEP Technology and Engineering Literacy Framework will be developed to elaborate the generic policy definitions of Basic, Proficient, and Advanced achievement. Preliminary achievement level definitions have been developed for each of the three areas to be reported separately in the assessment and they will be used to guide item development and initial stages of standard setting for the 2014 NAEP Technology and Engineering Literacy Assessment.

The preliminary achievement level definitions will be revised when actual student responses have been collected and analyzed. The Governing Board will convene panels of experts to examine the preliminary achievement level definitions and to recommend final achievement level definitions for each grade level.

Conclusion

For generations students have been taught about technology and have been instructed in the use of various technological devices, but there has been no way to know exactly what students understand about technologies and their effective uses. The exploding growth in the world of technology led the Governing Board to sponsor the development of a framework for a National Assessment of Technology and Engineering Literacy. The Governing Board hopes that this TEL Framework will serve as a significant national measure of what students know and can do in technology and engineering, and support improvements in student achievement.

To view the complete Technology and Engineering Literacy Framework for the 2014 NAEP, or to view an interactive version of the framework, please visit http://nagb.org/publications/frameworks.htm or call us at 202.357.6938.
The National Assessment Governing Board is an independent, bipartisan board whose members include governors, state legislators, local and state school officials, educators, business representatives, and members of the general public. Congress created the 26-member Governing Board in 1988 to set policy for the National Assessment of Educational Progress (NAEP).
For more information on the National Assessment Governing Board, please visit www.nagb.org or call us at 202-357-6938.
Science Interactive Computer Tasks

Development work has begun on the 2015 Science Interactive Computer Tasks for grades 4, 8 and 12. A subset of the tasks will be administered and reported in 2015.

Similar to the NAEP Technology and Engineering Literacy (TEL) assessment, the Science Interactive Computer Tasks will be developed using an Evidence Centered Design (ECD) model. ECD draws from research in cognitive science, psychometrics, and other fields to link together what will be measured (the student model), the evidence that will be used to make inferences (the evidence model), and the tasks that be will used to collect this evidence (the task model).

During the November 2012 meeting, the Assessment Development Committee (ADC) will review Interactive Computer Tasks currently under development. The session will be divided into three sections shown below.

1. Overview of the Science Interactive Computer Tasks Reporting Goals
   The first section will begin with a description of the reporting goals for the tasks and how the tasks will be used in conjunction with Hands-on Tasks and Discrete items to provide the richest possible measurement of students’ science knowledge, skills, and abilities.

2. Orientation to Design Patterns and Task Templates
   The second section will describe how the ECD model is used to translate the NAEP Science Framework into a set of design patterns and task templates that will both guide and document the task development process.

3. Review of Draft Task Outlines and Storyboards
   The third section will present task outlines for the first set of tasks under development. Some of these tasks will include draft storyboards with artwork to help the ADC see the flow and look/feel of the tasks. Several aspects of the task designs will be highlighted during this period, including how the elements of the designs link back to the design patterns and task templates, documenting the underlying validity argument for each task. The Committee will be asked to provide feedback on all task outlines.
## Assessment Development Committee
### Item Review Schedule
**November 2012 – August 2013**
*(Updated 10/30/12)*

<table>
<thead>
<tr>
<th>Review Package to Board</th>
<th>Board Comments to NCES</th>
<th>Background/Cognitive</th>
<th>Review Task</th>
<th>Approx Number Items</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/26/12</td>
<td>12/5/12</td>
<td>Cognitive</td>
<td>2015 Pilot Science Interactive Computer Tasks (SICTs) (4, 8, 12)</td>
<td>8 task outlines</td>
<td>Review on Nov. 30</td>
</tr>
<tr>
<td>12/21/12</td>
<td>1/17/13</td>
<td>Cognitive</td>
<td>2015 Pilot Science (4,8,12)</td>
<td>310 items</td>
<td></td>
</tr>
<tr>
<td>1/11/13</td>
<td>1/22/13</td>
<td>Cognitive</td>
<td>2015 Pilot SICTs (4, 8, 12)</td>
<td>8 task outlines 8 alpha versions*</td>
<td></td>
</tr>
<tr>
<td>2/14/13 (Tentative/TBD)</td>
<td>3/7/13</td>
<td>Background</td>
<td>2015 Science (4, 8, 12)</td>
<td>169 items</td>
<td></td>
</tr>
<tr>
<td>2/11/13</td>
<td>2/25/13</td>
<td>Cognitive</td>
<td>2017 Reading (12)</td>
<td>TBD Reading Passages</td>
<td></td>
</tr>
<tr>
<td>2/25/13</td>
<td>3/5/13</td>
<td>Cognitive</td>
<td>2015 Pilot SICTs (4, 8, 12)</td>
<td>8 task outlines 8 alpha versions</td>
<td></td>
</tr>
<tr>
<td>3/28/13</td>
<td>4/18/13</td>
<td>Cognitive</td>
<td>2015 Pilot SICTs (4, 8, 12)</td>
<td>3 task outlines 8 alpha versions 3 beta versions</td>
<td></td>
</tr>
<tr>
<td>4/5 (Tentative/TBD)</td>
<td>TBD</td>
<td>Background</td>
<td>2014 TEL</td>
<td>TBD</td>
<td></td>
</tr>
<tr>
<td>5/02/13</td>
<td>5/23/13</td>
<td>Cognitive</td>
<td>2015 Pilot SICTs (4, 8, 12)</td>
<td>3 alpha versions 8 beta versions</td>
<td></td>
</tr>
<tr>
<td>6/13/13</td>
<td>7/5/13</td>
<td>Cognitive</td>
<td>2015 Pilot SICTs (4, 8, 12)</td>
<td>3 alpha versions 8 beta versions</td>
<td></td>
</tr>
<tr>
<td>7/25/13</td>
<td>8/8/13</td>
<td>Cognitive</td>
<td>2015 Pilot SICTs (4, 8, 12)</td>
<td>8 beta versions</td>
<td></td>
</tr>
</tbody>
</table>

*Alpha and beta versions are the first- and second-draft versions of the rendered task, respectively.*
## AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:30 pm</td>
<td>Welcome, Introductions, and Agenda Overview</td>
<td>David Driscoll, Chair</td>
</tr>
<tr>
<td>4:35 pm</td>
<td>Committee Issues and Challenges</td>
<td>Committee Chairs</td>
</tr>
<tr>
<td>4:45 pm</td>
<td>Report on First Meeting of Ad Hoc Committee on NAEP Background Information</td>
<td>Terry Holliday, Chair, Ad Hoc Committee</td>
</tr>
<tr>
<td>4:50 pm</td>
<td>Governing Board 25th Anniversary Planning</td>
<td>Mary Crovo, Deputy Executive Director</td>
</tr>
<tr>
<td>4:55 pm</td>
<td>Update: Education Summit for Parent Leaders</td>
<td>Cornelia Orr, Executive Director</td>
</tr>
<tr>
<td>5:00 pm</td>
<td>Executive Committee September 2012 Planning Meeting</td>
<td>David Driscoll</td>
</tr>
<tr>
<td>5:05 pm</td>
<td>Updating Governing Board Policy: Reviewing the Past, Looking to the Future</td>
<td>Ray Fields, Assistant Director for Policy and Research</td>
</tr>
<tr>
<td>5:25 pm</td>
<td><strong>ACTION ITEM</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Status of FY 2013 Appropriation for NAEP and the Governing Board</td>
<td>Ray Fields</td>
</tr>
<tr>
<td>5:30 pm</td>
<td>Adjourn</td>
<td></td>
</tr>
</tbody>
</table>

**Tentative:** CLOSED SESSION 5:30 – 6:00 p.m.

<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:30 pm</td>
<td>NAEP Contracts, Budget, and Schedule for 2013 and Beyond</td>
<td>Cornelia Orr, Peggy Carr, Associate Commissioner, NCES</td>
</tr>
<tr>
<td>6:00 pm</td>
<td>Adjourn</td>
<td></td>
</tr>
</tbody>
</table>
Governing Board 25th Anniversary Planning Committee

Current Board Members

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

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

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

Tonya Miles
General Public Representative
Mitchellville, MD

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

Board Alumni*

Amanda Avallone
8th Grade Teacher
Boulder, Colorado

Michael Guerra
Non-Public School Representative
Washington, DC

Mark Musick
General Public Representative
Atlanta, Georgia

Michael Nettles
Testing and Measurement Expert
Ann Arbor, Michigan

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

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

*Alumni are listed with their category and geographic location while serving on the Board.
Education Summit for Parent Leaders

In March 2011, the Governing Board established the Ad Hoc Committee on NAEP Parent Engagement. The charge to the Ad Hoc Committee was to develop recommendations for steps and strategies the Governing Board and representatives of the NAEP program can take directly, and/or support the efforts of others to increase parent awareness about the urgency to improve the levels of student achievement in the U.S. and the urgency to reduce the size of achievement gaps by race, ethnicity, and income levels, using NAEP data and resources.

The Ad Hoc Committee submitted its report\(^1\) to the Board on March 2, 2012. The report contained five recommendations, which the Governing Board accepted at the May 2012 meeting:

1. Specify the Target Audience: National, State, and Local Parent Leaders and Parent Organizations
2. Establish Relationships with Recognized Parent and Community-based Organizations
3. Develop Presentations and Materials Targeted to Parents for Use by Governing Board Members and Others
4. Develop Parent Pages on the Governing Board and NAEP Websites
5. Conduct a Parent Education Summit in Late Summer/Early Fall 2012

Planning for the fifth recommendation is underway and will be discussed at the Executive Committee meeting on November 29, 2012 (n.b., the timeframe for the summit is now February-April 2013).

The Reporting and Dissemination Committee will begin planning for implementation of the other four recommendations at their meeting on November 30, 2012.

\(^1\) [http://www.nagb.org/content/nagb/assets/documents/what-we-do/quarterly-board-meeting-materials/2012-03/Final%20Ad%20Hoc%20Committee%20Report.pdf](http://www.nagb.org/content/nagb/assets/documents/what-we-do/quarterly-board-meeting-materials/2012-03/Final%20Ad%20Hoc%20Committee%20Report.pdf)
Education Summit for Parent Leaders

The National Assessment Governing Board is planning a one-day education summit for parent leaders in the February-April 2013 timeframe. The summit will be conducted in Washington, D.C. and available across the nation via live-streaming internet video, with the potential for live TV and radio coverage.

The objective of the summit is to convey the urgency of improving student achievement in the United States for all children and the urgency of reducing achievement gaps between student subgroups.

The audience of 150-300 would consist primarily of parent and community leaders, parent organizations, and leaders in education, business, civil rights, the religious community, and legislative policy.

To bring attention to its importance for the nation’s future, as well as to help convey the non-partisan, universal interest in achieving the summit objective, current and former First Ladies (e.g., Michelle Obama and Laura Bush) would be invited to share the podium in delivering the keynote address.

One or more distinguished journalists or media representatives, acknowledged for intellect and freedom from bias, would be invited to moderate and provide a concluding summary.

A respected education advocate, with a strong reputation for compelling presentations on student achievement, would be invited to present the NAEP data as evidence of the need to address the summit objective.

Individual and panel presentations would be made to address the national imperative for improving achievement and closing achievement gaps from a wide range of perspectives. The intent is for these perspectives, taken together, to provide a compelling, unassailable argument for the urgent need to take action.

For example (not listed in priority order):

- Religious leaders to provide the moral perspective
- Economists to provide the national economic perspective
- Civil rights leaders to provide the equity perspective
- Military leaders to address the national security imperative
- Business leaders to address the human capital and employment imperative
- Scholars from nationally recognized policy institutions and foundations, representing a diverse range of philosophical orientations, to provide societal perspectives
- Demographers to address the implications from the perspective of a changing population
- Parent leaders to address the imperative for families and students
- Educators to describe actions that are needed to improve academic achievement overall and close achievement gaps

The National Assessment Governing Board is an independent, bipartisan organization created by Congress in 1988 to set policy for NAEP. The Governing Board oversees NAEP, identifies subjects to be tested, determines test content, sets performance standards called achievement levels for each assessment, approves test questions, and releases NAEP results in The Nation’s Report Card. The Board also works to improve the reporting of results to make sure they are communicated effectively to a wide range of Americans. The Governing Board is committed to making NAEP an accessible, useful resource for parents.

The National Assessment of Educational Progress (NAEP), also referred to as The Nation’s Report Card, is the only continuing, nationally representative measure of achievement in core subjects at grades 4, 8, and 12. NAEP provides achievement results and reveals trends over time; compares performance among states, urban districts, public and private schools, and student demographic groups; and informs the public about elementary and secondary school student academic performance.
Executive Committee September 2012 Planning Meeting

The Executive Committee met on September 6, 2012 in Washington, DC for a day-long planning meeting. The Committee members included those who would be retiring from the Governing Board on September 30 and new, incoming Committee members. The purpose of the meeting was to begin discussing long-term issues facing NAEP and the Board.

On the following pages are a summary overview of the meeting and background papers addressing key issues discussed during the meeting: NAEP reauthorization and the NAEP schedule of assessments.
National Assessment Governing Board
Executive Committee

Meeting of September 6, 2012

Board Members in Attendance: David Driscoll, Chair, Mary Frances Taymans, Vice-Chair Retiring, Susan Pimentel, Vice Chair Incoming, David Alukonis, Lou Fabrizio, Alan Friedman, Tonya Miles, Fielding Rolston, Eileen Weiser. Governing Board Staff: Cornelia Orr, Mary Crovo, Ray Fields, Angela Scott.

Meeting Outline

The Executive Committee convened at 8:30 a.m. The purpose of the meeting was to receive information about and discuss issues that are long-term in nature. No decisions were made or votes taken by the Executive Committee during the meeting.

The topics addressed were:

- Reauthorization of NAEP and ESEA
- NAEP and the Roles of the Governing Board, IES, and NCES
- The NAEP Schedule—Setting Priorities in an Uncertain Budget Environment
- What the Board Should Do to Direct Public Attention to Improving Achievement and Closing Achievement Gaps
- NAEP Proficient and 12th Grade Academic Preparedness
- Updating Board Policy on: Redesigning the National Assessment of Educational Progress
- Planning for the Board’s 25th Anniversary
- The Role of NAEP and the Governing Board in Relation to Common Core State Standards and Assessments

The Executive Committee adjourned at 3:30 p.m.
The NAEP authorizing legislation expired at the end of FY 2009. NAEP and the Governing Board have continued to operate through congressional appropriations in FY 2010, FY 2011, FY 2012, and now in FY 2013. This not an uncommon practice, but not ideal.

It is possible that the next Congress will address reauthorization of NAEP and the Governing Board, so it is timely to begin considering recommendations the Governing Board may want to make (N.B., Congress is likely to take up reauthorization of the Elementary and Secondary Education Act of 1965 first).

The purpose of this discussion draft is to provide background on legislation relevant to NAEP reauthorization and to pose some issues for discussion by the Executive Committee.

There are three laws that have provisions that affect NAEP directly. One additional law does not currently bear on NAEP, but may be of interest to amend, as will be discussed below. The three laws with NAEP-related provisions are:

- The National Assessment of Educational Progress Authorization Act (the NAEP Act)
- The Education Sciences Reform Act (ESRA--the authorizing legislation for the Institute of Education Sciences)
- Title I of the Elementary and Secondary Education Act of 1965, also referred to as the No Child Left Behind Act (ESEA/NCLB)

The additional law deserving attention is the Individuals with Disabilities Education Act (IDEA). Maximizing the participation in NAEP of students with disabilities and reducing the variability in exclusion rates across jurisdictions are continuing objectives of the NAEP program. Proposed amendments to IDEA intended to advance these objectives are presented for discussion.

**THE NAEP ACT/ESRA—Governance of NAEP**

**Background**

The NAEP Act and ESRA are inter-related laws. For example, the NAEP Act authorizes NAEP, establishes the Governing Board to oversee and set policy for NAEP (along with a number of specific responsibilities), and calls for the conduct of NAEP by the National Center for Education Statistics. ESRA establishes IES and the National Center for Education Statistics within IES, subject to IES policies and priorities.

Even though ESRA contains specific provisions intended to exempt NAEP and the Governing Board from the policies and priorities that otherwise affect IES activities, a number of provisions in the respective acts are either at odds or create ambiguity about the governance of NAEP.

Among these contradictory provisions, the ones most elemental and profound in their effect relate to the NAEP budget and annual budget requests. Under the NAEP Act, the Governing Board’s responsibilities include “select[ing] the subject areas to be assessed.” Accordingly, the Governing
Board has, since its inception in 1988, set the NAEP schedule of assessments—that is, determined the subjects and grades to be assessed in specific years, including whether they are to be assessed at the national, state, and urban district levels.

The Governing Board sets the schedule with a 10-year outlook. This enables states to have a clear and timely basis to plan for participation in NAEP and permits the NAEP program to plan in advance for the contracts and budgets needed for implementation.

The schedule of assessments is a primary expression of Governing Board policy and the fundamental cost-driver for the NAEP budget. The types of test items used and the costs for their development and scoring (e.g., multiple choice, constructed response, computer administered, hands-on science tasks, etc.); the number of grades and subjects assessed; and the scope of the samples (e.g., national, state, urban district) all have direct impact on costs.

However, the annual request for the NAEP budget is developed and proposed within the Department of Education by IES without consultation with the Governing Board. Once proposed by IES, Department staff do not share the NAEP request with the Governing Board, just as they would not share information about proposed requests across any Principal Office of the Department.

Even though the NAEP legislation requires “the Commissioner for Education Statistics to report to the [Governing] Board on the Department's actions to implement the decisions of the [Governing] Board,” this has not occurred with respect to the annual request for the NAEP budget until after the President’s annual budget request is delivered to Congress and made public.

**Conclusion**
The lack of a role in the development of the annual NAEP budget request is counter to the Board’s responsibility under law for setting the schedule of assessments. It is potentially detrimental to the effective operation of the very complex NAEP program. Intimate knowledge of the budget requests and budget outlook for NAEP is essential for informing the Board’s decisions about the schedule (as well as other policy matters), and having this information as it is being developed and proposed is consistent with the intent of Congress in assigning this responsibility to the Board.

**THE NAEP ACT/ESRA—Reporting NAEP Results**

**Background**
The reporting of NAEP results is a second key area in which the NAEP Act and ESRA create ambiguity about roles and responsibilities between the Board and IES/NCES. The NAEP Act provides that the Governing Board shall “plan and execute the initial public release of [NAEP]… reports...[and that the NAEP]…data shall not be released prior to the release of the reports [by the Governing Board].”
Further, the NAEP Act requires the Governing Board to “develop guidelines for reporting and disseminating results...develop standards and procedures for regional and national comparisons...and take appropriate actions needed to improve the form, content, use, and reporting of [NAEP] results.”

With respect to NCES, the NAEP Act states that “The Commissioner for Education Statistics shall, with the advice of the [Governing] Board...carry out [NAEP], through grants, contracts, or cooperative agreements...use widely accepted professional testing standards, objectively measure academic achievement, knowledge, and skills[, and]...collect and report assessment data...in a valid and reliable manner...”

Two important points about these provisions:

First, the role of the Board is focused on determining the policy affecting the reporting of results. With all of the data that NAEP collects, the decisions about what to include in NAEP reports and how conclusions about the results are framed are crucial policy matters. This is an appropriate assignment for the Board, consistent with the requirement that the Board shall “…[exercise] its independent judgment, free from inappropriate influences and special interests…” This provision is designed to protect NAEP's integrity and credibility from partisan and other internal and external influences in all aspects of NAEP, but particularly in reporting NAEP results.

Second, the role of NCES is focused on achieving high technical quality in the data collection procedures and in the reliability, validity, and accuracy of the data that are to be reported. This, too, is an appropriate assignment, consistent with the role and functions of a federal statistical agency.

The NAEP legislation does not specifically authorize the Commissioner to release NAEP results; it is the Governing Board that is assigned the responsibility of planning and executing the initial public release of NAEP reports, and it is particularly germane that the law prohibits any party from releasing NAEP data prior to this release of NAEP reports—this includes IES, NCES, and the Secretary.

However, NCES has maintained that it has the responsibility for releasing NAEP reports and for controlling the content of NAEP reports. This is due in part to provisions of ESRA related to the IES Director’s authority to publish reports and the NCES Commissioner’s authority to establish related “procedures to ensure that [NCES] reports issued...are relevant, of high quality, useful to customers, subject to rigorous peer review, produced in a timely fashion, and free from any partisan political influence.”

It is noteworthy that the verbs used in ESRA with respect to reports are “issue” and “publish”; the word “release” does not appear anywhere in ESRA. It is only in the NAEP Act that the word “release” is used—and only in relation to the Governing Board’s role.
Conclusion
There is a need for greater clarification of respective roles and responsibilities regarding NAEP data review and report preparation. This clarification should result in changes to both the NAEP Act and ESRA, with the aim of achieving coherence between the two related laws and the clear assignment of appropriate roles and responsibilities between the Governing Board and IES/NCES with respect to NAEP.

Taken together, the division of responsibilities in the NAEP/ESRA provisions cited above suggest that the role of NCES should be focused on (1) ensuring that data collection procedures follow professional standards and (2) certifying the accuracy, reliability and validity of the NAEP data. Once the NAEP data are certified by NCES, the Governing Board should have a greater role in determining what will be reported.

ESEA/NCLB

Background
There are two provisions in ESEA/NCLB that directly affect NAEP. The first requires state education agencies (SEAs) to assure, in the plans they submit to the Department of Education for Title I funding, that they will participate every two years in the NAEP state 4th and 8th grade reading and mathematics assessments. The second provision requires local education agencies (LEAs) receiving Title I funds to assure, in the plans they submit to their SEA, that they will participate every two years in the NAEP state 4th and 8th grade reading and mathematics assessments if their schools are selected for the sample. These two provisions make participation in state NAEP mandatory for the 4th and 8th grade reading and mathematics assessments.

Companion provisions in the NAEP legislation ensure that these mandatory assessments are scheduled and conducted consistent with the ESEA/NCLB. These provisions require NAEP national and state 4th and 8th grade reading and mathematics assessments to be conducted every two years and set the conduct of these assessments as the first priority for the use of funds appropriated for NAEP.

It is important to be aware that both the state and LEA assurances required under ESEA/NCLB are the essential drivers of mandatory participation in state NAEP and that there are companion provisions in NAEP that support implementation. Therefore, should Congress decide to change the ESEA provisions related to participation in NAEP (e.g., changing the subjects and/or grades assessed, or the frequency of the assessments), both the state and LEA assurances should be revised accordingly, and the companion provisions in the NAEP legislation should likewise be revised.

ESEA and NAEP both define Puerto Rico as a state; consequently, Puerto Rico is required to participate in NAEP 4th and 8th grade reading and mathematics. Puerto Rico has participated in NAEP mathematics assessments, but not reading. The Department of Education has waived the requirement that Puerto Rico participate in the NAEP reading assessments, because reading is defined in the NAEP Reading Framework as “reading in English” and the language of instruction in Puerto Rico is Spanish.
Examples of ESEA policy options that could be considered include:
- Continue (or change) mandatory biennial participation in reading and mathematics at grades 4 and 8 at the state level
- Add mandatory participation at 12th grade at the state level
- Add subjects (e.g., science, writing, U.S. history) for mandatory state participation
- Exempt Puerto Rico from participation in the NAEP reading assessment

Individuals with Disabilities Education Act (IDEA)

ISSUE: IDEA and Participation in State NAEP

Background
Under the current IDEA legislation, “All children with disabilities are included in all general State and district-wide assessment programs, including assessments described under section 1111 of the Elementary and Secondary Education Act of 1965, with appropriate accommodations and alternate assessments where necessary and as indicated in their respective individualized education programs (section 612(a)(16)(a).”

This requirement applies to state-adopted and district-adopted assessment programs only and does not apply to NAEP. NAEP is not considered a “State [or] district-wide assessment [program],” although it is “[an assessment] described under section 1111 of [ESEA]…” However, NAEP has for more than a decade carried out initiatives to increase the inclusion of students with disabilities, while at the same time recognizing that participation of individual students is voluntary under the NAEP legislation. An issue that has surfaced is that IDEA, unlike ESEA, does not explicitly acknowledge or address participation in NAEP.

Examples of IDEA policy options that could be considered include:
- Text added to IDEA explicitly acknowledging that
  - ESEA requires states to participate in NAEP,
  - participation of students with disabilities, while voluntary, is important to ensure that NAEP samples are state-representative,
  - NAEP is not a State or district-wide assessment for the purposes of ESEA and IDEA (and thus is exempt from providing alternative assessments), and
  - students with disabilities (other than those with the most significant cognitive disabilities) should be encouraged to participate in NAEP if selected for the sample and if their participation can be accomplished validly, with or without accommodations permitted on NAEP.
- Text added to IDEA explicitly authorizing State Educational Agencies (SEA) to provide technical assistance and information about NAEP to schools and Local Education Agencies (LEA) As under section 611(e)(2)(C), which describes the activities SEA may perform under their state IDEA grants.
Setting Priorities in an Uncertain Budget Environment:  
The NAEP Schedule of Assessments

Background

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

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

On the last page of this attachment is a staff proposal, developed in consultation with NCES, for the NAEP schedule of assessments through 2022. This proposal was presented at the Executive Committee meetings in May and August 2011, to provide a vehicle for starting a discussion about priorities for assessments, taking a very long-range view.

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

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

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

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

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

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

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

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

Other additions to the schedule include:

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

- Should 4th and 8th grade reading and mathematics assessments at the state level continue every two years?
- Should state level writing and science continue once every four years?
- Should the frequency of some state-level subjects be reduced so that other state-level subjects can be added?
- Should 12th grade state assessments be given more/less frequently?
- Should the number of 12th grade states be expanded?
- Should the frequency of TUDA be changed?
- Should the frequency of long-term trend be changed?
- Are there subjects that should be added/dropped or be administered more/less frequently?
- Should the frequency of the High School Transcript Study be changed?
<table>
<thead>
<tr>
<th>Year</th>
<th>National</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>Reading</td>
<td>Reading (4, 8)</td>
</tr>
<tr>
<td></td>
<td>MATHEMATICS</td>
<td>MATHE (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td>Science (4, 8)</td>
</tr>
<tr>
<td></td>
<td>High School Transcript Study</td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>U.S. History</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ECONOMICS (12)</td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>Reading (4, 8)</td>
<td>Reading (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Mathematics (4, 8)</td>
<td>Math (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Writing (8, 12)</td>
<td>Writing (8)</td>
</tr>
<tr>
<td>2008</td>
<td>Arts (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long-term trend</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>READING</td>
<td>READING (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Mathematics*</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>SCIENCE**</td>
<td>SCIENCE (4, 8)</td>
</tr>
<tr>
<td></td>
<td>High School Transcript Study</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>U.S. History</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geography</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>Reading (4, 8)</td>
<td>Reading (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Mathematics (4, 8)</td>
<td>Math (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Science (8)**</td>
<td>Science (8)</td>
</tr>
<tr>
<td></td>
<td>WRITING (8, 12)**</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>Economics (12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long-term trend</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Reading</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td>2014</td>
<td>U.S. History</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Civics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geography</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TECHNOLOGY AND ENGINEERING LITERACY (8) **</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Reading</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Science**</td>
<td>Science (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>High School Transcript Study</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>Arts (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long-term trend</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>Reading</td>
<td>Reading (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td>Math (4, 8, 12)</td>
</tr>
<tr>
<td></td>
<td>Writing**</td>
<td>Writing (4, 8, 12)</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Year</th>
<th>National</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>U.S. History, Civics, Geography</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>Reading (4, 8), Mathematics (4, 8), Science (8)<strong>, WRITING (8, 12)</strong></td>
<td>Reading (4, 8), Math (4, 8), Science (8)</td>
</tr>
<tr>
<td>2012</td>
<td>Economics (12), Long-term trend</td>
<td>Reading (4, 8, 12), Math (4, 8, 12)</td>
</tr>
<tr>
<td>2013</td>
<td>Reading, Mathematics</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>U.S. History, Civics, Geography, TECHNOLOGY AND ENGINEERING LITERACY (8)**</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Reading, Mathematics, Science**, High School Transcript Study</td>
<td>Reading (4, 8, 12), Math (4, 8, 12), Science (4, 8, 12)</td>
</tr>
<tr>
<td>2016</td>
<td>Arts (8), Add: Economics (12), Long-term trend</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>Reading, Mathematics, Writing**, MATHEMATICS** (nat’l and state)</td>
<td>Reading (4, 8, 12), Math (4, 8, 12)<strong>, Writing (4, 8, 12)</strong></td>
</tr>
<tr>
<td>2018</td>
<td>U.S. History, Civics, Geography, CIVICS**, Technology and Engineering Literacy (8, 12)**</td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>Reading, Mathematics**, Science**, High School Transcript Study</td>
<td>Reading (4, 8, 12)<strong>, Math (4, 8, 12)</strong>, Science (4, 8, 12)**</td>
</tr>
<tr>
<td>2020</td>
<td>Long-term trend, Economics (12)<strong>, FOREIGN LANGUAGE (12)</strong></td>
<td>NOTE: administer by computer?</td>
</tr>
<tr>
<td>2021</td>
<td>Reading**, Mathematics**, Writing**</td>
<td>Reading (4, 8, 12)<strong>, Math (4, 8, 12)</strong>, Writing (4, 8, 12)**</td>
</tr>
<tr>
<td>2022</td>
<td>U.S. HISTORY**, Civics**, GEOGRAPHY**, WORLD HISTORY (12)<strong>, Technology And Engineering Literacy (4, 8, 12)</strong></td>
<td></td>
</tr>
</tbody>
</table>

**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 2015, 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.**
**Updating Governing Board Policy: Reviewing the Past, Looking to the Future**

The Executive Committee will be reviewing and updating the Governing Board policy “Redesigning the National Assessment of Educational Progress.”

This seminal policy, adopted in 1996, has served as a compass for the Board and NAEP. It contains the underlying basis for many of the fundamental positions the Board holds today. For example, it is the original source of 6 months as the goal for reporting NAEP results, the definition of the "general public" as the primary audience for NAEP reports, and the rationale for the 10-year outlook for the schedule of assessments, to name a few.

However, the educational environment NAEP is to serve has changed substantially since 1996, with the advent of mandated state participation in reading and mathematics assessments under No Child Left Behind; the development of the Common Cores State Standards and the associated work of the two state-based assessment consortia; and the Governing Board’s initiative to make NAEP an indicator of 12th grade student academic preparedness for college and job training.

While the fundamentals of the policy are still intact and valid, a number of provisions are out of date. In addition, the NAEP legislation has been amended, rendering some aspects of the policy obsolete.

The NAEP Redesign Policy appears on the following pages with comments and questions in track-changes format to prompt discussion at the Executive Committee meeting. The goal of the discussion is to provide general guidance to staff on the approach to take in updating or replacing the policy.

In addition to the comments and questions embedded in the document, please consider the following over-arching questions:

1. What is most important to convey about the current policy context and the role of NAEP?
2. What elements of the policy should remain and which should be amended?
3. Is the overall tone appropriate or should there be a change?
Adopted: August 2, 1996

National Assessment Governing Board

Redesigning the National Assessment of Educational Progress

Policy Statement

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

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

A Better Way to Measure Educational Progress in America

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

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

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

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

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

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

For example, where a single 120 page mathematics report once sufficed, mathematics reporting in 1992 consisted of seven volumes totaling almost 1,800 pages, not including individual state reports. Also, there are now two separate testing programs for reading, writing, math, and science. One monitors trends using tests developed during the 1970's; the other reflects current views on instruction and uses performance standards to report whether achievement is good enough.

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

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

National Assessment Redesign

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

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

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

Purpose and Objectives of the National Assessment of Educational Progress

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

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

Thus, the central concern of the National Assessment is to inform the nation on the status of student achievement. The National Assessment Governing Board believes that this should be accomplished through the following objectives:
1. To measure national and state progress toward the third National Education Goal and provide timely, fair, and accurate data about student achievement at the national level, among the states, and in comparison with other nations;

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

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

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

The Audience for the National Assessment

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

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

Limitations: What the National Assessment Is Not

The National Assessment is intended to describe how well students are performing, but not to explain why. The National Assessment only provides group results; it is not an individual student test. The National Assessment tests academic subjects and does not collect information on individual students' personal values or attitudes. Each National Assessment test is developed through a national consensus process. This national consensus process takes into account education practices, the results of education research, and changes in the curricula. However, the National Assessment is independent of any particular curriculum and does not promote specific ideas, ideologies, or teaching techniques. Nor is the National Assessment an appropriate means, by itself, for improving
instruction in individual classrooms, evaluating the effects of specific teaching practices, or determining whether particular approaches to curricula are working.

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

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

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

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

- The National Assessment shall be conducted annually, two or three subjects per year, in order to cover all required subjects at least twice a decade.
- The National Assessment shall assess all subjects listed in the third National Educational Goal—reading, writing, mathematics, science, history, geography, civics, the arts, foreign language and economics—according to a publicly released schedule adopted by the National Assessment Governing Board, covering eight to ten years, with reading, writing, mathematics, and science tested more frequently than the other subjects.
• The National Assessment Governing Board shall consult with technical experts and with education policymakers, in conjunction with the development of assessment frameworks, to determine the feasibility, desirability, and costs of combining several related subjects into a single assessment.

Provide National Assessment results for states

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

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

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

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

There is a strong interest among states to participate in the National Assessment to get state level information at grades 4 and 8 in reading, writing, mathematics, and science. The level of interest in participating in the National Assessment varies with respect to the other subjects (i.e., history, geography, civics, economics, the arts, and foreign language) and at grade 12, where state officials say that obtaining cooperation from high schools and 12th grade students is difficult.

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

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

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

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

Vary the amount of detail in testing and in reporting results

More subjects can be assessed if different strategies are used. Currently, each time the National Assessment is conducted, it uses a similar approach, regardless of the nature of the subject or the number of times an assessment in a subject has been administered. This approach is locked-in through 1998 under current contracts. Under this approach, a larger number of students is tested in order to provide not just overall results, but fine-grained details as well (e.g. the achievement scores of 4th grade students whose teachers that year had five hours or more of in-service training). The National Assessment also collects "background" information through questionnaires completed by students, teachers, and principals. The questionnaires ask about teaching practices, school policies, and television
watching, to name a few. Data analyses are elaborate. Reports are detailed and exhaustive, involving as many as seven separate reports per subject. Although the National Assessment has been praised for this thoroughness, the cost of this thoroughness is that fewer subjects are assessed, assessments occur less frequently, and reports take longer to produce.

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

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

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

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

The use of background surveys also would be varied. The three kinds of background surveys—student, teacher and principal questionnaires—would not necessarily all be employed each time a subject is assessed. Instead, the use of such surveys would be limited and selective, with reports of results focused on a core of background questions addressing the most essential issues. Also, background surveys used for standard report cards in a particular year would be designed to complement, rather than duplicate, background surveys used for comprehensive reports in the same year.
• National Assessment testing and reporting shall vary, using standard report cards most frequently, comprehensive reporting in selected subjects about once every ten years, and special, focused assessments.
• National Assessment results shall be timely, with the goal being to release results within 6 months of the completion of testing for standard report cards and within 9 months for comprehensive reports.

Simplify the National Assessment design

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

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

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

Simplify the way the National Assessment reports trends in student achievement

From its beginning in 1969, monitoring achievement trends has been a central mission of the National Assessment of Educational Progress. Monitoring long-term trends in educational achievement, both for the population as a whole and for significant subgroups, is a capacity unique to the National Assessment and should be continued as a central mission. However, as the National Assessment approaches its third decade, it must address the problem of how to assess trends in achievement when curricula continue to evolve and change. An assessment in a subject must be kept stable to monitor trends. However, stable assessments may not reflect important changes in curricula. Over time, there develops a legitimate concern about the relevance of the content of the assessment versus the ability to track change in achievement.
As a solution to this problem, since 1990, the National Assessment has reported achievement trends using two unconnected assessment programs. The tests, criteria for selecting students, and reporting are all different. The first program, "the main National Assessment," tests at grades 4, 8, and 12 and covers ten subjects. The assessments are based on a national consensus representing current views of each subject. Performance standards are used to report whether student achievement on the National Assessment is "good enough." The schedule of subjects to be assessed in the main National Assessment is unrelated to the schedule of subjects under the second testing program.

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

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

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

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

In reporting on "educational progress," the National Assessment has, until recently, only considered current student performance compared to student achievement in previous years. Under this approach, the only standard was how well students had done previously, not how well they should be doing on what is measured by the National Assessment. Although this approach has been useful, it began to change in 1988 from a sole focus on "where we have been" to include "where we want to be" as well.

Comment [AU27]: At the time of the adoption of the policy in 1996, the trend data for “main NAEP” were only a few years old; today we are about to begin the third decade of trend data for 4th and 8th grade reading and math. The current long-term trend NAEP in reading and math goes back to the early 1970’s. Is the time right to begin the “carefully planned transition” for main NAEP?

Comment [AU28]: This section, assuming it should be retained, should be moved under objective 2.
In 1988, Congress created a non-partisan citizen's group—the National Assessment Governing Board—and authorized it to set explicit performance standards, called achievement levels, for reporting National Assessment results.

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

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

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

Use international comparisons

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

**Emphasize reporting for grades 4, 8, and 12**

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

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

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

---

Comment [AU31]: Are these still the right grades?
Should these be augmented?
Is the argument for grade versus age-based testing still relevant?
Much has been done about 12th grade participation and motivation over the years since this report—should this be incorporated?
The National Assessment shall continue to test in and report results for grades 4, 8, and 12; however, in selected subjects, one or more of these grades may not be tested.

Age-based testing and reporting shall be permitted when deemed appropriate and when necessary for international comparisons and for long-term trends, should the National Assessment Governing Board decide to continue long-term trends in their current form.

Grade 12 results shall be accompanied by clear, highlighted statements about school and student participation, student motivation, and cautions, where appropriate, about interpreting 12th grade achievement results.

The National Assessment design shall seek to improve school and student participation rates and student motivation at grade 12.

The National Assessment shall provide practical incentives for school and district participation at grades 4, 8, and 12.

Use innovations in measurement and reporting

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

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

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

Where warranted, the National Assessment shall implement such advances in order to reduce costs and/or improve test administration, measurement, and reporting.

The next competition for National Assessment contracts, for assessments beginning around the year 2000, shall ask bidders to provide a plan for
(1) conducting testing by computer in at least one subject at one grade, and
(2) making use of technology to improve test administration, measurement, and reporting.

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

**Keep test frameworks and specifications stable**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Draft Resolution**

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

Whereas, absent action by Congress, the Budget Control Act of 2011 would, as of January 2, 2013, result in cuts estimated at approximately 8 percent of the NAEP and Governing Board appropriations in FY 2013; and

Whereas, the National Center for Education Statistics (NCES) is reviewing contract bids for NAEP operations that are scheduled to be awarded by the end of December 2012; and

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

Therefore, the Executive Committee requests a delegation of authority to act on behalf of the Governing Board, prior to the March 2013 Board meeting, to make decisions on changes to the NAEP schedule of assessments in 2014 and beyond, as necessary, in consultation with NCES.
Tentatively Scheduled Closed Session on

NAEP Contracts, Budget, and Schedule for 2013 and Beyond

Governing Board staff have made provisions for a closed session of the Executive Committee from 5:30 p.m. to 6:00 p.m. on November 29, 2012.

The closed session will be needed only in the event that Congress takes action on the FY 2013 appropriation by that date, in a manner that would require Executive Committee review of the impact on the NAEP contracts, budget and schedule.
# National Assessment Governing Board Committee on Standards, Design and Methodology

**November 30, 2012**  
**9:30 am – 12:00 pm**

**AGENDA**

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Attachments</th>
</tr>
</thead>
</table>
| 9:30 – 9:40 am     | Introductions  
Welcome New COSDAM Member, Andrew Ho  
Agenda Overview  
*Lou Fabrizio, COSDAM Chair* |             |
| 9:40 – 10:45 am    | NAEP 12th Grade Preparedness Research  
*Cornelia Orr, NAGB Executive Director*  
*Ray Fields, NAGB Assistant Director for Policy and Research*  
*Michelle Blair, NAGB Senior Research Associate*  
- Additional Updates from 2009 Research  
- Plans for Reporting 2009 Research Studies  
- Plans for 2013 Research Studies | Attachment A |
| 10:45 – 11:50 am   | Closed Session  
NAEP/TIMSS Linking Study  
*Laurie Wise, HumRRO* | Attachment B |
| 11:50 am – 12:00 pm| Open Session  
Recommendations for Future COSDAM Agenda Topics  
*COSDAM Members* |             |
NAEP 12th Grade Preparedness Research
Updates with 2009 Reporting Plans and 2013 Research Proposals

Based on the Program of Preparedness Research adopted by the Governing Board in March 2009, four categories of research studies were conducted to produce evidence to develop and support the validity of statements for NAEP reporting on the academic preparedness in reading and mathematics of 12th grade students for college and job training.

▪ content alignment studies;
▪ statistical relationship studies;
▪ judgmental standard setting studies; and
▪ surveys

Additionally, the Texas Commissioner of Higher Education offered the opportunity to conduct a benchmarking study with Texas higher education institutions, and a pilot study to examine the feasibility was conducted.

Based on discussions at its quarterly meetings in May 2012 and August 2012, the Governing Board has determined that the research studies completed to date should be made available through an online technical report. Dissemination through this format will be useful to the research community as well as policymakers and interested members of the general public.

In addition, the NAEP 12th Grade Preparedness Commission will conduct a symposium in Washington, DC in early 2013 focused on the Board’s preparedness research results and the plans for the next phase of the research.

The November 30, 2012 COSDAM briefing and committee discussion on preparedness will address:

▪ Progress updates related to the research being conducted in connection with the 2009 grade 12 NAEP results (Attachment A-1) ......................................................... Page 3

▪ Update and committee feedback on the reporting plans for the 2009 grade 12 NAEP preparedness research (Attachment A-2 embargoed, pending Board review, and distributed under separate cover) .................................................................Page 136

▪ Updates and committee feedback on proposed research projects for the preparedness research to be conducted in relation to the 2013 grade 12 NAEP assessment (Attachment A-3) .................................................................Page 138

Additionally, background materials to describe each major study category are appended as Attachment A-4 (see page 140).
Attachment A-1

Progress Updates on Research with 2009 Grade 12 NAEP Results

Higher Education Survey
A survey of two-year and four-year post-secondary institutions was conducted in Fall 2011 to gather information regarding (1) the tests used and (2) the cut scores on those tests in reading and mathematics below which need was indicated for remedial/developmental courses in reading and mathematics, and at or above which placement in credit-bearing entry level courses was indicated. The sample of accredited postsecondary education institutions was nationally representative. A weighted response rate of 81% was achieved. Attached is a copy of the report presenting the findings from the research (see page 6). Ray Fields will join the COSDAM meeting to discuss these results.

Job Training Program Content Analysis
In October 2011, the Governing Board began work with WestEd and its subcontractor, the Education Policy Improvement Center (EPIC), to conduct follow-up research relative to the NAEP preparedness judgmental standard setting (JSS) research. The research results from this project are intended to supplement the JSS research findings by providing a clearer understanding of the knowledge and skills required for entry- and exit-level coursework in designated occupational programs. By reviewing course artifacts such as syllabi, text books, and assignments, this study will help to determine if the knowledge, skills, and abilities (KSAs) required of students in the training programs are appropriately represented by the borderline preparedness descriptions (developed in the JSS research), by all the items on the 2009 NAEP, and by the 2009 NAEP items in the scale score ranges identified by panelists in the JSS research project.

For reading and mathematics, this project addresses the following research questions.

1) What is the degree of agreement between borderline preparedness on NAEP and those KSAs that are prerequisite to or taught within the training program?
2) What is the degree of agreement between the performance assessed by the NAEP items, and in particular the NAEP items at the cut scores resulting from the standard-setting process, and the KSAs that are prerequisite to and taught within each training program?

Attached is a status report further detailing the methodology and providing a summary of the project’s progress (see page 122).

College Course Content Analysis
In September 2012, the Governing Board awarded a contract to the Education Policy Improvement Center (EPIC) to conduct research and analysis of college course content in order to determine relationships between the prerequisite knowledge and skills in reading and mathematics for entry-level college courses and the content of grade 12 NAEP. This project addresses academic preparedness for college only—a separate parallel research project addresses preparedness for job training (described above).
In this project, EPIC will determine the entry-level (introductory) credit-bearing courses most frequently taken by entering students that are reflective of college-level reading and mathematics demands and that satisfy general education requirements. These introductory courses should have no college-level prerequisite course requirements, and only non-remedial courses that satisfy general education requirements should be included in the analysis. Further, in cases where multiple versions of a course are offered for majors and non-majors, only the course for non-majors should be included.

After this determination, EPIC will obtain course artifacts for a generally representative sample of institutions, and conduct several content analyses, e.g., an analysis of the introductory course artifacts for commonalities and differences in the reading and mathematics prerequisites (i.e., the prerequisite KSAs) needed to qualify for placement into the course. From these analyses, EPIC will develop descriptions of the knowledge, skills, and abilities needed for students to qualify for placement into the introductory course, based on an analysis of the course artifacts. And as part of a set of comparative analyses, EPIC will then use this description to review:

- the description of minimal requirements for placement into college-level coursework as developed in the NAEP preparedness judgmental standard setting (JSS) research
- KSAs represented by 2009 grade 12 items that map to the NAEP scale with a response probability of .67 and fall within the range of cut scores set by the two replicate panels in the JSS research
- 2009 and 2013 grade 12 NAEP items
- the KSAs represented by 2009 items that map in the range of the NAEP score scale from the mid-range of the Basic level to the mid-range of the Proficient level; and
- the NAEP achievement level descriptions.

For reading and mathematics, this project will address the following research questions.

1) What are the prerequisite knowledge, skills, and abilities (hereafter referred to as “prerequisite KSAs”) in reading and mathematics to qualify for entry-level credit-bearing courses that satisfy general education requirements?
2) How do these prerequisite KSAs compare with the 2009 and 2013 NAEP reading and mathematics frameworks and item pools?
3) How do these prerequisite KSAs compare with previous NAEP preparedness research, i.e., the descriptions of minimal academic preparedness requirements produced in the JSS research?
4) How can these prerequisites inform future NAEP preparedness research, i.e., planning and analysis efforts relative to the 2013 grade 12 NAEP reading and mathematics assessments?

A progress report is not attached at this time because the project is just starting.

**College Board Supplemental Analysis**
The statistical relationship studies include linking NAEP and the SAT in reading and mathematics. The College Board reports college readiness benchmarks related to the likelihood of earning a B- or better in terms of overall freshman year GPA. Hence, the college readiness benchmark is not content-specific performance; it may include different
ranges of subjects that may or may not be related to mathematics or reading, which is the focus of grade 12 NAEP research. At the request of the Governing Board, the College Board has calculated benchmarks that link SAT section scores to performance in multiple related freshmen college courses. For example, SAT Critical Reading section scores were linked to performance in courses which require extensive reading assignments and SAT Math section scores were linked to performance in math courses. See page 132 for further details on this analysis.
Tests and Cut Scores
Used for Student Placement
in Postsecondary Education: Fall 2011

Ray Fields
National Assessment Governing Board

Basmat Parsad
Westat

November 2012
National Assessment Governing Board

BOARD MEMBERSHIP (2012-2013)

Honorable David P. Driscoll
Chair
Former Commissioner of Education
Melrose, Massachusetts

Susan Pimentel
Vice Chair
Educational Consultant
Hanover, New Hampshire

Doris R. Hicks
Principal and
Chief Executive Officer
Dr. Martin Luther King, Jr. Charter
School for Science and Technology
New Orleans, Louisiana

Andrew Dean Ho
Assistant Professor
Harvard Graduate School
of Education
Harvard University
Cambridge, Massachusetts

Honorable Terry Holliday
Commissioner of Education
Kentucky Department of Education
Lexington, Kentucky

Richard Brent Houston
Principal
Shawnee Middle School
Shawnee, Oklahoma

Hector Ibarra
Eighth-Grade Teacher
Belin Blank International Center
and Talent Development
Iowa City, Iowa

Honorable Tom Luna
Idaho Superintendent
of Public Instruction
Boise, Idaho

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

Tonya Miles
General Public Representative
Mitchellville, Maryland

Dale Nowlin
Twelfth-Grade Teacher
Columbus North High School
Columbus, Indiana

Joseph M. O’Keefe, S.J.
Professor
Lynch School of Education
Boston College
Chestnut Hill, Massachusetts

W. James Popham
Professor Emeritus
University of California,
Los Angeles
Wilsonville, Oregon

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

Cary Sneider
Associate Research Professor
Portland State University
Portland, Oregon

Blair Taylor
Chief Community Officer
Starbucks Coffee Company
Seattle, Washington

Honorable Leticia Van de Putte
Senator
Texas State Senate
San Antonio, Texas

Ex-officio Member

John Q. Easton
Director
Institute of Education Sciences
U.S. Department of Education
Washington, D.C.

Suggested Citation:
Tests and Cut Scores
Used for Student Placement
in Postsecondary Education: Fall 2011

Ray Fields
National Assessment Governing Board

Basmat Parsad
Westat

November 2012
Acknowledgements

The authors wish to thank the many postsecondary institutions across the country that participated in the development of the survey, and the staff within the sampled institutions who gave their time and expertise to complete the survey forms.

We also wish to thank Jack Buckley, Peggy Carr, Patricia Etienne, Arnold Goldstein, Marilyn Seastrom, and Tom Weko of the National Center for Education Statistics for sage advice and the technical review of the data tables and technical report prepared by Westat, the contractor that conducted the survey.

We would like to thank Mildred Bazemore, Wayne Camara, and Michael Kirst, who reviewed the draft report and provided invaluable comments.

We are very grateful for the endorsement of the survey by the American Association of Community Colleges, American Association of State Colleges and Universities, American Council on Education, Association of Public and Land-grant Universities, National Association for College Admission Counseling, and State Higher Education Executive Officers.

Finally, many thanks go to Jacqui Lipson, Neby Ejigu, and the team at Widmeyer Communications, Inc., for proofreading and report design services.
## Acknowledgements


## List of Tables


## Executive Summary


## I. Introduction

Overview of the Governing Board Program of Preparedness Research .................................................. 1
Purpose of the Survey ........................................................................................................................................ 3
Policy Context .................................................................................................................................................. 3
Policy Considerations for the Survey .............................................................................................................. 4

## II. Survey Development, Sample, and Data Collection Methodology

Survey Development ........................................................................................................................................ 6
Sample ............................................................................................................................................................. 6
Data Collection Methodology ....................................................................................................................... 7

## III. Caveats and Limitations on Interpretation

Remedial/Developmental Coursework: Required or Recommended? .............................................................. 8
Use of Multiple Tests ..................................................................................................................................... 8
General Policy Versus Individual Student Placement Determination .......................................................... 9

## IV. Findings

Mathematics .................................................................................................................................................... 10
  Frequency of Use of Mathematics Tests ...................................................................................................... 10
  Mean Cut Scores on Mathematics Tests by Institution Level and Type ..................................................... 12
  Variability of Cut Scores on Mathematics Tests by Institution Level and Type ......................................... 14
  Other Criteria Used by Postsecondary Institutions ..................................................................................... 16

Reading ........................................................................................................................................................ 17
  Frequency of Use of Reading Tests ............................................................................................................. 17
  Mean Cut Scores on Reading Tests by Institution Level and Type ............................................................ 18
  Variability of Cut Scores on Reading Tests by Institution Level and Type ................................................ 20
  Other Criteria Used by Postsecondary Institutions .................................................................................... 22

## V. Conclusion

Summary of Findings ...................................................................................................................................... 23
Questions for Further Research .................................................................................................................. 25

## VI. References


## Appendices

Appendix A: Westat Technical Report ............................................................................................................. A-1
Appendix B: Standard Error Tables .................................................................................................................. B-1
Appendix C: Additional Tables ....................................................................................................................... C-1
List of Tables

Report Tables

Table 1  Estimated number of postsecondary institutions in the population and percentage of institutions using selected mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011 ................................. 11

Table 2  Mean mathematics test scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .................................................. 13

Table 3  Ranges of mathematics test scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .................................................. 15

Table 4  Percentiles for mathematics test cut scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .................................................. 15

Table 5  Estimated percentage of postsecondary institutions using criteria other than postsecondary mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011 .................................................. 16

Table 6  Estimated percentage of postsecondary institutions using selected reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011 .................................................. 17

Table 7  Mean reading test scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .................................................. 19

Table 8  Ranges of reading test scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .................................................. 21

Table 9  Percentiles for reading test cut scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .................................................. 21

Table 10 Estimated percentage of postsecondary institutions using criteria other than postsecondary reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011 .................................................. 22
Appendix B Tables

Table B.1 Standard errors for the estimated number of postsecondary institutions in the population and the percentage of institutions using selected mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011 ................. B-2

Table B.2 Standard errors for the mean mathematics test scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .............................................. B-2

Table B.3 Standard errors for the percentiles for mathematics test cut scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .............................................. B-3

Table B.4 Standard errors for the estimated percentage of institutions using criteria other than postsecondary mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011 .............................................. B-3

Table B.5 Standard errors for the estimated percentage of institutions using selected reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011 .............................................. B-4

Table B.6 Standard errors for the mean reading test scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .............................................. B-4

Table B.7 Standard errors for the percentiles for reading test cut scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011 .............................................. B-5

Table B.8 Standard errors for the estimated percentage of institutions using criteria other than postsecondary reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011 .............................................. B-5

Appendix C Tables

Table C.1 Percentage of postsecondary institutions using various mathematics tests to evaluate entering students for developmental or remedial courses in mathematics: Fall 2011 .............................................. C-2

Table C.2 Percentage of postsecondary institutions using various reading tests to evaluate entering students for developmental or remedial courses in reading: Fall 2011 .............................................. C-2

Table C.3 Standard errors for the percentage of postsecondary institutions using various mathematics tests to evaluate entering students for developmental or remedial courses in mathematics: Fall 2011 .............................................. C-3

Table C.4 Standard errors for the percentage of postsecondary institutions using various reading tests to evaluate entering students for developmental or remedial courses in reading: Fall 2011 .............................................. C-3
Executive Summary

The National Assessment Governing Board (NAGB) oversees and sets policy for the National Assessment of Educational Progress (NAEP). Since 1969, NAEP has provided information to the public on the achievement of the nation’s elementary and secondary students at grades 4, 8, and 12. As the only continuing measure of student achievement at grade 12, NAEP is uniquely positioned to report on the academic preparedness of 12th graders for college and job training.

To transform NAEP into a valid indicator of 12th grade academic preparedness, NAGB is conducting a comprehensive program of research. The fall 2011 survey of postsecondary education institutions’ use of tests and cut scores in college placement is one component of this larger research program. The survey’s central two-part research question is:

(1) Which national standardized tests are used by postsecondary education institutions to determine the need of entry-level students for remedial/developmental instruction in reading or mathematics?

(2) What are the cut scores on those national standardized tests below which students are deemed to need remedial/developmental instruction in reading or mathematics and at or above which students would be just academically qualified for entry-level credit-bearing college courses?

NAGB contracted with Westat to conduct the survey with a nationally representative sample of 2-year and 4-year public and private (both for-profit and not-for-profit) postsecondary education institutions. Slightly different survey forms were developed for 2-year and 4-year institutions (see Appendix A). In August 2011, Westat contacted the office of the president at each sampled institution to identify the appropriate respondent and sent survey materials to these individuals. Follow-up for survey recruitment, nonresponse, and data clarification was conducted via telephone and email by experienced interviewers from September 2011 through January 2012.

Mathematics Findings

During survey development, six nationally available standardized tests were consistently identified as being used by postsecondary education institutions in making determinations about student need for remedial/developmental instruction for mathematics: the ACT and SAT admissions tests, the ACCUPLACER Elementary Algebra and College-level Mathematics placement tests, and the COMPASS Algebra and College Algebra placement tests.

Seventy-one percent of postsecondary education institutions reported using some mathematics test for determining the need of entry-level students for remedial courses in mathematics. The range was from 4 percent for the COMPASS College Algebra test to 23 percent for ACT Mathematics.
The overall mean cut scores reported for the six standardized mathematics tests were:

- 19 on ACT mathematics on a scale of 1 to 36
- 471 on SAT mathematics on a scale of 200 to 800
- 70 on ACCUPLACER Elementary Algebra on a scale of 20 to 120
- 57 on ACCUPLACER College-Level Mathematics on a scale of 20 to 120
- 49 on COMPASS Algebra on a scale of 1 to 99
- 43 on COMPASS College Algebra on a scale of 1 to 99

**Reading Findings**

During survey development, five nationally available standardized tests were consistently identified as being used by postsecondary education institutions in making determinations about student need for remedial/developmental instruction for reading: the ACT and SAT admissions tests, and the ACCUPLACER Reading Comprehension, ASSET Reading Skills, and COMPASS Reading placement tests.

About half (53 percent) of postsecondary education institutions reported using some reading test for determining the need of entry-level students for remedial courses in reading. The range was from 9 percent for the ASSET Reading test to 22 percent for the COMPASS reading test.

The overall mean cut scores reported for the five standardized reading tests were:

- 18 on ACT Reading on a scale of 1 to 36
- 456 on SAT Critical Reading on a scale of 200 to 800
- 76 on ACCUPLACER Reading Comprehension on a scale of 20 to 120
- 41 on ASSET Reading Skills on a scale of 23 to 55
- 76 on COMPASS Reading on a scale of 1 to 99

**Variability of Cut Scores**

In addition to the mean cut score for each test, the overall range and 25th, 50th, and 75th percentiles were reported. The size of the interquartile range on each test for all institutions was converted to standard deviation units as a way of comparing the variability in cut scores across tests.

A key assumption in this survey is that each postsecondary education institution’s conception of the reading and mathematics knowledge and skills needed to be “just academically prepared” for credit-bearing entry-level courses is operationalized in the cut scores used.
The variability in cut scores is appreciable. It is smallest for ASSET Reading Skills, with one score point across the interquartile range and standard deviation of .15, and largest for COMPASS Algebra, with 26 score points across the interquartile range and standard deviation of 1.35.

The data in this analysis suggest that postsecondary education institutions across the nation do not hold a single, common conception of “just academically prepared.”

<table>
<thead>
<tr>
<th>Mathematics Tests</th>
<th>Interquartile Range in Standard Deviation Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>.57</td>
</tr>
<tr>
<td>SAT</td>
<td>.51</td>
</tr>
<tr>
<td>ACCUPLACER Elementary Algebra</td>
<td>.73</td>
</tr>
<tr>
<td>ACCUPLACER College-level Mathematics</td>
<td>.91</td>
</tr>
<tr>
<td>COMPASS Algebra</td>
<td>1.35</td>
</tr>
<tr>
<td>COMPASS College Algebra</td>
<td>.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reading Tests</th>
<th>Interquartile Range in Standard Deviation Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>.32</td>
</tr>
<tr>
<td>SAT</td>
<td>.44</td>
</tr>
<tr>
<td>ACCUPLACER Reading Comprehension</td>
<td>.41</td>
</tr>
<tr>
<td>ASSET Reading Skills</td>
<td>.15</td>
</tr>
<tr>
<td>COMPASS Reading</td>
<td>.48</td>
</tr>
</tbody>
</table>
I. Introduction

The survey of postsecondary education institutions’ use of tests and cut scores in college placement is one component of a larger research program of the National Assessment Governing Board (NAGB). This introductory section provides an overview of the Governing Board’s research program, the purpose of this study, the policy context, policy considerations in survey design, the survey methodology, and study limitations.

Overview of the Governing Board Program of Preparedness Research

Since 1969, the National Assessment of Educational Progress (NAEP) — also known as the Nation’s Report Card — has reported to the public on the academic achievement of United States elementary and secondary school students. The National Assessment is authorized and funded by Congress. Assessments are conducted in key subjects, such as reading, writing, mathematics, science, U.S. history, civics, geography, and the arts at grades 4, 8, and 12.

NAGB — an independent body of state and local educators, policymakers, technical experts, business leaders, parents, and the general public — oversees and sets policy for NAEP.

In 2004, the Governing Board received a report from a national blue-ribbon commission it had established. The Commission on NAEP 12th Grade Assessment and Reporting was composed of leaders in K-12 and postsecondary education, business, and the military. Their charge was “To review the current purpose, strengths, and weaknesses of 12th grade NAEP…” After more than a year of deliberation, the Commission made five recommendations. One key recommendation was to transform NAEP to enable it to report on the academic preparedness of 12th grade students for college and job training (National Commission on NAEP 12th Grade Assessment and Reporting, 2004).

The National Commission (2004), recognizing that 12th grade is the transition point to adult pursuits, stated that:

America needs to know how well prepared its high school seniors are to become productive citizens and to compete in a global economy — how well they can read, write and compute, and what they know about science, history, civics, and other important disciplines (p. 2).

As the only source of nationally representative data on student achievement at grade 12, the Commission concluded that NAEP is uniquely positioned to report on the academic preparedness of 12th graders.

The Governing Board accepted the Commission’s report and acted on it. The Board determined that questions of validity are at the heart of transforming NAEP from a measure of current student achievement to an indicator of academic preparedness. Specifically, was the NAEP content appropriate for measuring academic preparedness for college and job training? What evidence would be necessary to support statements about preparedness to be made in NAEP reports?

1 More information about the Governing Board is available at www.nagb.org or www.nagb.gov.
The first step in addressing validity questions was to determine whether the 12th grade NAEP assessment content in reading and mathematics is relevant and appropriate for measuring academic preparedness for college and job training. Consequently, the Governing Board contracted with the Achieve, Inc. to review the NAEP 12th grade assessment frameworks in reading and mathematics. The test frameworks define what will be measured and how it will be measured for each NAEP assessment, i.e., the assessment content.


In 2007, the Governing Board established the Technical Panel on 12th Grade Preparedness Research. The panel’s charge was to assist “the National Assessment Governing Board in planning research and validity studies that will enable...[NAEP] to report on the preparedness of 12th graders for postsecondary education and job training” (NAGB Technical Panel on 12th Grade Preparedness Research, 2009, p. iii). The panel’s recommendations were presented to the Governing Board in November 2008. The first phase of the research was to be conducted in connection with the 2009 NAEP 12th grade reading and mathematics assessments.

For the purpose of designing the research program, the NAGB Technical Panel (2009) recommended the following working definition of preparedness:

Preparedness for college refers to the reading and mathematics knowledge and skills necessary to qualify for placement into entry level college credit coursework without the need for remedial coursework in those subjects. Preparedness for workplace training refers to the reading and mathematics knowledge and skills needed to qualify for job training; it does not mean that the student is ready to be hired for a job (p. 3).

This working definition focuses on academic preparedness to qualify for entry, not on success in first year courses or completion of a certificate or degree.

The NAGB Technical Panel (2009) recommended studies in five areas:

- **Content alignment**, to determine the degree of overlap between NAEP and other relevant tests
- **Statistical linking**, to examine how student performance on NAEP compares with performance on other relevant tests
- **Judgmental standard-setting**, in which experts in college placement and occupational training identify the point(s) on the score scale for NAEP reporting that represents “just academically prepared”

---

2 Achieve, Inc. “...is a bipartisan, non-profit organization that helps states raise academic standards, improve assessments, and strengthen accountability to prepare all young people for postsecondary education, work, and citizenship.” www.achieve.org

3 Achieve, working through state leaders, established the American Diploma Project (ADP) Network in 2005 “...to improve postsecondary preparation by aligning high school standards, graduation requirements and assessment and accountability systems with the demands of college and careers.” www.achieve.org/adp-network
Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011

• **Benchmarking,** in which NAEP is administered to a representative group of interest, such as college freshmen or individuals newly enrolled in specific job training programs

• **Survey** of postsecondary education institutions’ use of tests and cut scores for entry-level college course placement

The intent was to conduct a range of types of studies, examine the degree to which results were mutually confirming, and, on the basis of that examination determine whether: (a) it is feasible for NAEP to report on 12th grade students’ academic preparedness, and (b) the results are sufficient to support valid statements about 12th grade students’ academic preparedness in NAEP reports.

Ultimately, the goal is to identify points on the NAEP 12th grade reading and mathematics reporting scales at or above which represent the knowledge and skills needed to qualify academically for placement into entry-level, credit-bearing college courses or job training, and below which indicate a likely need for remediation.

It is important to note that NAEP, by law, does not provide individual student results, only group results for the nation, states, and 21 urban districts. The NAEP results are also disaggregated by race, ethnicity, gender, income level, and for students with disabilities and English language learners. None of this will change as a result of the preparedness research program. The intention is solely to make NAEP 12th grade results more relevant and useful as a national indicator of academic preparedness for college and job training.

This report provides findings from the fifth type of research study cited above: the survey of postsecondary education institutions’ use of tests in entry-level college placement for credit-bearing versus remedial/developmental courses.

**Purpose of the Survey**

The primary purpose of this survey is to provide a source of nationally representative data for use as potential validity evidence for NAEP reporting on 12th grade student’s academic preparedness for college. The survey’s central two-part research question is:

1. Which national standardized tests are used by postsecondary education institutions to determine the need of entry-level students for remedial/developmental instruction in reading or mathematics?

2. What are the cut scores on those national standardized tests below which students are deemed to need remedial/developmental instruction in reading or mathematics and at or above which students would be just academically qualified for entry-level credit-bearing college courses?

The survey is intended as a component of the Governing Board’s program of preparedness research. However, because recent research on the use of tests in college placement decisions has been focused at the state level or presented as case studies (e.g., Brown & Niemi, 2007; Belfield & Crosta, 2012; Sommo, Mayer, Rudd, & Cullinan, 2012), the nationally representative aspect of the survey results fills an information gap that may be of wider interest to K-12 and postsecondary education policymakers and researchers.

**Policy Context**

The Governing Board’s initiative to transform NAEP into an indicator of academic preparedness for college and job training comes amid wider

---

4 In this report, the terms “remedial/developmental” and “remedial” are used synonymously.
recognition by state and national leaders of the need to ensure that 12th grade students leave high school “college and career ready.”

In 2008, the National Governors Association and the Council of Chief State School Officers joined together to develop the Common Core State Standards Initiative (CCSSI). These are K-12 education standards for mathematics and English language arts that:

[D]efine the knowledge and skills students should have within their K-12 education careers so that they will graduate high school able to succeed in entry-level, credit-bearing academic college courses and in workforce training programs (“About the Standards,” 2012, para. 4).

As of October 2012, the Common Core State Standards for mathematics and English language arts had been adopted formally by 45 states and 3 territories (CCSSI, “In the States”). Two consortia of states are developing assessments aligned to the Common Core State Standards.1

In 2010, the Obama Administration proposed that college- and career-ready standards be included in the reauthorization of the Elementary and Secondary Education Act (White House, Office of the Press Secretary, 2010). The U.S. Senate Committee on Health, Education, Labor and Pensions (2011) approved a bill to reauthorize the Elementary and Secondary Education Act that “makes college- and career-readiness for all students a central goal” (p. 20).

Similarly, the Business Coalition for Student Achievement (2012), coordinated by the Business Roundtable and the U.S. Chamber of Commerce, has stated the U.S. business community’s commitment to “Ensuring that all students graduate academically prepared for college, citizenship and the 21st century workplace…” The basis for this commitment is their perspective that “improving the performance of the K-12 education system in the United States is necessary to provide a strong foundation for both U.S. competitiveness and for individuals to succeed in our rapidly changing world.”

In addition to the view that inadequate achievement erodes individual opportunity and global economic competitiveness, a recent report prepared by the Independent Task Force on U.S. Education Reform and National Security of the Council on Foreign Relations (2012), co-chaired by Joel Klein and Condoleezza Rice, argues that inadequate achievement is a threat to national security.

**Policy Considerations for the Survey**

In conceptualizing the survey design and developing the survey items, a number of key policy issues and assumptions were addressed.

**Admission Versus Placement**

The first issue involved distinguishing between policies for admission to postsecondary education institutions and policies for placement into first-year courses, whether credit-bearing or remedial/ developmental. The intention was to avoid conflating admissions policy and placement policy, because they are not necessarily synonymous. Many postsecondary institutions admit students who may need remediation. Highly selective institutions generally do not admit students who need remediation; they limit admission to students whose academic proficiency often exceeds “just academically prepared” to an exceptional degree. Accordingly, test cut scores used for admission would not necessarily be an indicator of “just academically prepared.”

---

1 The consortia are the Smarter Balanced Assessment Consortium (www.smarterbalanced.org) and the Partnership for Assessment of Readiness for College and Careers (www.parcconline.org).
The object of the Governing Board’s preparedness research is to locate the points on the NAEP 12th grade reading and mathematics scales:

- at or above which represent the knowledge and skills needed to be “just academically prepared” for entry-level credit-bearing coursework;
- below which represent a likely need for remediation.

Therefore, the choice was made to focus on policies used for placement, because these policies are assumed to be more closely aligned with the institution’s conceptualization of being “just academically prepared” for college level work. This distinction was also important because some tests (e.g., ACT and SAT) could be used both for admissions and placement.

**Target Academic Subjects**

The second key issue was the choice of reading and mathematics as the subjects of interest for the survey. Writing, for example, which is also assessed by NAEP, could have been a candidate for inclusion in the survey. Along with reading and mathematics, adequate writing skills are necessary for college-level work and inadequate skills a prescription for remediation. However, the Governing Board’s initial set of preparedness research studies was designed solely in connection with the NAEP 12th grade reading and mathematics assessments. To minimize burden on respondents, the survey asked for information that was essential for the Board’s immediate research needs.

**Other Criteria Used for Placement**

The third issue considered was that multiple factors, such as high school course taking, class rank, teacher recommendations, and grade point average, may be a factor in placement. In each stage of survey development activities, some postsecondary participants strongly urged reporting these other factors, since they viewed doing so as providing a more comprehensive picture of the information used in evaluating students’ preparedness for college level courses. While the focus of the survey remained the institution’s conceptualization of necessary reading and mathematics skills in relation to a specific score on a test, survey respondents were afforded the opportunity to provide information about ancillary data that are considered by their institution in determining student need for remediation.

**Preparedness Versus Readiness**

The fourth key issue was the distinction between academic preparedness for college and “college readiness.” College readiness is a much broader concept than academic preparedness. Conley (2007), for example, has developed a conception of college readiness that includes time management, persistence, disciplined study habits, and the ability to negotiate college bureaucratic procedures, as well as performing the higher order intellectual tasks generally expected of entry-level college students. As noted above, the Governing Board’s working definition is limited to academic preparedness. Accordingly, academic preparedness, rather than “readiness,” is the focus of the survey.

**One Survey Form or Two?**

The fifth key issue was whether to have a single survey form or separate forms for 2-year institutions and for 4-year institutions. For the sake of clarity and simplicity, the decision was made to have separate forms with minor variances to account for differences in the respective missions of 2-year and 4-year institutions (e.g., the multiple roles played by 2-year institutions as a path to a college degree, job training, and adult education versus the target of a college degree for 4-year institutions).
II. Survey Development, Sample, and Data Collection Methodology

Survey Development

The survey forms were developed through a multi-stage process involving expert review and testing with potential survey respondents. A draft survey developed under preliminary research conducted by the Governing Board was refined by survey development specialists at Westat, the Governing Board's contractor for the conduct of the survey. The new draft was reviewed by the contractor's Technical Review Panel for this project and further refined. The refined draft was piloted with six institutions to assess the clarity of the instructions and questions, the time needed to complete the survey, whether there were other extant sources of the requested data, and the most efficient way to identify appropriate respondents within institutions.

It was anticipated that there would be a high degree of variability among institutions in terms of the office or individual most knowledgeable about the policy on tests and cut scores for placement and that identifying the appropriate respondent would be a challenge in conducting the survey. This was confirmed at every stage of survey development.

After the pilot test, another exploratory study was conducted with eight volunteering institutions. Key among the questions addressed, where admission tests like the ACT and SAT were also used to identify students in need of remediation, was how to avoid errors due to respondents confusing the different uses of the tests.

The next phase of survey development involved review of the draft survey by an external panel of content experts in a half-day structured discussion by teleconference and then a small-scale field test with a diverse sample of 120 postsecondary institutions. This work confirmed the use of separate forms for 2-year and 4-year institutions and the president's office of the respective institutions as the place to begin the process of identifying the appropriate respondent for the survey. As a final step for survey quality and clarity, a small-scale cognitive lab and usability study was conducted with participants from nine institutions.

Sample

The sample of approximately 1,670 Title IV postsecondary education institutions was drawn from the 2009–10 Institutional Characteristics (IC) component of the Integrated Postsecondary Education Data System (IPEDS) maintained by the National Center for Education Statistics (NCES). It included 2-year and 4-year degree-granting institutions in the 50 states and the District of Columbia that offer at least an undergraduate degree.

The 1,670 institutions in the sample was reduced, yielding a final sample of about 1,560, due to five institutions having closed and approximately 100 not meeting one of the following eligibility criteria in fall 2011:

- Enrollment of entering students in a degree program designed to transfer to a 4-year institution (applies to 2-year institutions only); or
- Enrollment of entering students in an undergraduate degree program in the liberal arts and sciences (applies to 4-year institutions only).

---

6 For additional detail, see Technical Report in Appendix A.
7 A cognitive lab and usability study involves in-depth interviews with individuals representing likely respondents, to assess the degree of agreement between the intended meaning of the survey instructions and items and the meaning as interpreted by the respondents.
Stratification was by level (four-year, two-year), control (public, private not-for-profit, private for-profit), highest level of offering (doctor’s/dottoral-professional practice, master’s, bachelor’s, less than bachelor’s), and total fall enrollment. Within each stratum, institutions were sorted by region (Northeast, Southeast, Central, West) and by level of minority enrollment (high black enrollment, high total minority enrollment but not high black enrollment, and low minority enrollment).

A weighted response rate of 81 percent was achieved. Weighted response rates by sector were:

<table>
<thead>
<tr>
<th>Institution Type</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-year public institutions</td>
<td>91%</td>
</tr>
<tr>
<td>2-year public institutions</td>
<td>89%</td>
</tr>
<tr>
<td>4-year private, not-for-profit institutions</td>
<td>84%</td>
</tr>
<tr>
<td>2-year private, not-for-profit institutions</td>
<td>74%</td>
</tr>
<tr>
<td>4-year private for-profit institutions</td>
<td>68%</td>
</tr>
<tr>
<td>2-year private for-profit institutions</td>
<td>56%</td>
</tr>
</tbody>
</table>

A non-response bias analysis was conducted and non-response adjusted weights were estimated. Based on the non-response bias analysis, it appears that the estimates reported in the study, using the nonresponse adjusted weights, are nationally representative.  

Data Collection Methodology

Separate survey forms were used for 2-year and for 4-year institutions, with only minor differences in the forms. The primary difference is in the description of the target population for the institution’s general policy about entry-level placement into remedial/developmental or credit-bearing courses.

Two-year institutions serve students on a wide array of diverse paths, from taking a single course of interest, to obtaining training for an occupation, to enrolling in a program that would transfer to a 4-year institution for a bachelor’s degree. Depending on the path chosen by the student, the 2-year institution may employ different criteria for determining student need for remedial courses. Therefore, and consistent with the eligibility criteria described above, the 2-year survey asked respondents to report “based on the tests your institution uses to evaluate entering students who are pursuing a degree program that is designed to transfer to a four-year institution.” For 4-year institutions, to capture the general policy for entry-level students, as opposed to students in programs that may have additional and/or more rigorous enrollment requirements, the survey asked respondents to report “based on the tests your institution uses to evaluate entering students who are enrolled in an undergraduate degree program in the liberal arts and sciences.”

Letters with background information on the survey were sent to the office of the president at each sampled institution in August 2011. The letters identified the survey sponsor (the National Assessment Governing Board), explained the purpose of the survey, and asked the president to identify an appropriate respondent.

Survey materials (background information, instructions, and the survey) were sent directly to the identified respondents via email, mail, or fax. Four options were offered for completing the survey: online, and via mail, email, or fax. Follow-up for survey recruitment, nonresponse, and data clarification was conducted via telephone and email by experienced, trained interviewers from September 2011 through January 2012.

---

See Technical Report in Appendix A.

See pages A-1 through B-10, Technical Report in Appendix A.

See Technical Report in Appendix A.
III. Caveats and Limitations on Interpretation

The work that was done in developing and executing the survey consistently pointed to the complexity and variability in the use of tests and cut scores by postsecondary education institutions in determining student need for remediation. The procedures in use across the country are not monolithic. However, the data in this report are intended to reflect the institutions’ general policies; they do not necessarily convey the complexity of the determination process as it applies to individual students. Readers are therefore cautioned to interpret the data in this report in light of the following information.

Use of Multiple Tests

Survey development activities also confirmed that individual institutions may use more than one test, either alone or in combination, in determining student need for remedial/developmental instruction. In the cases where respondents reported using more than one test, project interviewers followed up with the respondent to clarify which test or tests, and associated cut score, signified the institution’s policy for distinguishing those students who were “just academically prepared” from those needing remedial/developmental instruction.

In cases where performance on both admissions tests (i.e. ACT and SAT scores) and placement tests were used, respondents confirmed which tests and cut scores represented the institution’s general policy for placement. In other cases where multiple tests were reported, respondents confirmed whether only one test and cut score was used to make this distinction and whether the other tests and cut scores were used for determining the level of remediation needed. In most instances, these institutions offer graduated levels of remediation, reflecting higher and lower gaps between the institution’s conception of “just academically prepared” and the student’s test performance. Thus, multiple cut scores on a test or multiple subtests may be used to identify the level of the remedial course most appropriate for the student.

Remedial/Developmental Coursework: Required or Recommended?

During survey development, discussions with the technical experts and the respondents in the small-scale studies confirmed that there is variability in the use of tests in placement procedures for remedial courses. In some institutions, performance below a certain cut score leads to a requirement for students to take remedial courses. In other institutions, remedial coursework is recommended, but students are permitted to take credit-bearing courses, either at their own discretion or the discretion of the institution. If different scores were used for either requiring or recommending remediation, the survey asked respondents to report the highest score used.
For these reasons, the survey asked respondents to report the highest score below which the need for remedial course work is indicated and at or above which students would be placed into entry-level credit-bearing courses.

**General Policy Versus Individual Student Placement Determination**

The experts and respondents in the small scale studies who participated during the survey development process emphasized that the placement determination of entry-level students was often handled on a case-by-case basis, with a process employing personal advising by an admission or placement counselor, taking into account test performance and other factors. The aim of this approach is to maximize the chances for the individual student’s success. Examples of this approach can be found among the Achieving the Dream “Leader Colleges,” but certainly are not limited to these institutions.

It is important to be aware of such individual student-centered approaches to student placement, while recognizing that the narrow focus of this survey is on postsecondary education institutions’ conceptualization of being “just academically prepared” as operationalized by performance on a test, for the purpose of the Governing Board’s program of preparedness research.

---

11 A list of Achieving the Dream “Leader Colleges” can be found at: [www.achievingthedream.org/initiatives/PRESS](http://www.achievingthedream.org/initiatives/PRESS).
IV. Findings

The survey findings are presented first for mathematics and then for reading. For both subjects, national estimates are followed by breakouts by institution level and type. The tests displayed in the tables are the most relevant for statistical linking with NAEP because of the level of their content and their frequency of use. See Appendix C for the frequency of use of all tests for which data were collected.

Mathematics
During survey development, four nationally available standardized testing programs were consistently identified by advisors and preliminary study participants as being used in making determinations about student need for remedial/developmental instruction in mathematics. The four testing programs are the ACT and SAT admissions tests, and the ACCUPLACER and COMPASS placement tests. The placement test programs each have specific tests that are used for assessing different levels of student need for remediation.

Respondents were also given the opportunity to report on whether other tests are used. They were not asked to report the cut scores on those tests because it would not be practical or useful to link performance on NAEP with performance on those tests. For example, many of the tests could be developed by the institution itself or a state-wide postsecondary education entity and, therefore, not permit national-level comparisons.

Frequency of Use of Mathematics Tests
Table 1 displays the frequency of use of various tests. Seventy-one percent of postsecondary education institutions reported using some mathematics test for determining the need of entry-level students for remedial courses in mathematics. The range was from 4 percent for the COMPASS College Algebra test to 23 percent for ACT Mathematics.

Twenty percent of institutions reported using the COMPASS Algebra test; 17 percent the SAT Mathematics test; 16 percent the ACCUPLACER Elementary Algebra test; and 5 percent the ACCUPLACER College-level Mathematics test. Twenty-two percent of respondents reported using other mathematics tests than the standardized national tests for which data were collected.

Considering differences by institution type, 100 percent of 2-year public institutions and 85 percent of 4-year public institutions reported using some mathematics test for placement. Forty to 58 percent of private 2-year and 4-year institutions use some mathematics test for placement. Details on the frequency of use of specific national standardized tests or other tests by institution level and type are displayed in Table 1.
Table 1  Estimated number of postsecondary institutions in the population and percentage of institutions using selected mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Estimated number of institutions in the population</th>
<th>Percentage of institutions using any mathematics test</th>
<th>Percentage of institutions using specific mathematics tests</th>
<th>Other mathematics tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACT</td>
<td>SAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mathematics</td>
<td>Mathematics</td>
</tr>
<tr>
<td>All institutions</td>
<td>3,650</td>
<td>71</td>
<td>23</td>
<td>17</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>1,470</td>
<td>80</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>4-year</td>
<td>2,180</td>
<td>65</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>970</td>
<td>100</td>
<td>32</td>
<td>17</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>500</td>
<td>40</td>
<td>—</td>
<td>3!</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>620</td>
<td>85</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>1,230</td>
<td>58</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>330</td>
<td>52</td>
<td>9!</td>
<td>10</td>
</tr>
</tbody>
</table>

* Interpret data with caution; the coefficient of variation is greater than or equal to 30 percent but less than 50 percent.

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

**NOTE:** Details for the number of institutions may not sum to totals because of rounding.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.
Mean Cut Scores on Mathematics Tests by Institution Level and Type

Overall Estimates
Table 2 displays the mean mathematics test scores reported by postsecondary education institutions that demarcate the point at or above which students were deemed to be academically prepared for credit-bearing entry-level mathematics courses and below which student need for remediation was indicated.

The overall mean cut scores reported for the most frequently used standardized national mathematics tests are:

- 19 on ACT mathematics on a scale of 1 to 36
- 471 on SAT mathematics on a scale of 200 to 800
- 70 on ACCUPLACER Elementary Algebra on a scale of 20 to 120
- 57 on ACCUPLACER College-Level Mathematics on a scale of 20 to 120
- 49 on COMPASS Algebra on a scale of 1 to 99
- 43 on COMPASS College Algebra on a scale of 1 to 99

Statistically Significant Differences by Institution Level and Institution Type
There were five tests for which comparisons of statistical significance could be made by institution level and by institution type: the ACT and SAT mathematics tests, the ACCUPLACER Elementary Algebra and College-level Mathematics tests, and the COMPASS Algebra test.

Table 2 shows the following differences by institution level:

- The mean scores for 2-year institutions were higher than those for 4-year institutions for each of the ACT and SAT mathematics tests and the COMPASS Algebra test.
- The mean score for 2-year institutions was significantly lower than the mean score for 4-year institutions on the ACCUPLACER Elementary Algebra test.

Similar results were found by institution type:

- The mean scores for 2-year public institutions were higher than those for 4-year public institutions for each of the ACT and SAT mathematics tests and the COMPASS Algebra test.
- The mean score for 2-year public institutions was lower than the mean score for 4-year public institutions on the ACCUPLACER Elementary Algebra test.
- Public 2-year institutions had higher mean scores than private not-for-profit 4-year institutions for each of the ACT and SAT mathematics tests.
Table 2  Mean mathematics test scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>ACT Mathematics</th>
<th>ACT Mathematics</th>
<th>ACCUPLACER Elementary Algebra</th>
<th>ACCUPLACER College-Level Mathematics</th>
<th>COMPASS Algebra</th>
<th>COMPASS College Algebra</th>
</tr>
</thead>
<tbody>
<tr>
<td>All institutions †</td>
<td>19</td>
<td>471</td>
<td>70</td>
<td>57</td>
<td>49</td>
<td>43</td>
</tr>
<tr>
<td>Institution level ‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>20</td>
<td>483</td>
<td>68</td>
<td>55</td>
<td>51</td>
<td>43</td>
</tr>
<tr>
<td>4-year</td>
<td>19*</td>
<td>467*</td>
<td>73*</td>
<td>59</td>
<td>44*</td>
<td>—</td>
</tr>
<tr>
<td>Institution type §</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>20</td>
<td>485</td>
<td>69</td>
<td>55</td>
<td>51</td>
<td>43</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>19**</td>
<td>474**</td>
<td>73**</td>
<td>57</td>
<td>45**</td>
<td>—</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>19**</td>
<td>459**</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

† Data for private for-profit 4-year institutions and all private 2-year institutions are included in the totals but are not shown by institution type because of small cell sizes.

‡ Indicates a significant difference between 2-year and 4-year institutions reporting by institution level.

§ Indicates a significant difference between 2-year public and the comparison group by institution type.

NOTE: For each test, mean scores are based on the number of institutions reporting the use of the test to evaluate students for remedial or developmental mathematics courses in fall 2011.

Variability of Cut Scores on Mathematics Tests by Institution Level and Type

Tables 3 and 4 display the range and percentiles of cut scores at or above which postsecondary education institutions reported students were deemed to be academically prepared and below which the need for remedial courses was indicated. The data suggest an appreciable amount of variability in the cut scores reported by postsecondary institutions, as evidenced by the overall cut score ranges and the interquartile ranges for the tests displayed in the tables. Because the overall cut score ranges appear extreme in at least several of the cases, the analysis below focuses on the interquartile ranges.

The ACT mathematics test has a score scale of 1 to 36 and a standard deviation of 5.3. For the ACT mathematics test, the cut scores reported range from 10 to 25. For all institutions, the 25th, 50th, and 75th percentiles are 17, 19, and 20 respectively, with an interquartile range of 3, or .57 of a standard deviation.

The SAT mathematics test has a score scale of 200-800 with a standard deviation of 117. The range of cut scores reported for the SAT mathematics test is 330 to 600. For all institutions, the 25th, 50th, and 75th percentiles are 440, 470, and 500 respectively, with an interquartile range of 60, or .51 of a standard deviation.

The ACCUPLACER mathematics tests have a score scale of 20 to 120. The standard deviation is 27.3 for the Elementary Algebra test and 19.8 for the College-level Mathematics test.

Cut scores reported for the ACCUPLACER Elementary Algebra test ranged from 25 to 110. For all institutions, the 25th, 50th, and 75th percentiles are 61, 71, and 81 respectively, with an interquartile range of 20, or .73 of a standard deviation.

For the ACCUPLACER College-level Mathematics test, the range was 30 to 93. For all institutions, the 25th, 50th, and 75th percentiles are 45, 51, and 63 respectively, with an interquartile range of 18, or .91 of a standard deviation.

The COMPASS mathematics tests have a score range of 1 to 99. The standard deviation is 19.2 for the COMPASS Algebra test and 20.1 for the COMPASS College Algebra test. For the COMPASS Algebra test, the reported cut scores ranged from 15 to 86. For all institutions, the 25th, 50th, and 75th percentiles are 38, 45, and 64 respectively, with an interquartile range of 26, or 1.35 standard deviations.

Cut scores reported for the COMPASS College Algebra test ranged from 20 to 76. For all institutions, the 25th, 50th, and 75th percentiles are 36, 40, and 46 respectively, with an interquartile range of 10, or .50 of a standard deviation.

<table>
<thead>
<tr>
<th>Mathematics Tests</th>
<th>Interquartile Range in Standard Deviation Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>.57</td>
</tr>
<tr>
<td>SAT</td>
<td>.51</td>
</tr>
<tr>
<td>ACCUPLACER Elementary Algebra</td>
<td>.73</td>
</tr>
<tr>
<td>ACCUPLACER College-level Mathematics</td>
<td>.91</td>
</tr>
<tr>
<td>COMPASS Algebra</td>
<td>1.35</td>
</tr>
<tr>
<td>COMPASS College Algebra</td>
<td>.50</td>
</tr>
</tbody>
</table>
### Table 3
Ranges of mathematics test scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Ranges of scores for mathematics tests</th>
<th>ACT</th>
<th>SAT&lt;sup&gt;1&lt;/sup&gt;</th>
<th>ACCUPLACER</th>
<th>COMPASS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>Elementary Algebra</td>
<td>College-Level Mathematics</td>
</tr>
<tr>
<td>All institutions&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>Lowest score</td>
<td>Highest score</td>
<td>Lowest score</td>
<td>Highest score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 25</td>
<td>330 600</td>
<td>25 110</td>
<td>30 93</td>
</tr>
<tr>
<td>Institution level&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>2-year</td>
<td>10 25</td>
<td>380 600</td>
<td>25 110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4-year</td>
<td>12 24</td>
<td>330 600</td>
<td>29 109</td>
</tr>
<tr>
<td>Institution type&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>Public 2-year</td>
<td>10 25</td>
<td>380 600</td>
<td>25 110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public 4-year</td>
<td>12 24</td>
<td>330 600</td>
<td>34 109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private not-for-profit 4-year</td>
<td>14 24</td>
<td>340 590</td>
<td>— —</td>
</tr>
</tbody>
</table>

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.
<sup>1</sup> Some institutions reported interpolated SAT mathematics scores. Where applicable, the scores were rounded to the nearest ten for presentation in this table.
<sup>2</sup> Data for private for-profit 4-year institutions and all private 2-year institutions are included in the totals but are not shown by institution type because of small cell sizes.

**SOURCE**: National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.

### Table 4
Percentiles for mathematics test cut scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentiles for mathematics test cut scores</th>
<th>ACT</th>
<th>SAT&lt;sup&gt;1&lt;/sup&gt;</th>
<th>ACCUPLACER</th>
<th>COMPASS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>Elementary Algebra</td>
<td>College-Level Mathematics</td>
</tr>
<tr>
<td>All institutions&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>25&lt;sup&gt;th&lt;/sup&gt;</td>
<td>50&lt;sup&gt;th&lt;/sup&gt;</td>
<td>75&lt;sup&gt;th&lt;/sup&gt;</td>
<td>25&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17 19 20</td>
<td>440 470 500</td>
<td>61 71 81</td>
<td>45 51 63</td>
</tr>
<tr>
<td>Institution level&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>2-year</td>
<td>18 19 21</td>
<td>450 480 500</td>
<td>57 67 76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4-year</td>
<td>17 18 19</td>
<td>440 460 500</td>
<td>62 72 84</td>
</tr>
<tr>
<td>Institution type&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td>Public 2-year</td>
<td>18 19 21</td>
<td>450 480 510</td>
<td>61 70 76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public 4-year</td>
<td>18 19 19</td>
<td>450 460 500</td>
<td>63 72 82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Private not-for-profit 4-year</td>
<td>17 18 19</td>
<td>430 460 495</td>
<td>— — —</td>
</tr>
</tbody>
</table>

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.
<sup>1</sup> Some institutions reported interpolated SAT mathematics scores. Where applicable, the scores were rounded to the nearest ten for presentation in this table.
<sup>2</sup> Data for private for-profit 4-year institutions and all private 2-year institutions are included in the totals but are not shown by institution type because of small cell sizes.

**SOURCE**: National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.
### Other Criteria Used by Postsecondary Institutions

To round out the factors that are considered in placement determinations by postsecondary education institutions, respondents were asked to report on the use of criteria other than the tests reported in the preceding subsection. Overall, 21 percent of institutions reported using some other criteria. Eleven percent reported using Advanced Placement or International Baccalaureate scores; 10 percent used high school grades; 10 percent considered highest mathematics course taken; 3 percent looked at high school graduation tests or end-of-course tests, or faculty recommendations; and 2 percent considered other criteria than these.

Similar patterns were found by institution level. Seven to twelve percent of 2-year and 4-year institutions reported using Advanced Placement or International Baccalaureate scores, high school grades, and highest mathematics course taken. Two to four percent reported using high school graduation tests or end-of-course tests, faculty recommendations, or other criteria. Again, similar patterns were found for public 2-year, public 4-year, and private not-for-profit 4-year postsecondary education institutions.

### Table 5 Estimated percentage of institutions using criteria other than postsecondary mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentage of institutions using any criteria other than mathematics tests</th>
<th>Percentage of institutions using specific mathematics tests</th>
<th>Faculty recommendation</th>
<th>Other criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution level</td>
<td></td>
<td>High school graduation tests or end-of-course tests</td>
<td>High school grades (including grade point average)</td>
<td>Highest school mathematics course completed</td>
</tr>
<tr>
<td>All institutions</td>
<td>21</td>
<td>3</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2-year</td>
<td>20</td>
<td>3</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>4-year</td>
<td>22</td>
<td>3</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td>High school graduation tests or end-of-course tests</td>
<td>High school grades (including grade point average)</td>
<td>Highest school mathematics course completed</td>
</tr>
<tr>
<td>Public 2-year</td>
<td>27</td>
<td>4</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>27</td>
<td>5</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>25</td>
<td>4!</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

† Interpret data with caution; the coefficient of variation is greater than or equal to 30 percent but less than 50 percent.

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

**NOTE:** Details for the number of institutions may not sum to totals because of rounding.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.
Reading

During survey development, five nationally available standardized testing programs were consistently identified by advisors and preliminary study participants as being used in making determinations about student need for remediation for reading. The five testing programs are the ACT and SAT admissions tests, and the ACCUPLACER, ASSET, and COMPASS placement tests. Respondents also were given the opportunity to report on whether other tests are used, but they were not asked to report the cut scores on those tests, because it would not be feasible to make national comparisons or link performance on NAEP with performance on these tests.

Frequency of Use of Reading Tests

Table 6 displays the frequency of use of the various reading tests. About half (53 percent) of postsecondary education institutions reported using some reading test for determining the need of entry-level students for remedial courses in reading. The range was from 9 percent for the ASSET Reading Skills test to 22 percent for the COMPASS Reading test.

Nineteen percent of institutions reported using the ACCUPLACER Reading Comprehension test; 16 percent reported using the ACT Reading test; and 11 percent reported using the SAT Critical Reading test. Ten percent of respondents reported using other reading tests.

Table 6 Estimated percentage of institutions using selected reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentage of institutions using any reading test</th>
<th>Percentage of institutions using specific reading tests</th>
<th>Other reading tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>All institutions</td>
<td>53</td>
<td>16 11 19 9 22 10</td>
<td></td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>73</td>
<td>21 10 29 19 43 12</td>
<td>12</td>
</tr>
<tr>
<td>4-year</td>
<td>39</td>
<td>13 12 12 2 9 8</td>
<td>8</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>94</td>
<td>28 14 39</td>
<td>61 10</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>33</td>
<td>61 31 8</td>
<td>— — 16!</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>51</td>
<td>18 16 18</td>
<td>6 21 12</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>31</td>
<td>14 13 5</td>
<td># 5 5</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>44</td>
<td>— — 26</td>
<td>— — 13</td>
</tr>
</tbody>
</table>

# Rounds to zero.

! Interpret data with caution; the coefficient of variation is greater than or equal to 30 percent but less than 50 percent.

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

NOTE: Details for the number of institutions may not sum to totals because of rounding or suppressed data.

Considering differences by institution type, 94 percent of 2-year public institutions compared with 51 percent of 4-year public institutions used some reading test for placement. In addition, 31 to 44 percent of private 2-year and 4-year institutions reported using some reading test for placement. Details on the frequency of use of the five specific national standardized reading tests or other tests by institution level and type are displayed in Table 6.

**Mean Cut Scores on Reading Tests by Institution Level and Type**

**Overall Estimates**
Table 7 displays the mean reading test cut scores reported by postsecondary education institutions that demarcate the point at or above which students were deemed to be academically prepared for credit-bearing entry-level courses that require college level reading and below which student need for remedial/developmental instruction was indicated.

The overall mean cut scores reported for the five standardized reading tests were:

- 18 on ACT Reading on a scale of 1 to 36
- 456 on SAT Critical Reading on a scale of 200 to 800
- 76 on ACCUPLACER Reading Comprehension on a scale of 20 to 120
- 41 on ASSET Reading Skills on a scale of 23 to 55
- 76 on COMPASS Reading on a scale of 1 to 99

**Statistically Significant Differences by Institution Level and Institution Type**

For the five tests in Table 7, comparisons of statistical significance were made by institution level and by institution type.

By institution level, the mean scores for 2-year institutions were higher than for 4-year institutions for the ACT and the SAT reading tests.

By institution type, the only instance of statistically significant differences was for the SAT Critical Reading test – the mean cut score for 2-year public institutions was higher than the scores for 4-year public institutions and 4-year private not-for-profit institutions.
<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Mean reading test scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
</tr>
<tr>
<td>All institutions³</td>
<td>18</td>
</tr>
<tr>
<td>Institution level³</td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>19</td>
</tr>
<tr>
<td>4-year</td>
<td>18*</td>
</tr>
<tr>
<td>Institution type³</td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>18</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>18</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>18</td>
</tr>
</tbody>
</table>

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

³ Data for private for-profit 4-year institutions and all private 2-year institutions are included in the totals but are not shown by institution type because of small cell sizes.

* Indicates a significant difference between 2-year and 4-year institutions reporting by institution level.

** Indicates a significant difference between 2-year public and the comparison group by institution type.

NOTE: For each test, mean scores are based on the number of institutions reporting the use of the test to evaluate students for remedial or developmental reading courses in fall 2011.

Variability of Cut Scores on Reading Tests by Institution Level and Type

Tables 8 and 9 display the range and percentiles, respectively, of cut scores at or above which postsecondary education institutions reported students were deemed to be academically prepared and below which the need for remedial courses was indicated. The data suggest an appreciable amount of variability in the cut scores reported by postsecondary institutions, as evidenced by the overall cut score ranges and the interquartile ranges for the tests displayed in the tables. Because the overall cut score ranges appear extreme in at least several of the cases, the analysis below focuses on the interquartile ranges.

The ACT Reading test has a score scale of 1 to 36 with a standard deviation of 6.2. For the ACT reading test, the cut scores range from 14 to 25. For all institutions, the 25\textsuperscript{th}, 50\textsuperscript{th}, and 75\textsuperscript{th} percentiles are 17, 18, and 19 respectively, with an interquartile range of 2, or about .32 of a standard deviation.

The SAT Critical Reading test has a score scale of 200 to 800 with a standard deviation of 114. The range of cut scores reported for the SAT reading test is 320 to 750. For all institutions, the 25\textsuperscript{th}, 50\textsuperscript{th}, and 75\textsuperscript{th} percentiles are 430, 450, and 480 respectively, with an interquartile range of 50, or about .44 of a standard deviation.

The ACCUPLACER Reading Comprehension test has a score scale of 20 to 120 with a standard deviation of 22. Cut scores reported for the ACCUPLACER Reading Comprehension test range from 50 to 106. For all institutions, the 25\textsuperscript{th}, 50\textsuperscript{th}, and 75\textsuperscript{th} percentiles are 71, 78, and 80 respectively, with an interquartile range of 9, or .41 of a standard deviation.

The ASSET Reading Skills test has a score scale of 23 to 53 with a standard deviation of 6.5. The range of scores reported for the ASSET Reading Skills test is 35 to 47. For all institutions, the 25\textsuperscript{th}, 50\textsuperscript{th}, and 75\textsuperscript{th} percentiles are 40, 40, and 41 respectively, with an interquartile range of 1, or about .15 of a standard deviation.

The Compass Reading test has a score scale of 1 to 99 with a standard deviation of 16.7. For the COMPASS Reading test, the reported cut scores range from 19 to 91. For all institutions, the 25\textsuperscript{th}, 50\textsuperscript{th}, and 75\textsuperscript{th} percentiles are 73, 79, and 81 respectively, with an interquartile range of 8, or .48 of a standard deviation.

<table>
<thead>
<tr>
<th>Reading Tests</th>
<th>Interquartile Range in Standard Deviation Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT</td>
<td>.32</td>
</tr>
<tr>
<td>SAT</td>
<td>.44</td>
</tr>
<tr>
<td>ACCUPLACER Reading</td>
<td>.41</td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
</tr>
<tr>
<td>ASSET Reading Skills</td>
<td>.15</td>
</tr>
<tr>
<td>COMPASS Reading</td>
<td>.48</td>
</tr>
</tbody>
</table>
### Table 8
Ranges of reading test scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>ACT Reading</th>
<th>SAT Reading</th>
<th>ACCUPLACER Reading Comprehension</th>
<th>ASSET Reading Skills</th>
<th>COMPASS Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowest score</td>
<td>Highest score</td>
<td>Lowest score</td>
<td>Highest score</td>
<td>Lowest score</td>
</tr>
<tr>
<td>All institutions*</td>
<td>14</td>
<td>25</td>
<td>320</td>
<td>750</td>
<td>50</td>
</tr>
<tr>
<td>2-year</td>
<td>14</td>
<td>25</td>
<td>340</td>
<td>550</td>
<td>50</td>
</tr>
<tr>
<td>4-year</td>
<td>14</td>
<td>25</td>
<td>320</td>
<td>750</td>
<td>52</td>
</tr>
<tr>
<td>Public 2-year</td>
<td>14</td>
<td>25</td>
<td>340</td>
<td>550</td>
<td>50</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>14</td>
<td>25</td>
<td>320</td>
<td>750</td>
<td>55</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>14</td>
<td>21</td>
<td>340</td>
<td>750</td>
<td>–</td>
</tr>
</tbody>
</table>

* Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

** Some institutions reported interpolated SAT mathematics scores. Where applicable, the scores were rounded to the nearest ten for presentation in this table.

*** Data for private for-profit 4-year institutions and all private 2-year institutions are included in the totals but are not shown by institution type because of small cell sizes.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.

### Table 9
Percentiles for reading test cut scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>ACT Reading</th>
<th>SAT Reading</th>
<th>ACCUPLACER Reading Comprehension</th>
<th>ASSET Reading Skills</th>
<th>COMPASS Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25th</td>
<td>50th</td>
<td>75th</td>
<td>25th</td>
<td>50th</td>
</tr>
<tr>
<td>All institutions*</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>430</td>
<td>450</td>
</tr>
<tr>
<td>2-year</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>450</td>
<td>470</td>
</tr>
<tr>
<td>4-year</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>420</td>
<td>440</td>
</tr>
<tr>
<td>Public 2-year</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>440</td>
<td>470</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>16</td>
<td>18</td>
<td>19</td>
<td>430</td>
<td>440</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>400</td>
<td>440</td>
</tr>
</tbody>
</table>

* Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

** Some institutions reported interpolated SAT mathematics scores. Where applicable, the scores were rounded to the nearest ten for presentation in this table.

*** Data for private for-profit 4-year institutions and all private 2-year institutions are included in the totals but are not shown by institution type because of small cell sizes.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.
Other Criteria Used by Postsecondary Institutions

Respondents were asked to report on other criteria that are considered in placement determinations in addition to the admissions and placement tests. Overall, 13 percent of institutions reported using some other criteria. Six percent reported using Advanced Placement or International Baccalaureate scores; 6 percent used high school grades; 3 percent considered highest English course taken; and 2 percent looked at each of the remaining criteria — high school graduation tests or end-of-course tests, faculty recommendations, and other criteria.

By institution level, 14 percent of 2-year and 12 percent of 4-year institutions reported using other criteria. Eight percent of 2-year and 4 percent of 4-year institutions reported using Advanced Placement or International Baccalaureate scores. High school grades were considered by 4 percent of 2-year institutions and 7 percent of 4-year institutions. All other criteria were used by 3 percent or less of 2-year and 4-year postsecondary education institutions. Similar patterns were found by institution type.

Table 10 Estimated percentage of institutions using criteria other than postsecondary reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentage of institutions using any criteria other than reading tests</th>
<th>Percentage of institutions using specific evaluation criteria other than reading tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High school graduation tests or end-of-course tests</td>
<td>High school grades (including grade point average)</td>
</tr>
<tr>
<td>All institutions</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>4-year</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Interpret data with caution; the coefficient of variation is greater than or equal to 30 percent but less than 50 percent.

— Reporting standards not met; too few cases in cell or the coefficient of variation is greater than or equal to 50 percent.

NOTE: Details for the number of institutions may not sum to totals because of rounding.

V. Conclusion

NAGB is conducting a comprehensive program of research to enable NAEP to report on the academic preparedness of 12th grade students for college and job training. The first phase of the research was conducted in connection with the 2009 NAEP grade 12 assessments in reading and mathematics. It included the survey of postsecondary education institutions’ use of tests and cut scores for placement. A second phase of the research is being planned in connection with the 2013 NAEP assessments.

The results of the survey of postsecondary education institutions’ use of tests and cut scores for placement will be examined in relation to the results of the other research studies, especially the statistical linking studies. Immediate attention will be paid to how the overall means and medians for each test compare with results from the linking studies. Subsequent analysis by the Governing Board will involve a finer grained examination of the cut scores and inter-quartile variation by institution type. The analysis by institution type may include subcategories within 2-year and 4-year institutions, as well. The typical 2-year versus 4-year and public versus private categories may mask important patterns within subcategories of institutions, especially the broad access institutions that enroll the majority of students. The goal will be to determine whether a “best fit” among the tests and cut scores exists for locating points on the NAEP reading and mathematics scales that represent academic preparedness for college without remediation.

Summary of Findings

Use of Tests

The majority of postsecondary education institutions use student performance on tests in determining entry-level students’ need for remedial courses in mathematics and reading. However, the frequency of use is higher for mathematics than for reading. Overall, 71 percent of postsecondary education institutions reported using some mathematics test and 53 percent some reading test in evaluating student need for remediation in those two subject domains.

Public 2-year and 4-year institutions are the most frequent users of tests for this purpose. All (100 percent) of 2-year public institutions reported using some mathematics test and 94 percent reported some reading test. Among 4-year public institutions, 85 percent reported using some mathematics test and 51 percent some reading test.

The most frequently used national standardized mathematics tests were the ACT (23 percent); COMPASS Algebra test (20 percent); SAT (17 percent); and ACCUPLACER Elementary Algebra test (16 percent). The remaining ACCUPLACER, and COMPASS mathematics tests were each used by 5 percent or less of the institutions.

Twenty-two percent of postsecondary education institutions used tests other than national standardized mathematics tests, including tests developed by the institution or state.
For reading, 22 percent of institutions reported using COMPASS Reading, 19 percent ACCUPLACER Reading Comprehension, 16 percent ACT Reading, 11 percent SAT Critical Reading, and 9 percent ASSET Reading Skills. Ten percent of postsecondary education institutions used tests other than national standardized reading tests.

Frequency of test use will be a factor the Governing Board considers in weighing the relevance, utility, and limitations of particular tests and cut scores for locating points on the NAEP scales that represent academic preparedness for college.

**Mean Cut Scores for Reading**
The overall mean cut scores reported for the most frequently used standardized national reading tests were:

- 18 on ACT Reading on a scale of 1 to 36
- 456 on SAT Critical Reading on a scale of 200 to 800
- 76 on ACCUPLACER Reading Comprehension on a scale of 20 to 120
- 41 on ASSET Reading Skills on a scale of 23 to 55
- 76 on COMPASS Reading on a scale of 1 to 99

Comparing 2-year and 4-year institutions, the mean cut scores for the ACT and SAT reading tests were higher for 2-year institutions. For the other reading tests, there were no instances of statistical significance in comparing the mean cut scores set by 2-year and by 4-year institutions.

**Use of Other Criteria**
Criteria other than college admissions and placement tests were used by 21 percent and 13 percent of institutions, respectively, for evaluating student need for remedial/developmental instruction in mathematics and reading. Other criteria include high school graduation and end-of-course tests; high school grades; highest mathematics or English course taken; Advanced Placement or International Baccalaureate scores; and faculty recommendations. In addition, respondents were given the option of describing any other criteria they use. Whether for determinations about academic preparedness in mathematics or reading, no single “other criterion” was used by more than 11 percent of institutions overall.

**Mean Cut Scores for Mathematics**
The overall mean cut scores reported for the most frequently used standardized national mathematics tests were:

- 19 on ACT mathematics on a scale of 1 to 36
- 471 on SAT mathematics on a scale of 200 to 800
- 70 on ACCUPLACER Elementary Algebra on a scale of 20 to 120
- 57 on ACCUPLACER College-Level Mathematics on a scale of 20 to 120
- 49 on COMPASS Algebra on a scale of 1 to 99
- 43 on COMPASS College Algebra on a scale of 1 to 99

Comparing 2-year and 4-year institutions, the mean cut scores for the ACT, SAT, and COMPASS Algebra tests were higher for 2-year institutions while the mean for the ACCUPLACER Elementary Algebra test was higher for 4-year institutions.
Variability of Cut Scores
In addition to the mean cut score for each test, the overall range and 25th, 50th, and 75th percentiles were reported. The size of the interquartile range on each test for all institutions was converted to standard deviation units as a way of comparing the variability in cut scores across tests.

As noted previously, an underlying assumption in this survey is that each postsecondary education institution’s conception of the reading and mathematics knowledge and skills needed to be “just academically prepared” for credit-bearing entry-level courses is operationalized in the cut scores used. The variability of those cut scores is exemplified in the size of the interquartile range, converted to standard deviation units, for each test.

The variability is appreciable in each case. It is smallest for ASSET Reading Skills, with one score point across the interquartile range and standard deviation of .15, and largest for COMPASS Algebra, with 26 score points across the interquartile range and standard deviation of 1.35.

To the extent that the underlying assumption holds, the data in this analysis support the proposition that postsecondary education institutions across the nation do not hold a single, common conception of “just academically prepared.”

The implications of this proposition are great, for individuals, families, and the nation. For example, Venezia, Kirst, and Antonio (2003) and others have documented the failure of postsecondary education institutions to convey postsecondary entry-level academic requirements clearly to students, parents, and K-12 educators. They point to the immediate victims of this absence of clarity: the students who graduate high school only to find themselves — often quite unexpectedly — placed into remedial/developmental non-credit courses, facing the added costs of money and time in pursuing a degree, and the attendant weight of discouragement. These students, disproportionately from minority backgrounds, are less likely to finish a degree. In addition, there is the associated waste of K-12 and postsecondary resources in graduating high school students who need remedial instruction and the harm to the nation in failing to maximize the human capital potential inherent in its citizens.

Questions for Further Research
This survey is part of the Governing Board’s broader program of research to transform NAEP into an indicator of academic preparedness for entry-level credit-bearing college courses and job training, without remediation. The survey results will be examined in relation to results from other components of the Board’s preparedness research.
program. Thus, this survey was designed to address a two-fold primary research question:

(1) Which national standardized tests are used by postsecondary education institutions to determine the need of entry-level students for remedial/developmental instruction in reading or mathematics?

(2) What are the cut scores on those national standardized tests below which students are deemed to need remedial/developmental instruction in reading or mathematics and at or above which students would be just academically qualified for entry-level credit-bearing college courses?

Further research could help determine the extent to which the academic knowledge and skills needed to qualify for job training in various occupations is similar to those needed to qualify for college.

What Explains the Variability of Cut Scores Within the Tests Used and the Higher Mean Cut Scores Where They Exist for 2-Year or 4-Year Institutions?

The survey documents variability in the cut scores set by postsecondary institutions from .15 standard deviations for the ASSET Reading Skills test to 1.35 standard deviations for COMPASS Algebra. In comparisons of ten reading and mathematics tests, 2-year institutions had higher mean cut scores than 4-year institutions on five tests, 4-year institutions had a higher mean cut score on one test, and there was no difference on four tests.

Of the five tests for which 2-year institutions had higher mean cut scores, four were the ACT and SAT mathematics and reading tests. There were no differences in mean cut scores on the three reading placement tests — ACCUPLACER, ASSET and COMPASS. The picture was mixed for the mathematics placement tests, with 2-year institutions higher for Compass Algebra and 4-year institutions higher for ACCUPLACER Elementary Algebra.

Further research could examine how institutions set cut scores on the tests; the predictive validity of cut scores set by institutions in terms of first-year student grades and degree completion; the relationship between the academic knowledge and skills postsecondary education institutions view as needed to qualify for entry-level credit bearing courses and the knowledge and skills represented by the cut score on the test; and the characteristics of institutions with higher and lower cut scores.

With gratitude to the many respondents that helped achieve a nationally representative sample of postsecondary institutions, there is now an answer to this question. However, there are some questions of importance that the survey did not address, and others that surfaced from the survey results that researchers and policymakers may want to pursue.

Further research could help determine the extent to which the academic knowledge and skills needed to qualify for job training in various occupations is similar to those needed to qualify for college.

What Tests and Cut Scores are used for Job Training Programs?

There is widespread consensus among national and state leaders that all high school students should graduate academically prepared for college and job training, without remediation. Implicit is the aspiration that 12th graders leave high school with the knowledge and skills needed to follow unimpeded whatever path they choose — be it college or job training. This survey was designed to gather information about the academic knowledge and skills needed to qualify for entry into college. It was not designed to answer this question for job training.
IV. References


  Retrieved from http://www.achievingthedream.org/initiatives/PRESS


  Retrieved from http://www.biz4achievement.org/


Appendix A: Technical Report

This appendix contains the technical report for the survey on “Evaluating Student Need for Developmental or Remedial Courses at Postsecondary Education Institutions.” The technical report was prepared by Westat, the contractor that conducted the survey for the National Assessment Governing Board.
National Assessment Governing Board Survey on Evaluating Student Need for Developmental or Remedial Courses at Postsecondary Education Institutions

Technical Report

Authors
Liam Ristow
Basmat Parsad

June 2012

Prepared for:
National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002

Prepared by:
Westat
1650 Research Boulevard
Rockville, Maryland 20850
(301) 251-1500
Contents

Page

Introduction ......................................................................................................................................  1
Survey Development ...................................................................................................................  2
  Technical Review Panel .........................................................................................................  2
  Pilot Test ..................................................................................................................................  2
  Feasibility Study ....................................................................................................................  3
  Content Expert Panel and Field Test .....................................................................................  3
  Cognitive Lab and Usability Study ..........................................................................................  4
Sample Design .............................................................................................................................  4
Data Collection and Response Rates .........................................................................................  5
  Nonresponse Bias Analysis ......................................................................................................  6
  Item Nonresponse ...................................................................................................................  8
Data Quality ................................................................................................................................  9
  Sampling Errors .......................................................................................................................  9
  Nonsampling Errors ................................................................................................................. 10
Definitions of Analysis Variables .............................................................................................. 10
Contact Information .................................................................................................................. 11

Appendix

A  Questionnaire for Two-Year Institutions .............................................................................. A-1
B  Questionnaire for Four-Year Institutions ............................................................................. B-1
C  Letter to the President or Chancellor .................................................................................. C-1
D  Enclosure About the National Assessment Governing Board ............................................ D-1
E  Letter to the Survey Respondent ............................................................................................ E-1
F  Nonresponse Bias Analysis .................................................................................................... F-1
G  Technical Review Panel Members ......................................................................................... G-1

Table

1  Number and percent of responding postsecondary education institutions in the study sample, and estimated number and percent of institutions the sample represents, by institution level and type: Fall 2011 .................................................................................................  6

2  Number of postsecondary education institutions with missing cut scores for mathematics and reading tests ..............................................................................................................  8
This page intentionally left blank
Introduction

The National Assessment Governing Board oversees and sets policy for the National Assessment of Educational Progress (NAEP), also known as the Nation’s Report Card. NAEP is the only continuing source of comparable national and state data available to the public on the achievement of students at grades 4, 8, and 12 in core subjects. The Governing Board is currently engaged in a research effort to support the validity of statements that might be made in NAEP reports about 12th-grade student academic preparedness in reading and mathematics for entry-level credit-bearing college courses and for job training. The survey on Evaluating Student Need for Developmental or Remedial Courses at Postsecondary Education Institutions is one component of this larger program of research. This particular study was aimed at collecting information on the tests and test scores used by postsecondary institutions to evaluate the need for remediation among degree-seeking undergraduate college students; it was not designed to collect information related to academic preparedness for job training.

The program of research consists of five types of studies:

1. Content coverage (in which the content of the NAEP 12th-grade reading and mathematics assessments is compared with the ACT, SAT, and ACCUPLACER reading and mathematics college admissions/placement tests);

2. Statistical relationship studies (in which students take both NAEP and one of the other admissions/placement tests);

3. Standard-setting studies (in which panels of experts identify the skills and knowledge in reading and mathematics on NAEP needed to qualify for entry-level credit-bearing courses without remediation for college and for selected job training programs);

4. Benchmarking studies (in which selected reference groups take NAEP); and

5. The survey of postsecondary institutions’ use of tests and cut-scores for determining student need for remediation described in this report.

Westat conducted the survey between August and December 2011 with a nationally representative sample of about 1,670 2-year and 4-year postsecondary education institutions. The questionnaire consisted of eight questions addressing the tests, cut-scores, and other criteria used in fall 2011 to evaluate entering students’ need for developmental or remedial courses in mathematics and reading. Slightly different questionnaires were used for 2-year and 4-year institutions to account for the differing academic structures of these institutions. The final questionnaire for 2-year institutions is provided in Appendix A, and the final 4-year questionnaire is in Appendix B.
Survey Development

Survey development took place between October 2009 and May 2011. Westat began survey development by revising an initial draft of the questionnaire prepared by the Governing Board. Work at this stage focused on revising the questionnaire format and question wording to reflect current best practices in survey design. Subsequent activities included reviews by expert panels, small-scale tests of draft questionnaires, and a field test with a diverse sample of 120 institutions designed to assess both questionnaire content and survey administration procedures. These activities are described below.

Technical Review Panel

A draft questionnaire was shared with an external Technical Review Panel (TRP) in November 2009 (see Appendix G). The TRP made a number of recommendations regarding the questionnaire content, sample design, and data collection methodology that were incorporated into the survey design going forward. Key recommendations included adding items to capture information about criteria other than tests used to evaluate student need for remediation (e.g., grades in high school) and suggestions for clarifying definitions of key survey terms (e.g., placement tests, remedial courses).

Pilot Test

Following the TRP meeting, a draft questionnaire was pilot tested with six institutions. The draft questionnaire addressed academic preparedness both for college and for job training. The pilot test respondents were asked to review the questionnaire and provide feedback about 1) the clarity of the project’s purpose as described on the instrument; 2) the clarity of instructions; 3) the clarity, quality, and appropriateness of the items to collect the information necessary for the study; 4) an estimation of the time necessary for completing the instrument; and 5) any suggestions for determining the entry-level programs for which respondents were asked to provide course-placement information. In addition, those participating in this review phase were asked to provide any information they had of other existing sources of the data needed for this study, including the name of the source and, if known, the frequency of collection, reporting format, and data elements recorded. Lastly, the pilot test was used to explore issues that could arise in data collection, notably the best way to identify the appropriate survey respondent. The pilot test findings pointed to problems with respondents’ interpretation of some questionnaire items and
definitions. Another major finding was that the survey did not provide adequate coverage of the various approaches used by institutions to evaluate student need for remediation. In addition, the pilot test demonstrated that inconsistency across institutions, particularly 2-year institutions, in categorizing career training programs and the academic prerequisites for certificate versus non-certificate programs would require the development of a detailed survey so burdensome as to deter respondents from completing it. Key questionnaire changes following the pilot test included the removal of a section of questions addressing evaluation of students enrolled in career training programs and the reorganization of survey items into two sections focused on mathematics and reading assessment.

**Feasibility Study**

To test the changes made following the pilot test, Westat conducted a second exploratory study of a revised questionnaire with eight additional postsecondary institutions in spring 2010. The feasibility study was designed to gain a better understanding of the problems encountered in the pilot test, with the goal of identifying those that could pose significant risk to a full-scale data collection. Among the key issues explored were the dual use of ACT and SAT tests for both admissions and placement purposes, how to define the appropriate student population for reporting on the survey, and the best way to operationalize in the questionnaire the study’s key objective of capturing cut scores that indicate student preparedness for entry-level college courses. Following the feasibility study, several changes were made to the questionnaire to address these and other areas found to be potentially problematic.

**Content Expert Panel and Field Test**

The issues uncovered in the pilot test and feasibility study prompted a recommendation for a larger scale test of the questionnaire to ensure that key problems had been resolved prior to conducting a full-scale survey administration. This work was carried out in two stages. First, feedback on the survey was sought from a panel of content experts during a half-day teleconference. Seven content experts participated to provide feedback on a revised questionnaire. Findings from the discussion confirmed the use of varied and complex approaches to evaluate student preparedness and the potential for further refinement of the questionnaire. Critically, the feedback from the panel resulted in a recommendation to use slightly different reporting instructions for 2-year and 4-year institutions, resulting in different questionnaires for the two institution types.
Second, a field test was conducted with a diverse sample of 120 postsecondary institutions in fall 2010 to explore questionnaire issues and potential hurdles to full-scale data collection. Those findings were used to inform additional changes to the survey instrument and data collection approaches. As a result of the field test, the use of separate forms for 2-year and 4-year institutions was supported; the test lists were revised; minor adjustments were made to ensure that the survey items are correctly understood; and the president’s office was confirmed as the appropriate starting point to identify the survey respondent.

**Cognitive Lab and Usability Study**

A final round of testing was conducted after the field test. Westat carried out a small “cognitive lab” and web usability study with nine institutions to confirm that changes made to the questionnaire (including a graphical instruction for reporting the correct test score) and the addition of new web-based data checks would help address the issues encountered during the field test and previous survey testing. The results of this activity indicated that the revised questionnaire and online data checks had effectively addressed many of the problems identified in previous testing. Minor changes, including a change in the reference period to fall 2011, were made to the questionnaire and web survey data checks following this round of tests.

**Sample Design**

The sample consisted of about 1,670 postsecondary education institutions. The sampling frame or respondent universe from which the sample of institutions was drawn was constructed from the 2009–10 Institutional Characteristics (IC) component of the Integrated Postsecondary Education Data System (IPEDS) maintained by the National Center for Education Statistics (NCES). To be eligible for the study, 2-year and 4-year degree-granting institutions had to be located in the 50 states or the District of Columbia and must have offered an undergraduate course of study, although the institution may have also offered graduate degrees. As with many other sample surveys of postsecondary institutions, the current study excluded non-Title IV institutions because these institutions are generally too few in the universe to be reported separately. The sampling frame for the study comprised 4,220 institutions.
The sampling frame was stratified by instructional level (four-year, two-year), control (public, private not-for-profit, private for-profit), highest level of offering (doctor’s/doctoral-professional practice, master’s, bachelor’s, less than bachelor’s), and total fall enrollment. Within each strata, institutions were sorted by region (Northeast, Southeast, Central, West) and by level of minority enrollment (high black enrollment, high total minority enrollment but not high black enrollment, and low minority enrollment). The sample of about 1,670 institutions was allocated to the strata in proportion to the aggregate square root of total enrollment. Institutions within a stratum were sampled with equal probabilities of selection.

Data Collection and Response Rates

Cover letters were mailed to the president of each sampled institution in August 2011. The letter introduced the study and requested that the president identify a respondent for the survey by completing and returning an enclosed respondent information form. Also accompanying the letter were a copy of the questionnaire and background information on the Governing Board and its program of research (see Appendix C for a copy of the president’s letter and Appendix D for the informational material about the Governing Board).

Once a respondent was identified by the president’s office, survey materials were sent directly to the respondent via email, mail, or fax, with email being the primary method. Materials sent to the survey respondent included a cover letter providing an overview of the survey and instructions for logging on to the online survey, a copy of the questionnaire, and the informational material about the Governing Board and its research program. A copy of the letter sent to respondents is provided in Appendix E. Respondents were encouraged to complete the survey online, but they could complete it via mail, email, or fax. Follow-up for survey recruitment, nonresponse, and data clarification was initiated in September 2011 and completed in January 2012. Telephone calls and email were used to follow up with respondents.

Of the about 1,670 institutions in the sample, approximately 100 were found to be ineligible for the survey because they did not meet one of the following eligibility criteria in fall 2011:

- Enrollment of entering students in a degree program designed to transfer to a 4-year institutions (applies to 2-year institutions only); or
- Enrollment of entering students in an undergraduate degree program in the liberal arts and sciences (applies to 4-year institutions only).
In addition to the approximately 100 ineligible institutions, about five institutions were found to be outside of the study’s scope because of closure, leaving about 1,560 eligible institutions in the sample. Completed questionnaires were received from about 1,340 institutions (Table 1). Of the institutions that completed the survey, 83 percent completed it via the web, 15 percent completed it by telephone, and 2 percent completed it by mail, fax, or email. The unweighted response rate for the survey was 86 percent, and the weighted response rate was 81 percent using the initial base weights. The weighted number of eligible institutions in the survey represents the estimated universe of postsecondary education institutions in the 50 states and the District of Columbia that offer a degree program at the bachelor’s or associate’s level and enroll entering students per the study’s eligibility criteria noted above.

Table 1. Number and percent of responding postsecondary education institutions in the study sample, and estimated number and percent of institutions the sample represents, by institution characteristics: Fall 2011

<table>
<thead>
<tr>
<th>Institution characteristic</th>
<th>Responding institutions (unweighted)</th>
<th>National estimate (weighted)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>All institutions</td>
<td>1,340</td>
<td>100</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>460</td>
<td>34</td>
</tr>
<tr>
<td>4-year</td>
<td>880</td>
<td>66</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>410</td>
<td>31</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>420</td>
<td>31</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>390</td>
<td>29</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>80</td>
<td>6</td>
</tr>
</tbody>
</table>

NOTE: Detail may not sum to total due to rounding.
SOURCE: The National Assessment Governing Board, survey on “Evaluating Student Need for Developmental or Remedial Courses at Postsecondary Education Institutions,” fall 2011.

Nonresponse Bias Analysis

The weighted response rate for the National Assessment Governing Board (NAGB) survey of institutions was 81 percent and the unweighted response rate was 86 percent. As specified in the NCES Statistical Standards (2002), a nonresponse bias analysis is required if the weighted unit response rate for a survey is less than 85 percent (Standard 4-4-1). Thus, a nonresponse bias analysis was conducted to (a) assess and document the impact nonresponse may have on estimates derived from the survey, and (b) assess the extent to which the non-response-adjusted sampling weights developed for analysis are effective in reducing potential nonresponse biases. A summary of the analysis is available in Appendix F.
Response rates varied considerably by sector, type of control, and enrollment size class (see Appendix F, Table 1). For example, by sector, the weighted response rates were 91 percent for four-year public institutions; 89 percent for two-year public institutions; 84 percent for four-year private, not-for-profit institutions; 74 percent for two-year private, not-for-profit institutions; 68 percent for four-year private for-profit institutions; and 56 percent for two-year, private for-profit institutions. To compensate for the differential survey response rates, weight adjustments were developed and applied to the base weights within appropriate weight adjustment classes. In general, such weight adjustments will reduce nonresponse bias if the variables used in forming the weight adjustment classes are correlated with response propensity (the probability that a sampled institution will respond to the survey) and with the characteristics obtained from the survey.

There are reasons to believe that the nonresponse-adjusted weights developed for the survey of institutions will be reasonably effective in reducing potential biases. First, the weight adjustments removed most of the disparities between the weighted distributions of the respondents and the distributions of the total sample. Second, a comparison of weighted estimates of selected data items available in the IPEDS files showed that the weight adjustment procedures were effective in reducing the difference between the weighted estimate for the respondent sample and the corresponding base-weighted estimate for the total sample. Further evidence of the potential bias reductions is given by a comparison of weighted estimates of selected survey items before and after nonresponse adjustment. For example, among the eight numeric variables examined, none of the differences between the adjusted and unadjusted estimates are statistically significant at the 0.05 level, and none are significant at the more stringent 0.01 level. This suggests that the degree of nonresponse experienced in the survey is unlikely to have an appreciable impact on estimates of mean test scores.

Based on this analysis, it appears that the estimates derived from the study using the nonresponse adjusted weights are nationally representative. Although it is possible to conduct more in-depth analysis and possibly refine the weighting procedures, the results of this analysis suggest that any potential improvements will be modest.
Item Nonresponse

Item nonresponse was reported for question 2 (mathematics tests) and question 6 (reading tests). For both questions, some respondents reported using tests to evaluate entering students’ need for developmental or remedial courses but did not provide cut scores below which such courses were needed. In the vast majority of these cases, respondents indicated they could not report a single cut score (e.g., because the cut score varied depending on some other factor). In other cases, the test score was unreported with no reason given. Table 2 displays the number of institutions unable to report a single cut score and the number with missing cut scores and with no reason provided for mathematics and reading tests on questions 2 and 6. As shown in the table, item nonresponse rates were low and do not require any adjustments.

Table 2. Number of postsecondary education institutions with missing cut scores for mathematics and reading tests

<table>
<thead>
<tr>
<th>Test name</th>
<th>Respondent unable to report a single cut score</th>
<th>Cut score missing with no reason provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 2: Mathematics tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT Mathematics</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>ACT Composite</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>SAT Mathematics</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>SAT total score including writing</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SAT total score excluding writing</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ACCUPLACER Arithmetic</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>ACCUPLACER Elementary Algebra</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>ACCUPLACER College-Level Mathematics</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>ASSET Numerical Skills</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ASSET Elementary Algebra</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>COMPASS Pre-Algebra placement domain</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>COMPASS Algebra placement domain</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>COMPASS College Algebra placement domain</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Question 6: Reading tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT Reading</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ACT Composite</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SAT Critical Reading</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>SAT total score including writing</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SAT total score excluding writing</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ACCUPLACER Reading Comprehension</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>ASSET Reading Skills</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>COMPASS Reading placement domain</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Nelson-Denny Reading</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

NOTE: The first data column displays the number of institutions that could not report a single cut score for a given test (e.g., because the score varied depending on some other factor). The second data column displays the number of institutions with a missing cut score and no reason given for the missing score.

SOURCE: The National Assessment Governing Board, survey on “Evaluating Student Need for Developmental or Remedial Courses at Postsecondary Education Institutions,” Fall 2011.
Data Quality

While the survey was designed to account for sampling error and to minimize nonsampling error, estimates produced from the data collected are subject to both types of error. Sampling error occurs because the data are collected from a sample rather than a census of the population, and nonsampling errors are errors made during the collection and processing of the data.

Sampling Errors

The responses were weighted to produce national estimates (see Table 1). The weights were designed to adjust for the variable probabilities of selection and differential nonresponse. The findings are estimates based on the sample selected and, consequently, are subject to sampling variability. General sampling theory was used to estimate the sampling variability of the estimates.

The standard error is a measure of the variability of an estimate due to sampling. It indicates the variability of a sample estimate that would be obtained from all possible samples of a given design and size. Standard errors are used as a measure of the precision expected from a particular sample. If all possible samples were surveyed under similar conditions, intervals of 1.96 standard errors below to 1.96 standard errors above a particular statistic would include the true population parameter being estimated in about 95 percent of the samples. This is a 95 percent confidence interval. For example, the estimated percentage of postsecondary institutions that used the ACT Mathematics test to evaluate entering students’ need for developmental or remedial mathematics courses is 23 percent and the standard error is 0.8 percent. The 95 percent confidence interval for the statistic extends from \(23 - (0.8 \times 1.96)\) to \(23 + (0.8 \times 1.96)\), or from 21 to 25 percent. The 1.96 is the critical value for a two-tailed Z test at the 0.05 significance level (where 0.05 indicates the 5 percent of all possible samples that would be outside the range of the confidence interval).

Because the data from the survey were collected using a complex sampling design, the variances of the estimates from this survey (e.g., estimates of proportions) are typically different from what would be expected from data collected with a simple random sample. Not taking the complex sample design into account can lead to an underestimation of the standard errors associated with such estimates. To generate accurate standard errors for the estimates in this report, standard errors were computed using a technique known as jackknife replication. As with any replication method, jackknife replication involves constructing a number of subsamples (replicates) from the full sample and computing the statistic of
interest for each replicate. The mean square error of the replicate estimates around the full sample estimate provides an estimate of the variance of the statistic. To construct the replications, 50 stratified subsamples of the full sample were created and then dropped one at a time to define 50 jackknife replicates. A computer program (WesVar) was used to calculate the estimates of standard errors.

Nonsampling Errors

Nonsampling error is the term used to describe variations in the estimates that may be caused by population coverage limitations and data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems like unit and item nonresponse, differences in respondents’ interpretations of the meaning of questions, response differences related to the particular time the survey was conducted, and mistakes made during data preparation. It is difficult to identify and estimate either the amount of nonsampling error or the bias caused by this error. To minimize the potential for nonsampling error, this study used a variety of procedures, including the various tests described in the Survey Development section above. In addition, extensive editing of the questionnaire responses was conducted to check the data for accuracy and consistency. Cases with missing or inconsistent items were recontacted by telephone to resolve problems. Data were keyed with 100 percent verification for surveys received by mail, fax, or telephone.

Definitions of Analysis Variables

- **Institution level**: 2-year and 4-year institutions. 2-year institutions are defined as institutions at which the highest level of offering is at least 2 but less than 4 years (below the baccalaureate degree); 4-year institutions are those at which the highest level of offering is 4 or more years (baccalaureate or higher degree).¹

- **Institution type**: public 2-year, private 2-year, public 4-year, private not-for-profit 4-year, private for-profit 4-year. Institution type was created from a combination of institution level (2-year, 4-year) and institution control (public, private not-for-profit, private for-profit). Private for-profit 2-year and private not-for-profit 2-year institutions are combined in the private 2-year category due to the small number of private not-for-profit 2-year institutions in the sample.

¹ Definitions for level are from the data file documentation for the IPEDS Institutional Characteristics file, U.S. Department of Education, National Center for Education Statistics.
Contact Information

For additional information about the study, contact Ray Fields, Assistant Director for Policy and Research, National Assessment Governing Board, 800 North Capitol Street, NW, Washington, DC 20002; telephone (202) 357-0395; e-mail Ray.Fields@ed.gov.
This page intentionally left blank.
Appendix A

Questionnaire for Two-Year Institutions
Please respond for the institution that matches the institution name and IPEDS ID number printed on the label below. If the information for the institution shown is incorrect, please update directly on the label before returning the survey questionnaire.

VERSION FOR TWO-YEAR INSTITUTIONS

This survey can be completed online at www.nagb-survey.org. We encourage you to complete the survey online if possible. You will need the User ID and Password shown on the label above to log in to the website. Please contact Liam Ristow at nagb-mailbox@westat.com, 1-888-429-6827 (toll-free), or 240-314-2456 if you do not have your User ID or Password.

If you prefer, you may complete this paper version. If you complete the paper version, please provide the following information, keep a copy of the completed questionnaire for your files, and return the original to Westat at the address shown below. We have enclosed a postage-paid envelope for your convenience. You may also fax a copy of the completed questionnaire to 1-800-254-0984.

Thank you. Please keep a copy of the survey for your records.

Please see page 8 for a list of Frequently Asked Questions (FAQs) regarding the purpose of this study.

Mail: NAGB Survey (8756.05.02)  
    Westat  
    1600 Research Boulevard, TA 1006F  
    Rockville, Maryland 20850-3195  
Fax: 1-800-254-0984

IF YOU HAVE ANY QUESTIONS OR COMMENTS, CONTACT:  
Liam Ristow at Westat  
1-888-429-6827 or 240-314-2456  
E-mail: nagb-mailbox@westat.com

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 3098-0006. The time required to complete this information collection is estimated to average 30 minutes per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. If you have any comments concerning the accuracy of the time estimate or suggestions for improving this form, OR if you have any comments or concerns regarding the status of your individual submission of this form, please write directly to National Assessment Governing Board, 800 North Capitol Street, NW, Washington, DC 20002.
Survey Overview and Instructions

This survey examines the test scores used by postsecondary institutions to evaluate whether entering students are academically prepared or in need of developmental or remedial courses in mathematics or reading. The goal of the survey is to identify the test scores below which students are deemed in need of developmental or remedial courses (i.e., academically unprepared for entry-level courses). The scores may come from a variety of tests, including:

- Admissions tests such as the ACT or SAT;
- Placement tests such as ACCUPLACER, ASSET, and COMPASS; or
- Institution- or state-developed tests.

Please report based on the tests your institution uses to evaluate entering students who are pursuing a degree program that is designed to transfer to a four-year institution.

Answer all relevant sections, even if your institution does not offer developmental or remedial courses or offers only one or two such courses. Use data from your institutional records whenever possible. If exact data are not available, then give your best estimate.

Definitions

**Developmental or remedial courses** are generally designed to improve the skills of entering students who are not academically prepared for entry-level college courses, as determined by your institution. This survey focuses on developmental or remedial courses that address skills in **mathematics** and **reading**.

Please note:

- In most cases, developmental or remedial courses are not credit-bearing and they do not count toward general education or degree requirements.
- Your institution may use other names such as "preparatory," "compensatory," or "basic skills" or some other term to refer to developmental or remedial courses.

**Entry-level college courses** are first-year credit-bearing courses that require college-level mathematics or reading skills, as these skills are defined by your institution. Entry-level college courses typically count toward general education or degree requirements. Please note that a variety of entry-level courses may require college-level mathematics or reading skills, including entry-level humanities, mathematics, and science courses.

**Entering students** include full-time and part-time students who are new to your institution and are subject to your institution’s policy for determining need for developmental or remedial courses.
Instructions for reporting test scores on questions 2 and 6

Questions 2 and 6 ask for the test scores below which entering students were identified as in need of remedial or developmental mathematics or reading courses in fall 2011. This is one of many student placement determinations using tests such as the ACT, SAT, or placement tests (i.e., ACCUPLACER, ASSET, COMPASS, or tests developed within an institution or state).

The example of a placement test score scale below shows different placement outcomes at different score ranges or points. In this example, the correct score point to report on questions 2 and 6 is indicated with an arrow. Please note that the scores shown do not represent those on an actual test and may not match your institution’s placement policy.

Example of a placement test score scale (0–100)

<table>
<thead>
<tr>
<th>Score</th>
<th>Placement outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>80 or above</td>
<td>Students are placed into college courses above entry-level or into academic programs with advanced skills requirements (e.g., engineering, physics, and mathematics programs)</td>
</tr>
<tr>
<td>50 to 79</td>
<td>Students are placed into entry-level college courses</td>
</tr>
<tr>
<td>50</td>
<td>Students scoring below this level are in need of remedial or developmental courses. Students scoring at or above this level are placed into entry-level college courses</td>
</tr>
<tr>
<td>40 to 49</td>
<td>Students are placed into the highest level of remedial or developmental courses</td>
</tr>
<tr>
<td>39 or below</td>
<td>Students are placed into lower levels of remedial or developmental courses</td>
</tr>
</tbody>
</table>

On questions 2 and 6, report only the score below which students needed developmental or remedial courses.
**Directions:** If your institution did not have any entering students in fall 2011 who were pursuing a degree program designed to transfer to a four-year institution, please check this box, complete the cover page, and return the questionnaire.

### Section A. Evaluating Need for Developmental or Remedial Mathematics Courses

1. In fall 2011, did your institution use ACT, SAT, or placement tests (i.e., ACCUPLACER, ASSET, COMPASS, or other tests developed by your institution or state) to evaluate whether entering students were in need of developmental or remedial mathematics courses (i.e., not academically prepared for entry-level courses that require college mathematics skills)?
   - **Consider tests used to evaluate entering students who were pursuing a degree program designed to transfer to a four-year institution.**
   - **Consider any use of ACT, SAT, or placement test scores to determine the need for remediation, even if your institution does not offer developmental or remedial mathematics courses or offers only one or two such courses.**

   Yes: 1 (Continue with question 2.)
   No: 2 (Skip to question 3.)

2. In Column B, please check the box for each ACT, SAT, or placement test that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial mathematics courses. In Column C, for each test used, write the score below which students were identified as in need of developmental or remedial mathematics courses.
   - **If different scores were used to identify students for different levels of developmental or remedial mathematics courses, report the score used for the highest level of remedial mathematics course.**
   - **If different scores were used for either requiring or recommending developmental or remedial mathematics courses, report the highest score used.**
   - **Please refer to the instructions on page 3 for an example of how to report test scores for this question.**

<table>
<thead>
<tr>
<th>A. ACT, SAT, or placement test (Score ranges shown in parentheses)</th>
<th>B. Test used to evaluate entering students</th>
<th>C. Score below which developmental or remedial mathematics courses were needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Subject Tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Mathematics (1-36) ..........................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>b. Composite score (1-36) .....................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>SAT Reasoning Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Mathematics (200-800) ....................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>d. SAT total score including Writing (600-2400) .....................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>e. SAT total score excluding Writing (400-1600) .....................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>ACCUPLACER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Arithmetic (20-120) ..........................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>g. Elementary Algebra (20-120) .............................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>h. College-Level Mathematics (20-120) ....................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>ASSET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. Numerical Skills (23-55) ...................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>j. Elementary Algebra (23-55) ...............................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>k. Intermediate Algebra (23-55) ............................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>l. College Algebra (23-55) ...................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>COMPASS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m. Pre-Algebra placement domain (1-99) ....................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>n. Algebra placement domain (1-99) ........................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>o. College Algebra placement domain (1-99) ............................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Other mathematics placement tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p. Other test 1 (specify): .....................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>q. Other test 2 (specify): .....................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>r. Other test 3 (specify): .....................................................</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

---

A-6
**COMMENT BOX FOR QUESTION 2:** Please provide additional details about your response to question 2 here. For example, if you reported more than one score for the ACCUPLACER, ASSET, or COMPASS tests, respectively, please confirm that these are scores below which students were in need of remedial or developmental mathematics.

---

3. In fall 2011, did your institution use any criteria other than ACT/SAT tests or placement tests to evaluate whether entering students were in need of developmental or remedial mathematics courses (i.e., not academically prepared for entry-level courses that require college mathematics skills)?

   - **Consider other criteria such as high school graduation or end-of-course exams, high school records, Advanced Placement (AP) or International Baccalaureate (IB) test scores, or faculty recommendations.**

   - Yes.................1  *(Continue with question 4.)*
   - No .................2  *(Skip to question 5.)*

4. Please check the box for each criterion that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial mathematics courses (i.e., not academically prepared for entry-level courses that require college mathematics skills).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Used to evaluate entering students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. High school graduation tests or end-of-course tests</td>
<td></td>
</tr>
<tr>
<td>b. High school grades (including grade point average)</td>
<td></td>
</tr>
<tr>
<td>c. Highest high school mathematics course completed</td>
<td></td>
</tr>
<tr>
<td>d. Advanced Placement (AP) or International Baccalaureate (IB) test scores</td>
<td></td>
</tr>
<tr>
<td>e. Faculty recommendation</td>
<td></td>
</tr>
<tr>
<td>f. Other (specify)</td>
<td></td>
</tr>
</tbody>
</table>

**COMMENT BOX FOR QUESTION 4:** If you have information that may explain how the criteria were used to evaluate student need for remedial or developmental mathematics, please provide it here.

---

A-7
Section B. Evaluating Need for Developmental or Remedial Reading Courses

5. In fall 2011, did your institution use ACT, SAT, or placement tests (i.e., ACCUPLACER, ASSET, COMPASS, or other tests developed by your institution or state) to evaluate whether entering students were in need of developmental or remedial reading courses (i.e., not academically prepared for entry-level courses that require college reading skills)?

- Consider tests used to evaluate entering students who were pursuing a degree program designed to transfer to a four-year institution.
- Consider any use of ACT, SAT, or placement test scores to determine the need for remediation, even if your institution does not offer developmental or remedial reading courses or offers only one or two such courses.

Yes .............................  1 (Continue with question 6.)
No ..............................  2 (Skip to question 7.)

6. In Column B, please check the box for each ACT, SAT or placement test that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial reading courses. In Column C, for each test used, write the score below which students were identified as in need of developmental or remedial reading courses.

- If different scores were used to identify students for different levels of developmental or remedial reading courses, report the score used for the highest level of remedial reading course.
- If different scores were used for either requiring or recommending developmental or remedial reading courses, report the highest score used.

Please refer to the instructions on page 3 for an example of how to report test scores for this question.

<table>
<thead>
<tr>
<th>A. ACT, SAT, or placement test (Score ranges shown in parentheses)</th>
<th>B. Test used to evaluate entering students</th>
<th>C. Score below which developmental or remedial reading courses were needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT Subject Tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Reading (1-36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Composite score (1-36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT Reasoning Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Critical Reading (200-800)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. SAT total score including Writing (600-2400)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. SAT total score excluding Writing (400-1600)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCUPLACER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Reading Comprehension (20-120)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASSET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. Reading Skills (23-55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPASS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h. Reading placement domain (1-99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson-Denny Reading Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. Nelson-Denny Reading test (0-172)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other reading placement tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>j. Other test 1 (specify):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k. Other test 2 (specify):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l. Other test 3 (specify):</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not applicable.
7. In fall 2011, did your institution use any criteria other than ACT/SAT tests or placement tests to evaluate whether entering students were in need of developmental or remedial reading courses (i.e., not academically prepared for entry-level courses that require college reading skills)?
- Consider other criteria such as high school graduation or end-of-course exams, high school records, Advanced Placement (AP) or International Baccalaureate (IB) test scores, or faculty recommendations.

Yes .................. 1 (Continue with question 8.)
No ................... 2 (Stop. Complete cover page and return questionnaire.)

8. Please check the box for each criterion that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial reading courses (i.e., not academically prepared for entry-level courses that require college reading skills).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Used to evaluate entering students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. High school graduation tests or end-of-course tests</td>
<td>☐</td>
</tr>
<tr>
<td>b. High school grades (including grade point average)</td>
<td>☐</td>
</tr>
<tr>
<td>c. Highest high school English course completed</td>
<td>☐</td>
</tr>
<tr>
<td>d. Advanced Placement (AP) or International Baccalaureate (IB) test scores</td>
<td>☐</td>
</tr>
<tr>
<td>e. Faculty recommendation</td>
<td>☐</td>
</tr>
<tr>
<td>f. Other (specify)</td>
<td>☐</td>
</tr>
</tbody>
</table>

COMMENT BOX FOR QUESTION 8: If you have information that may explain how the criteria were used to evaluate student need for remedial or developmental reading, please provide it here.

Thank you for completing the survey. Please remember to complete the information on the cover page (name and contact information) before returning the questionnaire.
FREQUENTLY ASKED QUESTIONS

Why is this survey important?
The academic preparedness of entering students for entry-level courses is a major concern for many higher education institutions. This survey will provide otherwise unavailable nationally representative data about the tests and cut-scores used by two-year and four-year institutions to determine student need for remediation in mathematics and reading. Thus, your institution’s participation will help address a knowledge gap on this vital issue.

Why was my institution selected? Do I have to do this?
Your institution was randomly chosen to participate in this study. Your institution’s participation is voluntary, and there is no penalty if you choose not to answer any or all questions in this survey. Your institution’s participation is very important for the success of the study because the answers you provide will be used to represent other institutions similar to yours.

Who is the sponsor of this survey?
The National Assessment Governing Board is the sponsor of this survey. An enclosure provides details about the Governing Board and how this study fits into its overall research program. The Governing Board was established by Congress in 1988 to oversee and set policy for the National Assessment of Educational Progress (NAEP), also known as the Nation’s Report Card. NAEP reports regularly to the public on the academic achievement of 4th, 8th and 12th grade students in reading, writing, mathematics, science, and other subjects, such U.S. history, civics, geography, economics, and the arts. For more information about NAEP, go to http://nces.ed.gov/nationsreportcard. For more information about the National Assessment Governing Board, go to http://www.nagb.gov.

Will responses from my institution be kept private?
Yes. The information provided by your institution will be kept private to the extent permitted by law. Data for this study will be reported in aggregate form; the information provided by your institution will be combined with other participating institutions to produce statistical summaries and reports. Your institution’s name or individual survey responses will not be reported.

Why is this study important? How will the information my institution provides be used?
The National Assessment Governing Board has undertaken a program of research to enable NAEP to report on the academic preparedness of 12th grade students for placement into entry-level credit-bearing college courses. This study is an important part of a program of research that involves more than 30 studies. An enclosure provides details about how this study fits into the overall research program. The data resulting from this survey will be used, along with the results of the other planned studies, to help develop valid statements that can be made about the preparedness of 12th grade students in NAEP reports. Survey results will be provided to your institution after they are finalized.

Who is conducting the survey?
Westat is conducting this survey under contract to the National Assessment Governing Board. Westat is a research company located in Rockville, Maryland.

How much time will it take to complete the survey?
The survey is designed to be completed in 30 minutes or less, including the time for reviewing instructions and completing and reviewing the collection of information.

Whom do I contact if I have questions?
For questions about the survey instructions or survey items, please contact Liam Ristow, the Westat survey manager, at 1-888-429-6827 (toll-free) or 240-314-2456, or by e-mail at nagb-mailbox@westat.com.

For questions about the National Assessment Governing Board and its research program, please contact Ray Fields, Assistant Director for Policy and Research, National Assessment Governing Board, at 202-357-0395, or by e-mail at Ray.Fields@ed.gov.
Appendix B

Questionnaire for Four-Year Institutions
This page intentionally left blank.
Please respond for the institution that matches the institution name and IPEDS ID number printed on the label below. If the information for the institution shown is incorrect, please update directly on the label before returning the survey questionnaire.

VERSION FOR FOUR-YEAR INSTITUTIONS

This survey can be completed online at www.nagb-survey.org. We encourage you to complete the survey online if possible. You will need the User ID and Password shown on the label above to log in to the website. Please contact Liam Ristow at nagb-mailbox@westat.com, 1-888-429-6827 (toll-free), or 240-314-2456 if you do not have your User ID or Password.

If you prefer, you may complete this paper version. If you complete the paper version, please provide the following information, keep a copy of the completed questionnaire for your files, and return the original to Westat at the address shown below. We have enclosed a postage-paid envelope for your convenience. You may also fax a copy of the completed questionnaire to 1-800-254-0984.

Name: ____________________________________________________________________________________________
Title/position: ______________________________________________________________________________________
Institution name: ____________________________________________________________________________________
Telephone number: ___________________________  E-mail: ________________________________

Thank you. Please keep a copy of the survey for your records.

Please see page 8 for a list of Frequently Asked Questions (FAQs) regarding the purpose of this study.

Mail: NAGB Survey (8756.05.02)
Westat
1600 Research Boulevard, TA 1006F
Rockville, Maryland 20850-3195
Fax: 1-800-254-0984

IF YOU HAVE ANY QUESTIONS OR COMMENTS, CONTACT:
Liam Ristow at Westat
1-888-429-6827 or 240-314-2456
E-mail: nagb-mailbox@westat.com

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number. The valid OMB control number for this information collection is 3098-0006. The time required to complete this information collection is estimated to average 30 minutes per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. If you have any comments concerning the accuracy of the time estimate or suggestions for improving this form, OR if you have any comments or concerns regarding the status of your individual submission of this form, please write directly to National Assessment Governing Board, 800 North Capitol Street, NW, Washington, DC 20002.
Survey Overview and Instructions

This survey examines the test scores used by postsecondary institutions to evaluate whether entering students are academically prepared or in need of developmental or remedial courses in mathematics or reading. The goal of the survey is to identify the test scores below which students are deemed in need of developmental or remedial courses (i.e., academically unprepared for entry-level courses). The scores may come from a variety of tests, including:

- Admissions tests such as the ACT or SAT;
- Placement tests such as ACCUPLACER, ASSET, and COMPASS; or
- Institution- or state-developed tests.

Please report based on the tests your institution uses to evaluate entering students who are enrolled in an undergraduate degree program in the liberal arts and sciences.

Answer all relevant sections, even if your institution does not offer developmental or remedial courses or offers only one or two such courses. Use data from your institutional records whenever possible. If exact data are not available, then give your best estimate.

Definitions

Developmental or remedial courses are generally designed to improve the skills of entering students who are not academically prepared for entry-level college courses, as determined by your institution. This survey focuses on developmental or remedial courses that address skills in mathematics and reading.

Please note:

- In most cases, developmental or remedial courses are not credit-bearing and they do not count toward general education or degree requirements.
- Your institution may use other names such as "preparatory," "compensatory," or "basic skills" or some other term to refer to developmental or remedial courses.

Entry-level college courses are first-year credit-bearing courses that require college-level mathematics or reading skills, as these skills are defined by your institution. Entry-level college courses typically count toward general education or degree requirements. Please note that a variety of entry-level courses may require college-level mathematics or reading skills, including entry-level humanities, mathematics, and science courses.

Entering students include full-time and part-time students who are new to your institution and are subject to your institution’s policy for determining need for developmental or remedial courses.
Instructions for reporting test scores on questions 2 and 6

Questions 2 and 6 ask for the test scores below which entering students were identified as in need of remedial or developmental mathematics or reading courses in fall 2011. This is one of many student placement determinations using tests such as the ACT, SAT, or placement tests (i.e., ACCUPLACER, ASSET, COMPASS, or tests developed within an institution or state).

The example of a placement test score scale below shows different placement outcomes at different score ranges or points. In this example, the correct score point to report on questions 2 and 6 is indicated with an arrow. Please note that the scores shown do not represent those on an actual test and may not match your institution’s placement policy.

Example of a placement test score scale (0–100)

<table>
<thead>
<tr>
<th>Score</th>
<th>Placement outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>80 or above</td>
<td>Students are placed into college courses above entry-level or into academic programs with advanced skills requirements (e.g., engineering, physics, and mathematics programs)</td>
</tr>
<tr>
<td>50 to 79</td>
<td>Students are placed into entry-level college courses</td>
</tr>
<tr>
<td>50</td>
<td>Students scoring below this level are in need of remedial or developmental courses. Students scoring at or above this level are placed into entry-level college courses</td>
</tr>
<tr>
<td>40 to 49</td>
<td>Students are placed into the highest level of remedial or developmental courses</td>
</tr>
<tr>
<td>39 or below</td>
<td>Students are placed into lower levels of remedial or developmental courses</td>
</tr>
</tbody>
</table>

On questions 2 and 6, report only the score below which students needed developmental or remedial courses.
Directions: If your institution did not have any entering students in fall 2011 who were enrolled in an undergraduate degree program in the liberal arts and sciences please check this box ☐, complete the cover page, and return the questionnaire.

Section A. Evaluating Need for Developmental or Remedial Mathematics Courses

1. In fall 2011, did your institution use ACT, SAT, or placement tests (i.e., ACCUPLACER, ASSET, COMPASS, or other tests developed by your institution or state) to evaluate whether entering students were in need of developmental or remedial mathematics courses (i.e., not academically prepared for entry-level courses that require college mathematics skills)?
   - Consider tests used to evaluate entering students who were enrolled in an undergraduate degree program in the liberal arts and sciences.
   - Consider any use of ACT, SAT, or placement test scores to determine the need for remediation, even if your institution does not offer developmental or remedial mathematics courses or offers only one or two such courses.

Yes .................................. 1 (Continue with question 2.)
No .................................. 2 (Skip to question 3.)

2. In Column B, please check the box for each ACT, SAT, or placement test that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial mathematics courses. In Column C, for each test used, write the score below which students were identified as in need of developmental or remedial mathematics courses.
   - If different scores were used to identify students for different levels of developmental or remedial mathematics courses, report the score used for the highest level of remedial mathematics course.
   - If different scores were used for either requiring or recommending developmental or remedial mathematics courses, report the highest score used.
   - Please refer to the instructions on page 3 for an example of how to report test scores for this question.

<table>
<thead>
<tr>
<th>A. ACT, SAT, or placement test (Score ranges shown in parentheses)</th>
<th>B. Test used to evaluate entering students</th>
<th>C. Score below which developmental or remedial mathematics courses were needed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACT Subject Tests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Mathematics (1-36) ..................................................</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>b. Composite score (1-36) .............................................</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td><strong>SAT Reasoning Test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Mathematics (200-800) ...............................................</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>d. SAT total score including Writing (600-2400) ..................</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>e. SAT total score excluding Writing (400-1600) ..................</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td><strong>ACCUPLACER</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Arithmetic (20-120) .................................................</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>g. Elementary Algebra (20-120)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>h. College-Level Mathematics (20-120)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td><strong>ASSET</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. Numerical Skills (23-55) ..........................................</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>j. Elementary Algebra (23-55)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>k. Intermediate Algebra (23-55)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>l. College Algebra (23-55)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td><strong>COMPASS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m. Pre-Algebra placement domain (1-99)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>n. Algebra placement domain (1-99)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>o. College Algebra placement domain (1-99)</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td><strong>Other mathematics placement tests</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p. Other test 1 (specify):</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>q. Other test 2 (specify):</td>
<td>☐</td>
<td></td>
</tr>
<tr>
<td>r. Other test 3 (specify):</td>
<td>☐</td>
<td></td>
</tr>
</tbody>
</table>

Not applicable.
COMMENT BOX FOR QUESTION 2: Please provide additional details about your response to question 2 here. For example, if you reported more than one score for the ACCUPLACER, ASSET, or COMPASS tests, respectively, please confirm that these are scores below which students were in need of remedial or developmental mathematics.

3. In fall 2011, did your institution use any criteria other than ACT/SAT tests or placement tests to evaluate whether entering students were in need of developmental or remedial mathematics courses (i.e., not academically prepared for entry-level courses that require college mathematics skills)?
   - Consider other criteria such as high school graduation or end-of-course exams, high school records, Advanced Placement (AP) or International Baccalaureate (IB) test scores, or faculty recommendations.

   Yes ................ 1 (Continue with question 4.)
   No ................. 2 (Skip to question 5.)

4. Please check the box for each criterion that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial mathematics courses (i.e., not academically prepared for entry-level courses that require college mathematics skills).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Used to evaluate entering students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. High school graduation tests or end-of-course tests</td>
<td>□</td>
</tr>
<tr>
<td>b. High school grades (including grade point average)</td>
<td>□</td>
</tr>
<tr>
<td>c. Highest high school mathematics course completed</td>
<td>□</td>
</tr>
<tr>
<td>d. Advanced Placement (AP) or International Baccalaureate (IB) test scores</td>
<td>□</td>
</tr>
<tr>
<td>e. Faculty recommendation</td>
<td>□</td>
</tr>
<tr>
<td>f. Other (specify)</td>
<td>□</td>
</tr>
</tbody>
</table>

COMMENT BOX FOR QUESTION 4: If you have information that may explain how the criteria were used to evaluate student need for remedial or developmental mathematics, please provide it here.
5. In fall 2011, did your institution use ACT, SAT, or placement tests (i.e., ACCUPLACER, ASSET, COMPASS, or other tests developed by your institution or state) to evaluate whether entering students were in need of developmental or remedial reading courses (i.e., not academically prepared for entry-level courses that require college reading skills)?
   - Consider tests used to evaluate entering students who were enrolled in an undergraduate degree program in the liberal arts and sciences.
   - Consider any use of ACT, SAT, or placement test scores to determine the need for remediation, even if your institution does not offer developmental or remedial reading courses or offers only one or two such courses.

   Yes .................................. 1 (Continue with question 6.)
   No ................................... 2 (Skip to question 7.)

6. In Column B, please check the box for each ACT, SAT or placement test that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial reading courses. In Column C, for each test used, write the score below which students were identified as in need of developmental or remedial reading courses.
   - If different scores were used to identify students for different levels of developmental or remedial reading courses, report the score used for the highest level of remedial reading course.
   - If different scores were used for either requiring or recommending developmental or remedial reading courses, report the highest score used.

   Please refer to the instructions on page 3 for an example of how to report test scores for this question.

<table>
<thead>
<tr>
<th>A. ACT, SAT, or placement test</th>
<th>B. Test used to evaluate entering students</th>
<th>C. Score below which developmental or remedial reading courses were needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Score ranges shown in parentheses)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT Subject Tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Reading (1-36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Composite score (1-36)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT Reasoning Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Critical Reading (200-800)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. SAT total score including Writing (600-2400)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. SAT total score excluding Writing (400-1600)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCUPLACER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. Reading Comprehension (20-120)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASSET</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g. Reading Skills (23-55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPASS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h. Reading placement domain (1-99)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson-Denny Reading Test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. Nelson-Denny Reading test (0-172)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other reading placement tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>j. Other test 1 (specify):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>k. Other test 2 (specify):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>l. Other test 3 (specify):</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Not applicable.
COMMENT BOX FOR QUESTION 6: If you have information that may explain how the ACT/SAT scores or any other test scores were used to evaluate student need for remedial or developmental reading, please provide it here.

7. In fall 2011, did your institution use any criteria other than ACT/SAT tests or placement tests to evaluate whether entering students were in need of developmental or remedial reading courses (i.e., not academically prepared for entry-level courses that require college reading skills)?
   
   • Consider other criteria such as high school graduation or end-of-course exams, high school records, Advanced Placement (AP) or International Baccalaureate (IB) test scores, or faculty recommendations.

   Yes .................. 1 (Continue with question 8.)
   No ................... 2 (Stop. Complete cover page and return questionnaire.)

8. Please check the box for each criterion that your institution used in fall 2011 to evaluate whether entering students were in need of developmental or remedial reading courses (i.e., not academically prepared for entry-level courses that require college reading skills).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Used to evaluate entering students</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. High school graduation tests or end-of-course tests</td>
<td>□</td>
</tr>
<tr>
<td>b. High school grades (including grade point average)</td>
<td>□</td>
</tr>
<tr>
<td>c. Highest high school English course completed</td>
<td>□</td>
</tr>
<tr>
<td>d. Advanced Placement (AP) or International Baccalaureate (IB) test scores</td>
<td>□</td>
</tr>
<tr>
<td>e. Faculty recommendation</td>
<td>□</td>
</tr>
<tr>
<td>f. Other (specify)</td>
<td>□</td>
</tr>
</tbody>
</table>

COMMENT BOX FOR QUESTION 8: If you have information that may explain how the criteria were used to evaluate student need for remedial or developmental reading, please provide it here.

Thank you for completing the survey. Please remember to complete the information on the cover page (name and contact information) before returning the questionnaire.
FREQUENTLY ASKED QUESTIONS

Why is this survey important?
The academic preparedness of entering students for entry-level courses is a major concern for many higher education institutions. This survey will provide otherwise unavailable nationally representative data about the tests and cut-scores used by two-year and four-year institutions to determine student need for remediation in mathematics and reading. Thus, your institution’s participation will help address a knowledge gap on this vital issue.

Why was my institution selected? Do I have to do this?
Your institution was randomly chosen to participate in this study. Your institution’s participation is voluntary, and there is no penalty if you choose not to answer any or all questions in this survey. Your institution’s participation is very important for the success of the study because the answers you provide will be used to represent other institutions similar to yours.

Who is the sponsor of this survey?
The National Assessment Governing Board is the sponsor of this survey. An enclosure provides details about the Governing Board and how this study fits into its overall research program. The Governing Board was established by Congress in 1988 to oversee and set policy for the National Assessment of Educational Progress (NAEP), also known as the Nation’s Report Card. NAEP reports regularly to the public on the academic achievement of 4th, 8th and 12th grade students in reading, writing, mathematics, science, and other subjects, such U.S. history, civics, geography, economics, and the arts. For more information about NAEP, go to http://nces.ed.gov/nationsreportcard. For more information about the National Assessment Governing Board, go to http://www.nagb.gov.

Will responses from my institution be kept private?
Yes. The information provided by your institution will be kept private to the extent permitted by law. Data for this study will be reported in aggregate form; the information provided by your institution will be combined with other participating institutions to produce statistical summaries and reports. Your institution’s name or individual survey responses will not be reported.

Why is this study important? How will the information my institution provides be used?
The National Assessment Governing Board has undertaken a program of research to enable NAEP to report on the academic preparedness of 12th grade students for placement into entry-level credit-bearing college courses. This study is an important part of a program of research that involves more than 30 studies. An enclosure provides details about how this study fits into the overall research program. The data resulting from this survey will be used, along with the results of the other planned studies, to help develop valid statements that can be made about the preparedness of 12th grade students in NAEP reports. Survey results will be provided to your institution after they are finalized.

Who is conducting the survey?
Westat is conducting this survey under contract to the National Assessment Governing Board. Westat is a research company located in Rockville, Maryland.

How much time will it take to complete the survey?
The survey is designed to be completed in 30 minutes or less, including the time for reviewing instructions and completing and reviewing the collection of information.

Whom do I contact if I have questions?
For questions about the survey instructions or survey items, please contact Liam Ristow, the Westat survey manager, at 1-888-429-6827 (toll-free) or 240-314-2456, or by e-mail at nagb-mailbox@westat.com.

For questions about the National Assessment Governing Board and its research program, please contact Ray Fields, Assistant Director for Policy and Research, National Assessment Governing Board, at 202-357-0395, or by e-mail at Ray.Fields@ed.gov.
This page intentionally left blank.
Dear PRESIDENT NAME:

INSTITUTION NAME has been selected to participate in a federally sponsored survey on the tests used by postsecondary institutions to determine whether students are academically prepared for college or in need of developmental or remedial courses.

I am writing to request your help in identifying the appropriate person at your institution to complete this survey. We believe that the results of this survey and other planned research may be of interest and use to postsecondary institutions that are addressing issues of student academic preparedness, persistence, and success. The survey results will provide otherwise unavailable nationally representative data on the tests and cut-scores used by postsecondary institutions in determining whether students are prepared for entry-level courses. An enclosure describes the study sponsor—the National Assessment Governing Board (an independent Federal board affiliated with the U.S. Department of Education)—and how this study fits into the Board’s overall research program on 12th grade academic preparedness.

The Federal Office of Management and Budget (OMB) has approved this survey (OMB No.: 3098-0006). This study has been endorsed by a number of national higher education organizations that encourage your participation (see reverse side of this letter). Westat, a research company in Rockville, Maryland, is contracted to conduct the survey.

Please use the enclosed Respondent Information Form to identify the person at your institution most knowledgeable about the evaluation of entering students to determine need for developmental or remedial courses in mathematics and reading. This may include individuals in offices of academic deans or provosts, academic departments (e.g., mathematics or English departments), or offices that handle institutional research, student assessment, student services, student counseling, new student orientation, or admissions. If more than one individual is involved, please identify one person who will have overall responsibility for completing the survey.

You may return the completed form to Westat in the enclosed postage-paid envelope or by e-mail or fax using the information provided on the form. The individual you identify will be asked to complete an eight-question survey questionnaire designed to be completed in 30 minutes or less. We are aware that you and the staff at your institution are confronted with many competing demands and survey requests and we have designed the survey, with input from an earlier small-scale study, to be as efficient as possible to complete. An informational copy of the questionnaire is enclosed. Answers to frequently asked questions (FAQs) about the study can be found on page 8 of the questionnaire.

Your institution’s participation is very important for the success of the study because the answers provided will be used to represent other institutions similar to yours. The information provided by your institution will be kept private to the extent permitted by law. Data for this study will be reported in aggregate form; the information provided by your institution will be combined with other participating institutions to produce statistical summaries and reports. Your institution’s name or individual survey responses will not be reported. Participation in this survey is voluntary. Survey results will be provided to your institution after they are finalized.

If you have any questions about the survey, please contact Liam Ristow, the Westat survey manager, at 1-888-429-6827 (toll-free) or 240-314-2456, or by e-mail at nagb-mailbox@westat.com. Thank you for your assistance.

Sincerely,

Ray Fields
Assistant Director for Policy and Research

800 NORTH CAPITOL STREET, NW, SUITE 825, WASHINGTON, DC 20002
The following organizations have endorsed this study and encourage your institution’s participation:

The American Association of Community Colleges
One Dupont Circle NW
Suite 410
Washington, DC 20036

The American Association of State Colleges and Universities
1307 New York Avenue NW
Washington, DC 20005

The American Council on Education
One Dupont Circle NW
Washington, DC 20036

The Association of Public and Land-grant Universities
1307 New York Avenue NW
Suite 400
Washington, DC 20005-4722

The National Association for College Admission Counseling
1050 N. Highland Street, Suite 400
Arlington, VA 22201

The State Higher Education Executive Officers
3035 Center Green Drive
Suite 100
Boulder, CO 80301-2205
Appendix D

Enclosure About the National Assessment Governing Board
This page intentionally left blank.
Program of Research on Academic Preparedness for Postsecondary Education and Training

The National Assessment of Educational Progress (NAEP) is also known as The Nation’s Report Card. NAEP is the only continuing source of comparable national and state data available to the public on the achievement of students at grades 4, 8, and 12 in core subjects. For more than 40 years, NAEP has reported to the public on the status and progress of student achievement in the United States.

The National Assessment Governing Board oversees and sets policy for NAEP. It is an independent, bipartisan Board, appointed by the U.S. Secretary of Education, comprising state and local policymakers, public and non-public educators, curriculum experts, measurement experts, representatives of business, and the general public. NAEP and the Governing Board are congressionally authorized under the National Assessment of Educational Progress Authorization Act (P.L. 107-279) (see www.nagb.gov).

Following the recommendations of a blue-ribbon commission in 2004, the Governing Board has embarked on a comprehensive program of research, with more than 30 studies authorized, to transform NAEP 12th grade reporting into an indicator of academic preparedness for college and job training. The commission concluded that having a measure of the “output” of K-12 education in the United States, as an indicator of the nation’s human capital potential at the transition point to adult pursuits, is essential for the economic well-being and security of the United States.

As the only source of nationally representative student achievement data at grade 12, NAEP has unique potential to serve as such an indicator.

The program of preparedness research for NAEP was developed by a panel of experts in measurement, research, industrial organizational psychology, and postsecondary policy, and adopted by the Governing Board. The survey in which you are being invited to participate is one component of this larger program of research.

The program of research consists of five types of studies: (1) content coverage (in which the content of the NAEP 12th grade reading and mathematics assessments is compared with the ACT, SAT, and ACCUPLACER reading and mathematics admissions/placement tests); (2) statistical relationship (in which students take both NAEP and one of the other admissions/placement tests); (3) standard setting (in which panels of experts identify the skills and knowledge in reading and mathematics on NAEP needed to qualify for entry-level credit-bearing courses without remediation for college and for selected job training programs); (4) benchmarking (in which selected reference groups take NAEP); and (5) the survey of postsecondary institutions’ use of tests and cut-scores for determining student need for remediation—the study that is part of this package.

As the Governing Board has been developing ways to implement the commission’s recommendations, there has been a wider recognition—among federal and state policymakers, K-12 and postsecondary educators, and the business community—of the importance of a rigorous high school program that results in meaningful high school diplomas and prepares students for college and for training for good jobs.

This study will provide valuable, otherwise unavailable information about the use of tests and test scores for placing first-year students into entry-level credit bearing courses or into remedial/developmental courses in mathematics and reading. The data resulting from this survey will be used, along with the results of the other planned studies, to help develop valid statements that can be made about the preparedness of 12th grade students in NAEP reports. The Governing Board believes that the results of the preparedness research program also will be of benefit to the K-12 and postsecondary communities, to inform their efforts in ensuring that our nation’s students are well-prepared for college and job training. More information about the research program and study results can be found under the Grade 12 and Preparedness headings at http://www.nagb.gov/publications/reports-papers.htm and http://www.nagb.gov/newsroom/press-releases.htm.

1 12th Grade Student Achievement in America: A New Vision for NAEP; www.nagb.gov/publications/12_gr_commission_rpt.pdf.
3 Note: By law, NAEP only reports group results; it does not produce individual student scores.

800 North Capitol Street, N.W. · Suite 825 · Washington, DC 20002-4233 · Telephone: (202) 357-6938 · Fax: (202) 357-6945 · www.nagb.org

D-3
This page intentionally left blank.
Appendix E

Letter to the Survey Respondent
This page intentionally left blank.
Dear RESPONDENT NAME:

I am writing to request your participation in a federally sponsored survey on the tests used by postsecondary institutions to determine whether students are academically prepared for college or in need of developmental or remedial courses. We received your name as the appropriate person to complete the survey through our contact with your institution’s president or chancellor.

We believe that the results of this survey and other planned research may be of interest and use to postsecondary institutions that are addressing issues of student academic preparedness, persistence, and success. The survey results will provide otherwise unavailable nationally representative data on the tests and cut-scores used by postsecondary institutions in determining whether students are prepared for entry-level courses. An enclosure describes the study sponsor—the National Assessment Governing Board (an independent Federal board affiliated with the U.S. Department of Education)—and how this study fits into the Board’s overall research program on 12th grade academic preparedness.

Answers to frequently asked questions (FAQs) about the study can be found on page 8 of the enclosed survey questionnaire. The survey consists of eight questions and is designed to be completed in 30 minutes or less. We are aware that you have many demands on your time, and we have designed the survey, with input from an earlier small-scale study, to be as efficient as possible for you to complete.

The Federal Office of Management and Budget (OMB) has approved this survey (OMB No.: 3098-0006). This study has been endorsed by a number of national higher education organizations that encourage your participation (see reverse side of this letter). Westat, a research company in Rockville, Maryland, is contracted to conduct the survey.

You were identified as the person most knowledgeable about your institution’s policy on evaluating entering students to determine need for developmental or remedial courses in mathematics and reading. We greatly appreciate your efforts to respond to the survey and, as needed, collaborate with other offices or personnel to gather the information requested on the survey.

Your participation is very important for the success of the study because the answers provided will be used to represent other institutions similar to yours. The information you provide will be kept private to the extent permitted by law. Data for this study will be reported in aggregate form; the information provided by your institution will be combined with other participating institutions to produce statistical summaries and reports. Your institution’s name or individual survey responses will not be reported. Participation in this survey is voluntary. Survey results will be provided to your institution after they are finalized.

We encourage you to complete the survey online at www.nagb-survey.org. To log in, use the User ID and Password shown on the accompanying Web Information Sheet.

If you prefer, the questionnaire may be completed and returned by mail using the enclosed postage-paid envelope or by toll-free fax at 1-800-254-0984. The survey should be completed only once, using either the online or paper version.

We ask that you complete the survey within three weeks, and that a copy is kept for your files. If you have any questions about the survey, please contact Liam Ristow, the Westat survey manager, at 1-888-429-6827 (toll-free) or 240-314-2456, or by e-mail at nagb-mailbox@westat.com. Thank you for your assistance.

Sincerely,

[Signature]

Ray Fields
Assistant Director for Policy and Research

800 NORTH CAPITOL STREET, NW, SUITE 825, WASHINGTON, DC 20002

Enclosures
The following organizations have endorsed this study and encourage your institution’s participation:

The American Association of Community Colleges
One Dupont Circle NW
Suite 410
Washington, DC 20036

The American Association of State Colleges and Universities
1307 New York Avenue NW
Washington, DC 20005

The American Council on Education
One Dupont Circle NW
Washington, DC 20036

The Association of Public and Land-grant Universities
1307 New York Avenue NW
Suite 400
Washington, DC 20005-4722

The National Association for College Admission Counseling
1050 N. Highland Street, Suite 400
Arlington, VA 22201

The State Higher Education Executive Officers
3035 Center Green Drive
Suite 100
Boulder, CO 80301-2205
Appendix F

Nonresponse Bias Analysis
This page intentionally left blank.
The weighted response rate for the National Assessment Governing Board (NAGB) survey of institutions was 81 percent and the unweighted response rate was 86 percent. As specified in the NCES Statistical Standards (2002), a nonresponse bias analysis is required if the weighted response rate for a survey is less than 85 percent (Standard 4-4-1). This report provides a summary of the findings of an analysis of nonresponse in the NAGB survey. The goals of the analysis are: (a) to assess and document the impact nonresponse may have on estimates derived from the survey, and (b) assess the extent to which the nonresponse-adjusted sampling weights developed for analysis are effective in reducing potential nonresponse biases.

This report is divided into eight sections. Section 1 provides an overview of the sample design and a description of the base weights used to compute the weighted response rates. In Section 2, the survey response rates by selected institution characteristics are summarized. In Section 3, the distributions of the responding institutions by selected characteristics are compared with the corresponding distributions of the nonresponding institutions. Section 4 summarizes the procedures used to adjust the sampling weights to compensate for nonresponse, and Sections 5 through 7 provide an assessment of the effectiveness of the weight adjustments in reducing potential nonresponse biases. Section 8 provides a summary and conclusions.

1. Sample Design and Construction of Base Weights

About 1,670 institutions were selected for the NAGB survey from a sampling frame derived from data files in the 2009 Integrated Postsecondary Education Data System (IPEDS). The sample was stratified by level (four-year versus two-year), type of control (public, private not-for-profit, and private for profit), highest level of offering (doctorate, masters, bachelors, other), and enrollment size class. Forty-three strata were specified for sampling purposes. The sampling rates used to select institutions varied by stratum, ranging from approximately 1 in 10 of small nonprofit institutions to 1 in 1 of large public and private institutions. Prior to sampling, institutions were sorted by region and minority status where feasible to induce additional implicit substratification. A systematic sample was then selected from the sorted file of institutions at the rate specified for the sampling stratum.

For subsequent weighting and analysis purposes, a base weight was computed for each sampled institution. The base weight, \( w_{hi}^{base} \), for institution \( i \) in sampling stratum \( h \) was computed as \( w_{hi}^{base} = 1/P_{hi} \), where \( P_{hi} \) is the corresponding probability of selecting the institution from the stratum. The base weights are theoretically unbiased in the absence of survey nonresponse. When survey nonresponse is relatively high, use of the base weights to derive estimates from the survey can result in biases. To minimize the potential for nonresponse bias, adjustments were made to the base weights to compensate for differential nonresponse losses (additional details about the weighting adjustments are provided in Section 4 below).

2. Response Rates by Selected Institution Characteristics

To examine the variation in response rates across different types of institutions, response rates were calculated for subsets of the sample based on selected characteristics of institutions. The characteristics included level, type of control, enrollment size class, highest level of offering, and geographic region. The results are summarized in Table 1. Of the 1,668 sample institutions, 105 were determined to be ineligible for the survey (e.g., closed, inactive, or did not enroll incoming students in a degree granting program) and are excluded from the calculation of the response rates. The last column of the table shows the p-value of a test of association between response status and each of the selected characteristics. A p-value of 0.05 or less indicates that there is a statistically significant association between the (weighted) response rate and the specified characteristic.
As can be seen in Table 1, sector (a cross-classification of institutions defined by level and type of control), type of control (across all levels), and enrollment size class are all strongly correlated with response status (p-value < 0.0005). By sector, (weighted) response rates are highest among the public four-year and two-year institutions and lowest among the two-year private for-profit institutions. Across all levels, public institutions generally had the highest response rates (90 percent) followed by private nonprofit institutions (83 percent) and private for-profit institutions (61 percent). By enrollment size class, response rates were generally higher among large institutions than smaller ones.

3. Comparison of Respondents and Nonrespondents by Selected Characteristics

The base-weighted distributions of the respondents and nonrespondents were compared for the same categories of institutional characteristics shown in table 1. The base-weighted distributions of responding institutions can be compared with the corresponding base-weighted distributions of the total sample to obtain a measure of the potential nonresponse bias. These comparisons, which are presented in table 2, provide an alternative way of examining the variation in response rates across selected subgroups of the sample. The p-value shown in the sixth column of the table corresponds to an overall test of the hypothesis that the base-weighted distribution of the respondent sample is the same as the distribution of the total sample for the given characteristic. A p-value of 0.05 or less indicates that the two distributions are significantly different, which implies that the distribution of respondents is significantly different from that of the nonrespondents. Shown in the fifth column of the table is the relative bias of the estimated percentage of a particular level of a characteristic if no adjustment is made to the base weights to compensate for nonresponse. (The tests associated with the p-values shown in the last column of this table are discussed later in Section 4.)

Consistent with the results of Section 2, the p-values given in column 6 of table 2 indicate that there are significant differences between the distributions of the respondents and nonrespondents by sector, type of control, and enrollment size class. These are essentially the same results presented earlier in table 1, but viewed in a different way. For example, by sector, the respondent sample has a much smaller percentage of two-year private for-profit institutions (12.7 percent) than the total sample (8.7 percent). Correspondingly, there are greater percentages of two-year (28.6 percent) and four-year public institutions (18.8 percent) in the respondent sample than in the total sample (26.1 and 16.7 percent, respectively). This disparity is also reflected in the relatively large spread of the relative biases shown in column 5 of the table. Similarly, by type of control, the percentage of public institutions in the respondent sample (47.4 percent) is higher than the percentage of public institutions in the total sample (42.8 percent), reflecting the generally higher response rates among the public institutions. By enrollment size class, the percentage of responding institutions with fewer than 1,000 students (26.1 percent) is lower than the corresponding percentage for the total sample (30.7 percent), reflecting the generally lower response rates among the small institutions.
4. Derivation of Nonresponse-Adjusted Weights

As noted in the previous section, the base-weighted distribution of the responding institutions differed significantly from the total sample for a number of characteristics. In general, weighting adjustments are used to compensate for distributional differences resulting from differential response rates. To be effective in reducing potential nonresponse biases, the nonresponse adjustment should be made within subsets of institutions (or “weighting classes”) that have similar propensities for responding to the survey. To start construct the weighting classes, the 43 strata specified for sampling was crossed by a four-level region code. Region was one of the variables used as an implicit stratifier in sampling and was expected to be correlated with response rates to a moderate degree. Where necessary, small cells were collapsed with other cells to ensure a minimum sample size per cell.

Since nonresponse could occur either before or after eligibility for the survey was ascertained, the weighting adjustment was conducted in two phases. The purpose of the first-phase adjustment was to compensate for the loss of about 140 sample institutions that did not return a questionnaire (and for which eligibility for the study is not known). Let \( w_{ki}^{\text{base}} \) denote the base weight for the \( i \)th sampled institution in weighting class \( k \) that returned a questionnaire. This group of institutions includes ineligible and nonresponding eligible institutions, as well as those that completed the NAGB survey. The first-phase adjusted weight for the \( i \)th sampled institution in weighting class \( k \) that returned a questionnaire was computed as:

\[
 w_{ki}^{R1} = \left(1/R_k\right) w_{ki}^{\text{base}},
\]

where \( R_k \) is the base-weighted percentage of institutions in weighting class \( k \) that returned a questionnaire. Within first-phase weighting class \( k \), the adjustment had the effect of distributing the weight of the cases that did not return a questionnaire to those cases that returned a questionnaire.

Let \( w_{gi}^{NR1} \) denote the first-stage adjusted weight of the \( i \)th responding institution in second-phase adjustment class \( g \) (the second-phase adjustment classes may differ from those used in the first-phase adjustment). The final weight for the \( i \)th responding institution in second-phase adjustment class \( g \) was then computed as:

\[
 w_{gi}^{\text{final}} = \left(1/S_g\right) w_{gi}^{R1},
\]

where \( S_g \) is the \( w_{gi}^{R1} \)-weighted percentage of institutions in weighting class \( k \) returning a questionnaire that were determined to be eligible for the study and completed the questionnaire. In this case, the second-phase adjustment had the effect of distributing a portion of adjusted weight of the cases that returned questionnaires to the eligible responding institutions in the weighting class. The \( w_{gi}^{\text{final}} \)'s defined above are the weights used to calculate estimates derived from the survey.

5. Comparisons Before and After Nonresponse Adjustment for Selected Distributions

The last three columns of table 2 summarize results related to distributions of the respondent sample using the nonresponse-adjusted weights described above. Column 7 shows the (nonresponse-adjusted) weighted distributions. Column 8 shows the corresponding relative bias. Column 9 shows the p-value for a test comparing the nonresponse-adjusted weighted distribution in column 7 with the corresponding base-weighted distribution in column 2. While significant differences were observed for some characteristics prior to nonresponse adjustment (see column 6), the differences for most of these characteristics have essentially disappeared after nonresponse adjustment, as can be seen by the small relative biases in column 8 and the non-significant p-values in column 9. The only exception is for
type of control, for which the difference remains significant; however, the corresponding relative biases are considerably smaller after the nonresponse adjustment, suggesting that the observed differences may not have an appreciable effect on the survey-based estimates.

6. **Comparisons of Estimates of IPEDS Data Items Before and After Nonresponse Adjustment**

Another way of gauging the effectiveness of the weighting procedures is to compare weighted estimates of institution-level data items available from the IPEDS sampling frame before and after nonresponse adjustment. Table 3 summarizes such a comparison. The statistics presented in these tables are based on items available (or derived) from the 2009 IPEDS data files. The p-value given in column 6 of the table corresponds to a test comparing the base-weighted estimate for respondents with the corresponding base-weighted estimate for the total sample (which is an unbiased estimate of the true population value in the IPEDS files). The p-value shown in the column 9 of the tables corresponds to a test comparing the nonresponse-adjusted estimate for respondents with the corresponding base-weighted estimate for the total sample. In Table 3, the six items listed under “numeric variables” are estimated means of selected counts reported in IPEDS. The five items under “attribute variables” are estimated percentages derived from categorical data reported in IPEDS.

For three of the six numeric variables presented in Table 3 (applications, admissions, and full-time enrollment), the base-weighted mean of the respondents is significantly different from the base-weighted mean of the total sample (p-value < 0.0005 in column 6 of the table). The corresponding relative biases range from 6 to 7 percent, indicating that the responding institutions tend to report higher IPEDS counts than nonresponding institutions. However, after nonresponse adjustment, it can be seen in column 8 of the table that the corresponding relative biases have been reduced considerably to around 2-3 percent. At the 0.01 significance level, none of the nonresponse-adjusted estimates shown in column 7 are significantly different from the corresponding unbiased estimate in column 2 (p-value = 0.047 or greater in column 9). This suggests that the nonresponse adjustments used to create the final weights may be effective in reducing the bias of survey estimates that are correlated with the variables listed in Table 3.

A similar comparison was made for the five attribute variables listed in Table 3. As indicated by the p-values in column 6 of the table, the unadjusted estimates for the respondent sample shown in column 3 are significantly different from the corresponding unbiased estimates in column 2 (p-value < 0.05) for three of the five items (dual credit, advance placement credit, and remedial services). However, after nonresponse adjustment, none of the weighted estimates (column 7) differ significantly from the unbiased estimate (column 9).

7. **Comparisons Before and After Nonresponse Adjustments for Selected Survey Results**

The final set of comparisons conducted in the nonresponse bias analysis involved a comparison of weighted estimates of selected survey characteristics using the base weights (unadjusted estimates) and nonresponse-adjusted weights (adjusted estimates). The results are summarized in Table 4. The p-value given in column 5 of this table corresponds to a test of the hypothesis that there is no difference between the two weighted estimates. The statistics under the heading “numeric variables” are estimates of mean test scores. The statistics under the heading “attribute variables” are estimates of the percentage of institutions using various tests.

Among the eight numeric variables examined, none of the differences between the adjusted and unadjusted estimates are statistically significant at the 0.05 level, and none are significant at the more stringent 0.01 level. This suggests that the degree of nonresponse experienced in the NAGB survey is
unlikely to have an appreciable impact on estimates of mean test scores. On the other hand, there are significant differences between the adjusted and unadjusted estimates for several of the attribute variables considered. Where there is a significant difference between the two estimates, the relative bias (column 4) provides a measure of the amount of bias that is potentially corrected for when using the adjusted estimate.

8. Summary and Conclusions

The weighted response rate for the NAGB survey of institutions was 81 percent. Response rates varied considerably by sector, type of control, and enrollment size class. To compensate for the differential survey response rates, weight adjustments were developed and applied to the base weights within appropriate weight adjustment classes (Section 4). In general, such weight adjustments will reduce nonresponse bias if the variables used in forming the weight adjustment classes are correlated with response propensity (the probability that a sampled institution will respond to the survey) and with the characteristics obtained from the survey.

There are reasons to believe that the nonresponse-adjusted weights developed for the survey of institutions will be reasonably effective in reducing potential biases. First, the weight adjustments removed most of the disparities between the weighted distributions of the respondents and the distributions of the total sample (Section 5). Second, a comparison of weighted estimates of selected data items available in the IPEDS files showed that the weight adjustment procedures was effective in reducing the difference between the weighted estimate for the respondent sample and the corresponding base-weighted estimate for the total sample (Section 6). Further evidence of the potential bias reductions is given by a comparison of weighted estimates of selected survey items before and after nonresponse adjustment (Section 7).

Based on this analysis, it appears that the estimates derived from the study using the nonresponse adjusted weights are nationally representative. Although it is possible to conduct more in-depth analysis and possibly refine the weighting procedures, the results of this analysis suggest that any potential improvements will be modest.
Table F-1. Sample sizes by response status, response rates, and test of association between response status and selected characteristics of sampled institutions

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample sizes by response status</th>
<th>Unweighted response rate (%)</th>
<th>Weighted response rate (%)</th>
<th>Test of association (p-value)&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Response</td>
<td>Nonresponse</td>
<td>Ineligible</td>
</tr>
<tr>
<td>All institutions</td>
<td>1,668</td>
<td>1,338</td>
<td>225</td>
<td>105</td>
</tr>
<tr>
<td>Sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four-year Public</td>
<td>469</td>
<td>415</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>Four-year Private, non-profit</td>
<td>468</td>
<td>385</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>Four-year Private, for profit</td>
<td>159</td>
<td>80</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>Two-year Public</td>
<td>468</td>
<td>410</td>
<td>52</td>
<td>1</td>
</tr>
<tr>
<td>Two-year Private, non-profit</td>
<td>22</td>
<td>13</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Two-year Private, for profit</td>
<td>82</td>
<td>35</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>Public</td>
<td>937</td>
<td>825</td>
<td>91</td>
<td>21</td>
</tr>
<tr>
<td>Private, non profit</td>
<td>490</td>
<td>398</td>
<td>69</td>
<td>23</td>
</tr>
<tr>
<td>Private, for profit</td>
<td>241</td>
<td>115</td>
<td>65</td>
<td>61</td>
</tr>
<tr>
<td>Four year</td>
<td>1,096</td>
<td>880</td>
<td>140</td>
<td>76</td>
</tr>
<tr>
<td>Two year</td>
<td>572</td>
<td>458</td>
<td>85</td>
<td>29</td>
</tr>
<tr>
<td>1,000 to 2,999</td>
<td>332</td>
<td>257</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>3,000 to 9,999</td>
<td>496</td>
<td>429</td>
<td>56</td>
<td>11</td>
</tr>
<tr>
<td>10,000 or more</td>
<td>579</td>
<td>511</td>
<td>64</td>
<td>4</td>
</tr>
<tr>
<td>Doctorate</td>
<td>470</td>
<td>397</td>
<td>54</td>
<td>19</td>
</tr>
<tr>
<td>Masters</td>
<td>369</td>
<td>314</td>
<td>44</td>
<td>11</td>
</tr>
<tr>
<td>Bachelors</td>
<td>257</td>
<td>169</td>
<td>42</td>
<td>46</td>
</tr>
<tr>
<td>Northeast</td>
<td>368</td>
<td>301</td>
<td>52</td>
<td>15</td>
</tr>
<tr>
<td>Southeast</td>
<td>408</td>
<td>313</td>
<td>68</td>
<td>27</td>
</tr>
<tr>
<td>Central</td>
<td>413</td>
<td>345</td>
<td>42</td>
<td>26</td>
</tr>
<tr>
<td>West</td>
<td>479</td>
<td>379</td>
<td>63</td>
<td>37</td>
</tr>
</tbody>
</table>

# Rounds to zero.
1 Weighted response rates are calculated using base weights.
2 Test of association between response status and institution characteristic.

NOTE: Detail may not sum to totals because of rounding. Institution characteristics are based on data available on the frame at the time of sampling and may differ from classification variables used in other reports.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th></th>
<th></th>
<th></th>
<th>Relative bias (percent)</th>
<th>Test of association (p-value)</th>
<th></th>
<th></th>
<th></th>
<th>Test of association (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Respondents</td>
<td>Non-respondents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All institutions</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four-year Public</td>
<td>16.7</td>
<td>18.8</td>
<td>7.7</td>
<td>12.5</td>
<td>17.1</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four-year Private, non-profit</td>
<td>33.2</td>
<td>34.3</td>
<td>28.5</td>
<td>3.2</td>
<td>33.7</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four-year Private, for profit</td>
<td>9.6</td>
<td>8.0</td>
<td>16.5</td>
<td>-16.7</td>
<td>9.0</td>
<td>-6.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-year Public</td>
<td>26.1</td>
<td>28.6</td>
<td>15.1</td>
<td>9.7</td>
<td>26.6</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-year Private, non-profit</td>
<td>1.7</td>
<td>1.6</td>
<td>2.4</td>
<td>-8.4</td>
<td>2.0</td>
<td>17.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-year Private, for profit</td>
<td>12.7</td>
<td>8.7</td>
<td>29.8</td>
<td>-31.1</td>
<td>11.5</td>
<td>-9.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>42.8</td>
<td>47.4</td>
<td>22.8</td>
<td>10.8</td>
<td>43.7</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private, non profit</td>
<td>34.9</td>
<td>35.8</td>
<td>30.9</td>
<td>2.7</td>
<td>35.8</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private, for profit</td>
<td>22.3</td>
<td>16.7</td>
<td>46.3</td>
<td>-24.9</td>
<td>20.6</td>
<td>-7.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four year</td>
<td>59.5</td>
<td>61.1</td>
<td>52.7</td>
<td>2.6</td>
<td>59.8</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two year</td>
<td>40.5</td>
<td>38.9</td>
<td>47.3</td>
<td>-3.9</td>
<td>40.2</td>
<td>-0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under 1,000</td>
<td>30.7</td>
<td>26.1</td>
<td>50.7</td>
<td>-15.0</td>
<td>29.2</td>
<td>-4.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000 to 2,999</td>
<td>27.7</td>
<td>28.4</td>
<td>24.5</td>
<td>2.7</td>
<td>28.5</td>
<td>2.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3,000 to 9,999</td>
<td>26.2</td>
<td>28.7</td>
<td>15.7</td>
<td>9.3</td>
<td>26.6</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10,000 or more</td>
<td>15.4</td>
<td>16.8</td>
<td>9.1</td>
<td>9.4</td>
<td>15.8</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doctorate</td>
<td>17.8</td>
<td>18.6</td>
<td>14.7</td>
<td>4.0</td>
<td>17.7</td>
<td>-1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masters</td>
<td>23.3</td>
<td>24.6</td>
<td>17.5</td>
<td>5.7</td>
<td>23.9</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelors</td>
<td>18.4</td>
<td>17.9</td>
<td>20.5</td>
<td>-2.7</td>
<td>18.2</td>
<td>-0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NA</td>
<td>40.5</td>
<td>38.9</td>
<td>47.3</td>
<td>-3.9</td>
<td>40.2</td>
<td>-0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>22.4</td>
<td>22.6</td>
<td>21.5</td>
<td>0.9</td>
<td>22.7</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>26.4</td>
<td>24.9</td>
<td>33.0</td>
<td>-5.8</td>
<td>25.2</td>
<td>-4.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>25.2</td>
<td>27.0</td>
<td>17.5</td>
<td>7.0</td>
<td>26.2</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>26.0</td>
<td>25.5</td>
<td>28.0</td>
<td>-1.8</td>
<td>25.9</td>
<td>-0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Relative bias defined to be 100*(B-A)/A, where A = base-weighted estimate for total sample and B = base-weighted estimate for respondent sample.
2 Test comparing distribution of total sample versus respondent sample using base weights.
3 Relative bias defined to be 100*(C-A)/A, where A = base-weighted estimate for total sample and C = nonresponse-adjusted estimate for respondent sample.
4 Test comparing distribution of respondent sample using nonresponse-adjusted weights with distribution of total sample using base weights.

NOTE: Detail may not sum to totals because of rounding. Institution characteristics are based on data available in either the sampling frame or IPEDS files at the time of sampling and may differ from classification variables used elsewhere in this report.
Table F-3. Comparison of weighted estimates of selected IPEDS statistics for sampled institutions, by response status

<table>
<thead>
<tr>
<th>IPEDS data item</th>
<th>Base-weighted data</th>
<th>Nonresponse-adjusted data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Respondents</td>
</tr>
<tr>
<td>Numeric variables(^5)</td>
<td></td>
<td>(Mean)</td>
</tr>
<tr>
<td>Total applications</td>
<td>2,316</td>
<td>2,484</td>
</tr>
<tr>
<td>Total admissions</td>
<td>4,067</td>
<td>4,325</td>
</tr>
<tr>
<td>Total part-time enrollment</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>Total full-time enrollment</td>
<td>768</td>
<td>822</td>
</tr>
<tr>
<td>Number of first-time degree/certificate seeking students submitting SAT scores</td>
<td>568</td>
<td>577</td>
</tr>
<tr>
<td>Number of first-time degree/certificate seeking students submitting ACT scores</td>
<td>479</td>
<td>485</td>
</tr>
<tr>
<td>Attribute variables(^3)</td>
<td></td>
<td>(Percent)</td>
</tr>
<tr>
<td>Institutions offering dual credit</td>
<td>77</td>
<td>80</td>
</tr>
<tr>
<td>Institutions with advanced placement credits</td>
<td>82</td>
<td>86</td>
</tr>
<tr>
<td>Institutions with remedial services</td>
<td>75</td>
<td>77</td>
</tr>
<tr>
<td>Institutions with academic/career counseling</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Institutions with employment services</td>
<td>88</td>
<td>89</td>
</tr>
</tbody>
</table>

\(^1\) Relative bias defined to be 100*(B-A)/A, where A = base-weighted estimate for total sample and B = base-weighted estimate for respondent sample.
\(^2\) Test comparing base-weighted estimate of total sample with base-weighted estimate of respondent sample.
\(^3\) Relative bias defined to be 100*(C-A)/A, where A = base-weighted estimate for total sample and C = nonresponse-adjusted estimate for respondent sample.
\(^4\) Test comparing nonresponse-adjusted estimate of respondent sample with base-weighted estimate of total sample.
\(^5\) Excludes missing values in IPEDS institutional characteristics(IC) file.
Table F-4. Comparison of selected weighted survey estimates for responding institutions before and after nonresponse adjustment

<table>
<thead>
<tr>
<th>Survey variable</th>
<th>Survey respondents</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Base-weighted</td>
<td>Nonresponse-</td>
<td>Relative bias</td>
<td>T-test</td>
<td></td>
</tr>
<tr>
<td></td>
<td>estimates¹</td>
<td>adjusted estimates¹</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Mean)</td>
<td>(Percent)</td>
<td>(P-value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numeric variables (mean score needing remediation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT mathematics</td>
<td>19</td>
<td>19</td>
<td>0.2</td>
<td>0.091</td>
<td></td>
</tr>
<tr>
<td>SAT mathematics</td>
<td>466</td>
<td>465</td>
<td>0.1</td>
<td>0.405</td>
<td></td>
</tr>
<tr>
<td>ACCUPLACER elementary algebra</td>
<td>68</td>
<td>68</td>
<td>0.6</td>
<td>0.221</td>
<td></td>
</tr>
<tr>
<td>COMPASS algebra</td>
<td>48</td>
<td>48</td>
<td>0.0</td>
<td>0.972</td>
<td></td>
</tr>
<tr>
<td>ACT reading</td>
<td>18</td>
<td>18</td>
<td>-0.1</td>
<td>0.556</td>
<td></td>
</tr>
<tr>
<td>Used SAT critical reading</td>
<td>450</td>
<td>450</td>
<td>0.0</td>
<td>0.989</td>
<td></td>
</tr>
<tr>
<td>Used ACCUPLACER reading comprehension</td>
<td>73</td>
<td>72</td>
<td>0.7</td>
<td>0.151</td>
<td></td>
</tr>
<tr>
<td>Used COMPASS reading</td>
<td>76</td>
<td>76</td>
<td>0.1</td>
<td>0.417</td>
<td></td>
</tr>
<tr>
<td>Attribute variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used any mathematics test to evaluate students</td>
<td>73</td>
<td>71</td>
<td>3.2</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Used ACT mathematics</td>
<td>34</td>
<td>33</td>
<td>1.9</td>
<td>0.013</td>
<td></td>
</tr>
<tr>
<td>Used SAT mathematics</td>
<td>24</td>
<td>24</td>
<td>0.6</td>
<td>0.538</td>
<td></td>
</tr>
<tr>
<td>Used ACCUPLACER elementary algebra</td>
<td>23</td>
<td>23</td>
<td>-0.2</td>
<td>0.741</td>
<td></td>
</tr>
<tr>
<td>Used COMPASS algebra</td>
<td>29</td>
<td>28</td>
<td>3.5</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Used any reading test to evaluate students</td>
<td>54</td>
<td>53</td>
<td>2.9</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Used ACT reading</td>
<td>31</td>
<td>31</td>
<td>2.1</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>Used SAT critical reading</td>
<td>22</td>
<td>21</td>
<td>1.1</td>
<td>0.452</td>
<td></td>
</tr>
<tr>
<td>Used ACCUPLACER reading comprehension</td>
<td>36</td>
<td>36</td>
<td>-0.3</td>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>Used COMPASS reading</td>
<td>44</td>
<td>43</td>
<td>3.4</td>
<td>0.003</td>
<td></td>
</tr>
</tbody>
</table>

¹For numeric variables, estimates are means. For attributes, estimates are percentages of institutions. Responses exclude missing values.
²Relative bias defined to be \(100(B - A)/A\), where \(B\) = base-weighted estimate for respondents and \(A\) = nonresponse-adjusted estimates for respondents.
³Test of difference between base-weighted and nonresponse-adjusted estimates using a variant of a nonparametric mean test.
This page intentionally left blank.
Appendix G

Technical Review Panel Members
This page intentionally left blank.
Technical Review Panel Members

Thomas Bailey
Director, Community College Research Center
Teachers College, Columbia University
New York, NY 10027

Wayne Camara
Vice President, Research and Development
College Board
New York, NY 10023

Tammy Halligan
Director, Regulatory Affairs
Government Relations
Career College Association
Washington, DC 20036

Richard Brown
Associate Professor, Education
University of Southern California
Los Angeles, CA 90089

James Chromy
Senior Fellow, Statistics
StatEpi
RTI International
Research Triangle Park, NC 27709

Bruce Vandal
Director, Postsecondary Education and Workforce Development
Education Commission of the States
Denver, CO 80203
Appendix B:
Standard Error Tables

This appendix contains companion tables of standard errors for the tables of estimates displayed in the report.
### Table B.1 Standard errors for the estimated number of postsecondary institutions in the population and the percentage of institutions using selected mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Estimated number of institutions in the population</th>
<th>Percentage of institutions using any mathematics test</th>
<th>Percentage of institutions using specific mathematics tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACT</td>
</tr>
<tr>
<td>All institutions</td>
<td>58.5</td>
<td>1.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>41.0</td>
<td>3.2</td>
<td>1.5</td>
</tr>
<tr>
<td>4-year</td>
<td>41.7</td>
<td>1.9</td>
<td>1.0</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>16.4</td>
<td>0.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>35.8</td>
<td>7.8</td>
<td>—</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>6.4</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>23.5</td>
<td>2.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>30.4</td>
<td>4.9</td>
<td>2.8</td>
</tr>
</tbody>
</table>

— Not applicable: estimate not reported.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.

### Table B.2 Standard errors for the mean mathematics test scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Mean mathematics test scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
</tr>
<tr>
<td>All institutions</td>
<td>0.1</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>0.3</td>
</tr>
<tr>
<td>4-year</td>
<td>0.1</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>0.2</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>0.1</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>0.2</td>
</tr>
</tbody>
</table>

— Not applicable: estimate not reported.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.
### Table B.3
Standard errors for the percentiles for mathematics test cut scores below which entering students were identified as in need of developmental or remedial courses in mathematics, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentiles for mathematics test cut scores</th>
<th>ACT</th>
<th>SAT</th>
<th>ACCUPLACER</th>
<th>COMPASS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mathematics</td>
<td>Mathematics</td>
<td>Elementary Algebra</td>
<td>College-Level Mathematics</td>
<td>Algebra</td>
</tr>
<tr>
<td></td>
<td>25th</td>
<td>50th</td>
<td>75th</td>
<td>25th</td>
<td>50th</td>
</tr>
<tr>
<td>All institutions</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>3.9</td>
<td>5.8</td>
</tr>
<tr>
<td>2-year</td>
<td>0.4</td>
<td>0.4</td>
<td>0.3</td>
<td>4.4</td>
<td>9.4</td>
</tr>
<tr>
<td>4-year</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>4.7</td>
<td>5.4</td>
</tr>
</tbody>
</table>

**Institution type**

| Public 2-year | 0.2 | 0.3 | 0.3 | 4.4 | 9.1 | 7.0 | 2.1 | 2.3 | 2.1 | 0.8 | 1.4 | 3.1 | 1.2 | 2.3 | 0.6 | 1.7 | 0.8 | 6.1 |
| Public 4-year | 0.2 | 0.1 | 0.2 | 3.1 | 0.7 | 3.7 | 1.6 | 0.7 | 1.1 | 3.4 | 3.0 | 2.4 | 0.6 | 0.7 | 1.4 | —   | —   | —   |
| Private not-for-profit 4-year | 0.2 | 0.2 | 0.3 | 9.5 | 9.4 | 3.3 | —  | —  | —  | —  | —  | —  | —  | —  | —  | —  | —  | —  |

— Not applicable: estimate not reported.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.

### Table B.4
Standard errors for the estimated percentage of institutions using criteria other than postsecondary mathematics tests to evaluate entering students for developmental or remedial courses in mathematics, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentage of institutions using any criteria other than mathematics tests</th>
<th>Percentage of institutions using specific evaluation criteria other than mathematics tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High school graduation tests or end-of-course tests</td>
<td>High school grades (including grade point average)</td>
</tr>
<tr>
<td>All institutions</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>2-year</td>
<td>1.8</td>
<td>0.5</td>
</tr>
<tr>
<td>4-year</td>
<td>1.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Public 2-year</td>
<td>2.3</td>
<td>0.8</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>1.7</td>
<td>0.4</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>2.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

— Not applicable: estimate not reported.

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.
Table B.5  Standard errors for the estimated percentage of institutions using selected reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentage of institutions using any reading test</th>
<th>Percentage of institutions using specific mathematics tests</th>
<th>Other reading tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ACT</td>
<td>SAT</td>
</tr>
<tr>
<td>All institutions</td>
<td>1.4</td>
<td>1.0</td>
<td>0.8</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>2.4</td>
<td>1.9</td>
<td>1.2</td>
</tr>
<tr>
<td>4-year</td>
<td>1.7</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>1.0</td>
<td>2.7</td>
<td>1.4</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>5.9</td>
<td>2.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>1.8</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>2.4</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>6.5</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

— Not applicable: estimate not reported.


Table B.6  Standard errors for the mean reading test scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Mean reading test scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACT</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
</tr>
<tr>
<td>All institutions</td>
<td>0.1</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>0.2</td>
</tr>
<tr>
<td>4-year</td>
<td>0.2</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>0.2</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>0.2</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>0.2</td>
</tr>
</tbody>
</table>

— Not applicable: estimate not reported.

Table B.7 Standard errors for the percentiles for reading test cut scores below which entering students were identified as in need of developmental or remedial courses in reading, for selected tests reported by postsecondary institutions, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentiles for reading test cut scores</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>All institutions</td>
<td>ACT</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>5.7</td>
<td>6.6</td>
<td>3.0</td>
<td>1.6</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>In 2-year</td>
<td>SAT</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>8.0</td>
<td>7.4</td>
<td>5.3</td>
<td>1.9</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Public 2-year</td>
<td>ACCUPLACER</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>7.9</td>
<td>2.4</td>
<td>4.1</td>
<td>1.9</td>
<td>0.7</td>
<td>0.8</td>
<td>--</td>
<td>0.8</td>
<td>--</td>
</tr>
<tr>
<td>4-year</td>
<td>ASSET</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>4.7</td>
<td>2.1</td>
<td>3.2</td>
<td>2.0</td>
<td>0.1</td>
<td>0.5</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>COMPASS</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>10.7</td>
<td>10.6</td>
<td>7.6</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

— Not available.
— Not applicable: estimate not reported.


Table B.8 Standard errors for the estimated percentage of institutions using criteria other than postsecondary reading tests to evaluate entering students for developmental or remedial courses in reading, by institution level and type: Fall 2011

<table>
<thead>
<tr>
<th>Institution level and type</th>
<th>Percentage of institutions using any criteria other than reading tests</th>
<th>Percentage of institutions using specific evaluation criteria other than reading tests</th>
<th>High school graduation tests or end-of-course tests</th>
<th>High school grades (including grade point average)</th>
<th>Highest school English course completed</th>
<th>Advanced Placement or International Baccalaureate scores</th>
<th>Faculty recommendation</th>
<th>Other criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>All institutions</td>
<td>0.9</td>
<td>0.3</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Institution level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-year</td>
<td>1.7</td>
<td>0.6</td>
<td>0.7</td>
<td>0.7</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>4-year</td>
<td>0.9</td>
<td>0.2</td>
<td>1.0</td>
<td>0.6</td>
<td>0.6</td>
<td>0.3</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public 2-year</td>
<td>2.0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.9</td>
<td>1.3</td>
<td>1.1</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>Private 2-year</td>
<td>1.8</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Public 4-year</td>
<td>1.2</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Private not-for-profit 4-year</td>
<td>1.6</td>
<td>0.3</td>
<td>1.7</td>
<td>1.1</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>Private for-profit 4-year</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

— Not applicable: estimate not reported.

Appendix C:
Additional Tables

This appendix contains tables for mathematics and reading, respectively, displaying estimates of the frequency of use of all tests for which data were collected. These tables are followed by the companion standard error tables for these estimates.
Table C.1 Percentage of postsecondary institutions using various mathematics tests to evaluate entering students for developmental or remedial courses in mathematics: Fall 2011

<table>
<thead>
<tr>
<th>Mathematics Test</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any mathematics test</td>
<td>71</td>
</tr>
<tr>
<td>ACT</td>
<td>Mathematics</td>
</tr>
<tr>
<td></td>
<td>Composite score</td>
</tr>
<tr>
<td>SAT</td>
<td>Mathematics</td>
</tr>
<tr>
<td></td>
<td>Total score including Writing</td>
</tr>
<tr>
<td></td>
<td>Total score excluding Writing</td>
</tr>
<tr>
<td>ACCUPLACER</td>
<td>Arithmetic</td>
</tr>
<tr>
<td></td>
<td>Elementary Algebra</td>
</tr>
<tr>
<td></td>
<td>College-Level Mathematics</td>
</tr>
<tr>
<td>ASSET</td>
<td>Numerical Skills</td>
</tr>
<tr>
<td></td>
<td>Elementary Algebra</td>
</tr>
<tr>
<td></td>
<td>Intermediate Algebra</td>
</tr>
<tr>
<td></td>
<td>College Algebra</td>
</tr>
<tr>
<td>COMPASS</td>
<td>Pre-Algebra</td>
</tr>
<tr>
<td></td>
<td>Algebra</td>
</tr>
<tr>
<td></td>
<td>College Algebra</td>
</tr>
<tr>
<td>Other mathematics tests</td>
<td>22</td>
</tr>
</tbody>
</table>


Table C.2 Percentage of postsecondary institutions using various reading tests to evaluate entering students for developmental or remedial courses in reading: Fall 2011

<table>
<thead>
<tr>
<th>Reading Test</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any reading test</td>
<td>53</td>
</tr>
<tr>
<td>ACT</td>
<td>Reading</td>
</tr>
<tr>
<td></td>
<td>Composite score</td>
</tr>
<tr>
<td>SAT</td>
<td>Critical Reading</td>
</tr>
<tr>
<td></td>
<td>Total score including Writing</td>
</tr>
<tr>
<td></td>
<td>Total score excluding Writing</td>
</tr>
<tr>
<td>ACCUPLACER</td>
<td>Reading Comprehension</td>
</tr>
<tr>
<td>ASSET</td>
<td>Reading Skills</td>
</tr>
<tr>
<td>COMPASS</td>
<td>Reading</td>
</tr>
<tr>
<td>Nelson-Denny</td>
<td>Reading</td>
</tr>
<tr>
<td>Other reading tests</td>
<td>10</td>
</tr>
</tbody>
</table>

### Table C.3
Standard errors for the percentage of postsecondary institutions using various mathematics tests to evaluate entering students for developmental or remedial courses in mathematics: Fall 2011

<table>
<thead>
<tr>
<th>Mathematics Test</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any mathematics test</td>
<td>1.6</td>
</tr>
<tr>
<td>ACT Mathematics</td>
<td>0.8</td>
</tr>
<tr>
<td>ACT Composite score</td>
<td>0.6</td>
</tr>
<tr>
<td>SAT Mathematics</td>
<td>0.8</td>
</tr>
<tr>
<td>SAT Total score including Writing</td>
<td>0.2</td>
</tr>
<tr>
<td>SAT Total score excluding Writing</td>
<td>0.4</td>
</tr>
<tr>
<td>ACCUPLACER Arithmetic</td>
<td>0.7</td>
</tr>
<tr>
<td>ACCUPLACER Elementary Algebra</td>
<td>0.8</td>
</tr>
<tr>
<td>ACCUPLACER College-Level Mathematics</td>
<td>0.4</td>
</tr>
<tr>
<td>ASSET Numerical Skills</td>
<td>0.3</td>
</tr>
<tr>
<td>ASSET Elementary Algebra</td>
<td>0.6</td>
</tr>
<tr>
<td>ASSET Intermediate Algebra</td>
<td>0.5</td>
</tr>
<tr>
<td>ASSET College Algebra</td>
<td>0.4</td>
</tr>
<tr>
<td>COMPASS Pre-Algebra</td>
<td>0.8</td>
</tr>
<tr>
<td>COMPASS Algebra</td>
<td>0.9</td>
</tr>
<tr>
<td>COMPASS College Algebra</td>
<td>0.5</td>
</tr>
<tr>
<td>Other mathematics tests</td>
<td>1.2</td>
</tr>
</tbody>
</table>

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.

### Table C.4
Standard errors for the percentage of postsecondary institutions using various reading tests to evaluate entering students for developmental or remedial courses in reading: Fall 2011

<table>
<thead>
<tr>
<th>Reading Test</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any reading test</td>
<td>1.4</td>
</tr>
<tr>
<td>ACT Reading</td>
<td>1.0</td>
</tr>
<tr>
<td>ACT Composite score</td>
<td>0.5</td>
</tr>
<tr>
<td>SAT Critical Reading</td>
<td>0.8</td>
</tr>
<tr>
<td>SAT Total score including Writing</td>
<td>0.2</td>
</tr>
<tr>
<td>SAT Total score excluding Writing</td>
<td>0.4</td>
</tr>
<tr>
<td>ACCUPLACER Reading Comprehension</td>
<td>0.9</td>
</tr>
<tr>
<td>ASSET Reading Skills</td>
<td>0.8</td>
</tr>
<tr>
<td>COMPASS Reading</td>
<td>1.1</td>
</tr>
<tr>
<td>Nelson-Denny Reading</td>
<td>0.3</td>
</tr>
<tr>
<td>Other reading tests</td>
<td>0.9</td>
</tr>
</tbody>
</table>

**SOURCE:** National Assessment Governing Board. (Fall 2011). Evaluating Student Need for Developmental or Remedial Coursework at Postsecondary Education Institutions [Survey]. Washington, DC: Author.
The survey on which this report is based — "Evaluating Student Need for Remedial or Developmental Courses at Postsecondary Education Institutions" — was conducted for the National Assessment Governing Board by Westat under contract ED-NAG-09-C-0002.
Job Training Program Curriculum Study for NAEP Preparedness Research

Submitted by WestEd
and Educational
Policy Improvement
Center (EPIC)
November 8, 2012

The National Assessment Governing Board (Governing Board) adopted a Program of Preparedness Research in March 2009 that included judgmental standard-setting (JSS) studies for the 12th grade National Assessment of Educational Progress (NAEP). These studies produced preparedness reference points on the NAEP scale for entry into job-training programs and for placement in college credit-bearing courses, representing the academic knowledge and skills required for postsecondary course and training program placement. A total of 180 job training programs were represented in the judgmental standard setting studies focusing on five occupations:

<table>
<thead>
<tr>
<th>Occupation</th>
<th>Number of Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automotive master technician</td>
<td>41</td>
</tr>
<tr>
<td>Computer support specialist</td>
<td>31</td>
</tr>
<tr>
<td>Heating, ventilation, and air conditioning technician</td>
<td>31</td>
</tr>
<tr>
<td>Licensed practical nurse</td>
<td>40</td>
</tr>
<tr>
<td>Pharmacy technician</td>
<td>37</td>
</tr>
</tbody>
</table>

The Governing Board requested additional research to examine the validity of findings obtained from the JSS studies and to better understand the knowledge, skills, and abilities in reading and mathematics required for these occupational training programs. This additional research is intended to provide a clearer understanding of the knowledge, skills, and abilities (KSAs) required for entry- and exit-level coursework in designated job training programs within these occupations. This study will help to determine if the KSAs required of students in the training programs are appropriately represented by the borderline preparedness descriptions (BPDs) and by the NAEP items near the reference points developed in the JSS studies to represent the minimal level of academic knowledge and skills in the subject matter necessary for a student to be prepared to enter the job training course.

**METHODOLOGY**

This study addresses the following research questions:

1. What mathematics and reading KSAs are prerequisite to the introductory-level courses, and what mathematics and reading KSAs are taught in the introductory courses for the job-training programs for each occupation?
2. What mathematics and reading KSAs are students expected to have attained at the conclusion of the job-training programs for each occupation?
3. How do the prerequisites (KSA expectations for entry) for job training programs in each occupation relate to descriptions of minimal academic preparedness on NAEP (as described by the BPDs from the JSS studies)?

4. How do prerequisites (KSA expectations for entry) for job training programs in each occupation relate to the content assessed by NAEP (as determined by NAEP items representing minimal academic preparedness)?

This study comprises three primary phases:

1. Identification and collection of course artifacts
2. Review of course artifacts by Review Teams
3. Analysis and reporting

**Identification and Collection of Course Artifacts**

Programs from the five occupations used in the JSS studies have comprised the population of programs for this study; from this population, a minimum of 20 programs per occupation have been recruited from the 180 programs represented on the JSS panels.

Occupational job-training instructors who served on the JSS panels were recruited to participate in this study. These job training instructors were asked to identify courses that best address the objectives of this study and to submit artifacts for those courses. These instructors also had the option of nominating colleagues who teach one or more courses selected for the study to participate in this activity. Course artifacts were collected for all programs in each occupational area that agreed to participate, with course submission remaining open until either materials were obtained from a minimum of 20 programs or the population of programs had been exhausted.

Each participating program instructor was asked to (1) identify foundational textbooks for her/his program; (2) verify program and institution information (e.g., accreditation status, course sequencing, school and department admission requirements, degree accreditation, and credit requirements); and (3) submit course artifacts for one introductory course. Course artifacts may have been submitted via a web-based upload tool, email, facsimile, or physical mail.

Preliminary analysis of curriculum sequence and course lists, provided evidence of skill building and job-specific math and reading courses within pharmacy technician job training programs. As a result, submitters from Pharmacy Technician programs were also asked to submit artifacts for one concluding course. The concluding course response rates for pharmacy technician were sufficient to allow a review of these artifacts.

**Introductory courses**

Introductory courses differed across programs within an occupation, and across occupations, in terms of standardization and sequencing. As such, “entry-level” courses could embody one or more of numerous definitions, including (1) those that occurred lowest in the course sequence for a program, regardless of course title; (2)
those that were core “Introduction to...” or “Foundations of...” courses that occurred across the majority of programs, and (3) those that were identified by instructors as being most representative of the mathematics and reading expectations for entry-level students in the program.

Because the study focuses on identifying the mathematics and reading skills expected upon entry into introductory-level courses in the job-training programs for each occupation, courses were selected for inclusion using the third definition.

Concluding courses
Concluding, or exit-level, courses also differed in level of standardization, and multiple options for identifying such courses also exist. For consistency, the same approach was used to identify the exit-level courses for inclusion in the study: instructors were asked to identify those courses that best represent the mathematics and reading knowledge and skills that students are expected to know upon program completion.

For each training program, a set of course materials was collected for introductory courses and a set for concluding courses. The following types of artifacts were submitted and assembled into a course packet (with only one of each type of artifact required):

1. Course syllabus
2. Textbook title(s) (with author and ISBN)
3. Textbook table of contents (instructor copied and uploaded or EPIC downloaded from publisher website)
4. Course exam (one or more), preferably the mid-term or earlier for introductory courses and the final exam for concluding courses
5. Text-based assignment (one or more), with corresponding passage, that best illustrates mathematics and reading KSAs needed by students—one or more for introductory courses and one or more for concluding courses
6. Stand-alone assignment (one or more) such as a lab, worksheet, problem sheet, essay, or group project that best represents mathematics and reading KSAs needed for students—one or more for introductory courses and one or more for concluding courses

Instructors representing institutions that offered more than one program within an occupational area were asked to complete a submission for one program and to complete submissions for additional degree programs if selected courses were different than those already submitted.
Review of Course Artifacts
Upon completion of gathering course artifacts, teams of content experts were trained to consistently and reliably apply a coding scheme to the course artifacts to identify prerequisite and taught content for each of the occupational training programs. Two Review Teams were recruited, one for mathematics and one for reading. Each team consisted of two mathematics or two reading experts and one occupational area expert (e.g., automotive master technician) who reviewed all packets within an occupation area. Licensed Practical Nursing and Pharmacy Technician had two occupational area experts representing proprietary and public institutions. The content experts in mathematics and reading were recruited from a pool of trained experts who have substantial experience in this type of work. Due to project timeline limitations, the content experts reviewed two occupational areas (e.g., HVAC and CSS, Pharm Tech and LPN). The occupational-area experts recruited for each of the Review Teams (one for mathematics and one for reading) were drawn from pools of mathematics and reading JSS panelists who were nominated by the JSS studies’ content and process facilitators as being well qualified for this type of work.

Review team members independently coded the course packets for their content area. In order to maximize the efficiency of the Review Teams, an initial set of foundational KSAs were used to analyze course materials. These foundational KSAs included the NAEP frameworks and additional KSAs derived from the National Career Clusters™ Essential Knowledge and Skill Statements, synthesized to reduce redundancy and to present only those KSAs relevant to mathematics and reading.

Once the Review Teams’ review of course materials was complete, EPIC staff aggregated the individual ratings for each course within each program to summarize the mathematics and reading KSAs prerequisite to and taught in introductory-level courses and that students are expected to have attained at program completion. Aggregated responses were displayed in overall content maps describing the relationship between frameworks and prerequisite KSAs for each occupation. In addition to tabular data displays, the data was displayed using color shading, as well as summary statistics, to show the extent of overlap in content between standards and programs. Content maps, grouped by key characteristics, were also created for programs to show the impact of key program characteristics that impacted findings. EPIC staff reviewed the content maps to identify similarities and differences across program types within occupations and noted the differences in findings due to program characteristics. Final results were provided both overall and by key program characteristics. EPIC staff also computed descriptive statistics to summarize the Review Teams’ demand ratings overall (by occupation) and by program type, in case program characteristics had an impact on the demand of occupational courses.

Review of Knowledge, Skills, and Abilities Required for Training Courses
Two NAEP Expert Teams, one team for mathematics and one for reading, each consisting of three experts, reviewed the prerequisite and taught KSAs (as identified by the Review Teams) in the context of NAEP. They were charged with describing the
relationships between the prerequisite content and both the BPDs and the content on the 2009 NAEP, evaluating the results of the Review Team analyses to describe KSAs assessed by NAEP that are not included in the job-training programs and KSAs included in Automotive Master Technician that are not part of the NAEP frameworks or assessments.

Comparison to BPDs
Using the Review Teams’ determination of KSA requirements and course artifacts, the NAEP Expert Teams were tasked with synthesizing and describing the relationship between the content that is prerequisite to and taught in occupational programs and the content described in the BPDs for that program. Conclusions were provided overall for Automotive Master Technician, identifying differences related to program characteristics.

Comparison to NAEP items
Each NAEP Expert Team was also tasked with comparing KSAs identified for each program’s introductory courses (drawing upon the content maps and BPD comparisons) to the NAEP item pools. Starting with a set of items near the cut scores identified in the JSS studies, they judged the correspondence between the course prerequisite KSAs and the KSAs needed to correctly respond to items with a .67 probability. They were asked to identify the items in the range of the cut score plus one standard deviation that are prerequisite to or required in the courses. They were also asked to examine items below the cut score and above the range in the first analysis to determine if the KSAs represented in the curricular requirements were largely above or below this range.

Due to qualities inherent with mathematics and reading content, the NAEP expert reviews yielded dissimilar results. For example, the mathematics team was able to complete comparisons of both the BPDs and NAEP items, while reading provided feedback specific to NAEP items and frameworks.

Consensus Meeting Process
During the project pilot, consensus discussions and decisions occurred via webinars. During these webinars, three content review members (i.e., two content experts and one occupational expert) for each content area (i.e., reading and math) participated in consensus discussions for each KSA-related point of disagreement. Time allocated for (i.e., 1-1.5 hours for math and 2-2.5 hours for reading)—and scheduling of (i.e., approximately 6 webinars over 8 weeks)—these webinars was constrained by content reviewer team members’ outside time commitments to regular work schedules and other professional and life demands. As a result, EPIC staff recommended that for the operational study, consultants convene for an onsite consensus meeting. Meeting in a central location allows for more focused group processing and eliminates the time lag between webinars, maximizing decision reliability and validity.

In order to accommodate the schedules of all participants, two meetings were scheduled over two weekends in October 2012. Meeting 1 consisted of six content review teams and Meeting 2 consisted of two content review teams. Each of these
teams was comprised of both content experts and occupational area experts, together with EPIC staff. The goal was to discuss all course packet review ratings with the purpose of identifying KSAs present in the course materials. Each team gathered over three sessions, totaling 1.5 work days (i.e., each individual session totaled .5 work days).

PILOT STUDY
In order to address unanticipated challenges that arose when implementing the proposed design, materials, and/or logistics, a pilot—or feasibility—study was implemented. The automotive master technician occupation had been selected for the pilot study. Lessons learned through the pilot study were used to refine the study design as needed for the subsequent four operational occupations.

Lessons Learned
Suggestions identified during the pilot study were discussed with the Governing Board, and decisions were documented as Lessons Learned and implemented for data collection and analyses within the remaining occupations for the operational study.

Lesson Learned: 1
Program context is essential to understanding results and interpreting similarities and differences across different types of programs. Program-level contextual data should be presented within the context of the research question to inform decisions related to sample size, structure and to better understand the context of content review team results.

Recommendations
We suggest that further, in depth review of program level data is needed to describe the population, the sample, and the representativeness of the programs included in the study. These data should be synthesized and consulted at each stage of analysis to provide context for interpreting findings, as well as to identify possible implications for sampling programs, or interpreting and reporting results. Such analysis will allow us to make informed decisions related to the total number and representativeness of programs sampled for each occupational area.

Lesson Learned: 2
The different types of artifacts collected may contribute differently to meeting study objectives.

Recommendations
Of the course artifacts collected so far, we suggest further, in depth review of content review team usefulness (i.e., helpfulness) ratings, and of the stand-alone assignment included in course packets within each occupation in contributing to answering the research questions.

We will use this analysis to confirm the most useful artifacts to include in course packets for the content review teams. This analysis also contributes to our understanding of differential weighting of artifacts.
Lesson Learned: 3
Additional training is necessary for the content review team members to maximize reliability.

Recommendations
To enhance agreement and monitor consistency across content review team members, we recommend:

1. Increased training for course instructors to ensure shared task understanding and increased reliability.
2. Requiring content reviewers to successfully complete a sample course artifact review to demonstrate their understanding before conducting further operational reviews.

Such training would include a high level walk-through of materials, thorough and easily applied definitions of the coding schemes, and successful completion of a qualifying course artifact review. Training and qualifying review completion would occur via Webinar or email prior to any onsite meetings. All content review team members successfully completed training.

Lesson Learned: 4
Focus on entry-level courses to more thoroughly address the primary research objective of identifying the prerequisite KSAs for each occupation.

Recommendation(s)
We recommend analyzing entry-level courses only to best focus the study on addressing the primary research objectives.

Lesson Learned: 5
NAEP experts are more qualified than content review team members to analyze BPD alignment to prerequisite KSA for each program and should be included in the review, synthesis, and incorporation (when necessary) of content review team member feedback on the relationship between the BPD and the prerequisite KSAs for each program.

Recommendations:
Because of their ability to synthesize content, we’d like to provide the NAEP experts with:

• The extent of overlap/consistency among content review team ratings.
• Comments from content review team about what’s missing from BPD, in lieu of content review team evaluating the BPD.
• Comments from the content review team about relative program rigor, based on prerequisite KSAs.
• Content review team lists of any KSA required by a program, not already included in the KSAs.
• Opportunities to evaluate course artifacts relative to each other, not relative to the BPD.
NAEP experts were asked to then apply their expertise in NAEP and in their content domain to interpret and analyze these ratings in order to identify the NAEP KSAs.

**Lesson Learned: 6**
The text-based artifact was the least provided artifact type and the most potentially useful. Make it clearer that we are not looking only for a textbook passage, but any reading material assigned at the beginning of the course.

**Recommendation:**
Update operational process and design document to make it clear that any reading passage is acceptable for a packet to be considered complete.

**Lesson Learned: 7**
We found that the data collected in the pilot was not presented in sufficient depth or detail for the NAEP experts to analyze.

**Recommendation:**
Clarify the process, data collected, and data to be used for both sets of review teams.

**Lesson Learned: 8**
To increase understanding and usefulness, clarity, timeliness, and comprehensiveness in project documentation and reporting need to be improved.

**Recommendation:**
Use the pilot report to document the final study approach and as the beginning of the final study report rather than as a stand-alone document.
KEY ACTIVITIES IN PROPOSED PROJECT SCHEDULE

<table>
<thead>
<tr>
<th>PLANNING ACTIVITIES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Submission Tool released to participants</td>
<td>1/12/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PILOT STUDY ACTIVITIES (INTRODUCTORY COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Course artifact collection(^1)</td>
<td>1/12/12–2/14/12</td>
</tr>
<tr>
<td>Review Teams course packet reviews</td>
<td>2/3/12–2/24/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>3/1/12–3/12/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REMAINING OCCUPATIONS ACTIVITIES (INTRODUCTORY COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Course artifact collection(^1)</td>
<td>2/13/12–3/9/12</td>
</tr>
<tr>
<td>Review Teams course packet reviews</td>
<td>3/2/12–4/20/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>3/12/12–5/7/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PILOT STUDY ACTIVITIES (CONCLUDING COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Teams course packet reviews</td>
<td>4/24/12–5/4/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>5/4/12–5/16/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REMAINING OCCUPATIONS ACTIVITIES (CONCLUDING COURSES)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Teams course packet reviews</td>
<td>5/8/12–6/29/12</td>
</tr>
<tr>
<td>NAEP Expert Teams reviews</td>
<td>5/18/12–7/12/12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REPORTING</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COSDAM update report submitted to Governing Board</td>
<td>2/9/12</td>
</tr>
<tr>
<td>Draft pilot report submitted to Governing Board</td>
<td>3/31/12</td>
</tr>
<tr>
<td>COSDAM update report submitted to Governing Board</td>
<td>4/19/12</td>
</tr>
<tr>
<td>Final pilot report submitted to Governing Board</td>
<td>4/30/12</td>
</tr>
<tr>
<td>COSDAM update report submitted to Governing Board</td>
<td>7/5/12</td>
</tr>
<tr>
<td>Draft final report submitted to Governing Board</td>
<td>12/28/12</td>
</tr>
<tr>
<td>Final report submitted to Governing Board</td>
<td>1/31/13</td>
</tr>
</tbody>
</table>

PROGRESS UPDATE

COURSE PACKET NUMBERS
The total number of Introductory Course Packets reviewed was as follows*:

- **Mathematics (Introductory):**
  - Computer Support Specialist—10
  - HVAC—18
  - Pharmacy Technician—22
  - Licensed Practical Nurse—14

- **Reading (Introductory):**
  - Computer Support Specialist—11
  - HVAC—14
  - Pharmacy Technician—22
  - Licensed Practical Nurse—15
Concluding Course Packets reviewed*:

- Mathematics (Concluding):
  - Pharmacy Technician—17

- Reading (Concluding):
  - Pharmacy Technician—19

Content review team members convened during an onsite meeting in Portland, Oregon, during the first two weekends of October to determine final introductory ratings using a convergent consensus model as follows:

1. October 4-7 HVAC (reading and math), Computer Support Specialist (reading and math), Pharmacy Technician (math) Licensed Practical Nurse (math)
2. October 11-14 Pharmacy technician (introductory reading), Licensed Practical Nurse (reading)

Reviewers conducted convergent consensus meetings for Pharmacy Technician concluding reading and math packets using a shared spreadsheet online and via webinars on November 3rd and November 6th, 2012.

During the meetings, final ratings were determined for discrepant independent ratings for applicability and importance.

**Analysis and Reporting**
All project activities are complete, and analysis and reporting are currently underway.
Introduction

In 2011, the College Board developed the SAT College and Career Readiness Benchmark to assist educators and policymakers in their efforts to better evaluate the college readiness of their students. This benchmark was designed to identify the point on the SAT scale that is indicative of students' having a high likelihood of success in college which was defined as a 65 percent probability of obtaining a first year GPA (FYGPA) of 2.67 (B-) or higher. This criteria was informed by a panel of expert educators and policy makers convened by the College Board in 2007 (Kobrin, Patterson, Wiley, and Mattern, 2012).

There are several advantages in using FYGPA as an outcome variable to measure college readiness. FYGPA encompasses all of the courses a student completes during his or her first term and often represents approximately 25 percent of the courses a student will complete during college. In addition, courses taken during a students' first year are typically more uniform than those taken during subsequent years, making it a more appropriate measure of general preparedness (Wiley, Wyatt and Camara, 2010). Research has established a strong correlation between FYGPA and retention, and the likelihood of continuing college for four years increases substantially for students with higher FYGPAs (Allen, 1999; Murtaugh, Burns, & Schuster, 1999).

One limitation of the college readiness benchmark is that it is not linked to content specific performance and may include a different range of subjects for different students. However, establishing a content related SAT section score benchmark based on a single specific course (e.g. SAT Math to College Algebra) would exclude a large percent of freshmen students who did not take that particular course (Shaw and Patterson, 2010). One way to address this limitation is to develop a benchmark which corresponds to performance in several freshmen courses within a content area(s). Accordingly, the College Board has calculated benchmarks that link SAT section scores to performance in multiple related freshmen college courses. Critical Reading section scores (SAT-CR) were linked to performance in courses which require extensive reading assignments, Math section scores (SAT-M) were linked to performance in math courses, and Writing section scores (SAT-W) were linked to performance in courses which typically require writing.

Data and Methodology

The data were obtained from a sample of 199,366 SAT takers who self reported their HSGPA, graduated high school in 2009 and attended one of the 131 four-year colleges and universities that participated in the College Board’s validity study. These institutions provided information on first year course titles, grades and credit hours earned. Titles were used to identify courses utilizing reading and writing skills as well as courses in math and related subjects (see Table 1). The decision to link SAT-CR to all courses likely to have extensive reading requirements rather than solely English courses was driven both by content considerations and empirical evidence. From a content perspective, both the Common Core College Readiness standards and the National Assessment of Educational Progress (NAEP) framework emphasize the use of informational text¹ that prepares students for reading material in a variety of areas including social science, history, science, and technical areas (Camara and Quanemoen, 2012; Common Core State Standards, n.d). From an empirical perspective, performance on AP Exams in English, history, and social science is moderately to strongly correlated with both PSAT CR and PSAT W section scores (Ewing, Camara, and Millsap, 2006). This suggests that reading and writing skills are related to success in entry level college courses in these subjects.

Thus, SAT section scores were linked to college level performance in coursework in the subject areas as described in Table 1. SAT section scores were linked to content specific FYGPA’s rather than to individual course grades to provide a more complete picture of student performance than would individual course grades alone.

¹ The NAEP framework recommends that 12th graders reading material be comprised of 70% informational text.
Table 1: Course Content areas that Comprise Content Area FYGPA’s

<table>
<thead>
<tr>
<th>Content Area FYGPA: Included Coursework</th>
<th>Reading</th>
<th>Writing</th>
<th>Math</th>
<th>Math &amp; Science</th>
<th>STEM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Business &amp; Communications</td>
<td>- Business &amp; Communications</td>
<td>- Math</td>
<td>- Math</td>
<td>- Math</td>
</tr>
<tr>
<td></td>
<td>- English (excluding writing)</td>
<td>- English</td>
<td>- Science</td>
<td>- Science</td>
<td>- Science</td>
</tr>
<tr>
<td></td>
<td>- History</td>
<td>- History</td>
<td>- Computer Science</td>
<td>- Computer Science</td>
<td>- Engineering</td>
</tr>
<tr>
<td></td>
<td>- Humanities</td>
<td>- Humanities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Social Science</td>
<td>- Social Science</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data Analyses

Logistic regression was used to compute the SAT content benchmarks. Logistic regression is a statistical method that uses one or more predictor variables (in this case, an SAT section score) to predict a binary outcome (e.g. achieving a content FYGPA of 2.67 or higher). A series of separate logistic regression equations were estimated for each of the five content areas (see Table 1) using the SAT section score as the predictor variable and the dichotomized content area FYGPA (e.g. 0 if below 2.67 and 1 if 2.67 or higher) as the outcome variable.

SAT section benchmarks were established for each of the 131 institutions participating in the College Board validity study. Any out of range institution-level benchmark (e.g. lower than 200 or higher than 800) was excluded\(^2\) and the remaining institution level benchmarks were averaged, weighted by the institution-level sample sizes. A total of eighteen benchmarks were computed for each content area using six probability levels (from 50% to 75%) and three content area FYGPA’s (i.e. 2.00, 2.67, and 3.00).

Results

Table 2 includes results for each of the logistic regressions. This table contains the maximum number of institutions (K) and the sample size (N) for each subject area analysis, the K and N used to create each of the eighteen benchmarks\(^3\), and the SAT content benchmark score. The content benchmark scores associated with a 65% probability of obtaining a 2.67 or higher are highlighted. These parameters were recommended by the College Board’s panel of educators and policy makers, although it should be noted that these recommendations were made based on overall FYGPA and not content specific FYGPA’s\(^4\). The content benchmark scores associated with a 65% probability of obtaining a content FYGPA of 2.67 or higher were 500 on SAT CR (to “reading” course FYGPA’s\(^5\)); 470 on SAT W (to “writing” course FYGPA’s); and, 610 - 630 on SAT M, depending on the college course composition. The benchmark SAT M score associated with a FYGPA comprised strictly of math courses was 630; the SAT M benchmark score associated with a combined math/science FYGPA is 620; and the SAT M benchmark score associated with a STEM FYGPA is 610.

The College Board now has a suite of college readiness benchmarks that provides information for a variety of purposes and objectives. The SAT College and Career Readiness benchmark of 1550 provides an indicator of overall student readiness while the SAT content benchmarks provide a measure of student readiness within specific content areas.

---

\(^2\) The number of institutions that were excluded because of out-of-range SAT values can be calculated from Table 2 by subtracting “K” from 131.

\(^3\) Schools whose benchmark score falls below 200 or above 800 were dropped and the number of valid institutions can differ between benchmarks. More schools had “out of range” benchmark scores when 2.00 was used as the content GPA outcome (than 2.67 or 3.00) as there was less variability associated with achieving this outcome.


\(^5\) See Table 1 for the course categorizations
## Table 2: SAT Scores Associated with a Given Probability of Obtaining First Year Course Content GPA

<table>
<thead>
<tr>
<th>Course</th>
<th>Probability</th>
<th>2.00</th>
<th>2.67</th>
<th>3.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>K</td>
<td>N</td>
<td>SAT</td>
</tr>
<tr>
<td><strong>SAT CR</strong></td>
<td>50%</td>
<td>55</td>
<td>78,156</td>
<td>260</td>
</tr>
<tr>
<td>To</td>
<td>55%</td>
<td>65</td>
<td>95,963</td>
<td>280</td>
</tr>
<tr>
<td><strong>“Reading” Courses</strong></td>
<td>60%</td>
<td>75</td>
<td>106,630</td>
<td>300</td>
</tr>
<tr>
<td>(K = 131)</td>
<td><strong>65%</strong></td>
<td>83</td>
<td>125,284</td>
<td>310</td>
</tr>
<tr>
<td>(N = 186,282)</td>
<td>70%</td>
<td>94</td>
<td>143,432</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>98</td>
<td>147,238</td>
<td>370</td>
</tr>
<tr>
<td><strong>SAT W</strong></td>
<td>50%</td>
<td>56</td>
<td>79,531</td>
<td>250</td>
</tr>
<tr>
<td>To</td>
<td>55%</td>
<td>67</td>
<td>95,671</td>
<td>270</td>
</tr>
<tr>
<td><strong>“Writing” Courses</strong></td>
<td>60%</td>
<td>72</td>
<td>106,630</td>
<td>290</td>
</tr>
<tr>
<td>(K = 131)</td>
<td><strong>65%</strong></td>
<td>83</td>
<td>122,802</td>
<td>300</td>
</tr>
<tr>
<td>(N = 193,974)</td>
<td>70%</td>
<td>93</td>
<td>141,746</td>
<td>320</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>104</td>
<td>156,964</td>
<td>350</td>
</tr>
<tr>
<td><strong>SAT Math</strong></td>
<td>50%</td>
<td>108</td>
<td>120,463</td>
<td>350</td>
</tr>
<tr>
<td>To</td>
<td>55%</td>
<td>115</td>
<td>129,834</td>
<td>370</td>
</tr>
<tr>
<td><strong>All Math Courses</strong></td>
<td>60%</td>
<td>121</td>
<td>134,553</td>
<td>400</td>
</tr>
<tr>
<td>(K = 131)</td>
<td><strong>65%</strong></td>
<td>124</td>
<td>141,195</td>
<td>430</td>
</tr>
<tr>
<td>(N = 143,665)</td>
<td>70%</td>
<td>126</td>
<td>141,728</td>
<td>470</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>129</td>
<td>142,509</td>
<td>510</td>
</tr>
<tr>
<td><strong>SAT Math</strong></td>
<td>50%</td>
<td>114</td>
<td>161,190</td>
<td>350</td>
</tr>
<tr>
<td>To</td>
<td>55%</td>
<td>117</td>
<td>163,215</td>
<td>380</td>
</tr>
<tr>
<td><strong>Math/ Science Courses</strong></td>
<td>60%</td>
<td>120</td>
<td>163,996</td>
<td>410</td>
</tr>
<tr>
<td>(K = 131)</td>
<td><strong>65%</strong></td>
<td>128</td>
<td>173,875</td>
<td>430</td>
</tr>
<tr>
<td>(N = 175,654)</td>
<td>70%</td>
<td>131</td>
<td>175,654</td>
<td>460</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>130</td>
<td>174,387</td>
<td>500</td>
</tr>
<tr>
<td><strong>SAT Math</strong></td>
<td>50%</td>
<td>115</td>
<td>165,538</td>
<td>350</td>
</tr>
<tr>
<td>To</td>
<td>55%</td>
<td>115</td>
<td>165,538</td>
<td>380</td>
</tr>
<tr>
<td><strong>STEM Courses</strong></td>
<td>60%</td>
<td>122</td>
<td>168,330</td>
<td>400</td>
</tr>
<tr>
<td>(K = 131)</td>
<td><strong>65%</strong></td>
<td>125</td>
<td>170,317</td>
<td>430</td>
</tr>
<tr>
<td>(N = 178,755)</td>
<td>70%</td>
<td>129</td>
<td>177,337</td>
<td>460</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td>129</td>
<td>177,100</td>
<td>500</td>
</tr>
</tbody>
</table>
References


Attachment A-2
Reporting Plans for Research Findings

The Board’s First Phase of Preparedness Research is essentially complete. To begin the reporting process, the Board discussed specific staff-developed reporting options at the May 2012 Board meeting. At the August 2012 meeting, the Board decided to release an online technical report that would describe:

- the research conducted,
- the main research findings, and
- plans for future research based on the 2013 NAEP.

The online technical report will not include statements about specific NAEP scores or score ranges as representing academic preparedness. At the August Board meeting, in both committee discussions and a plenary session, the Board addressed several core issues:

- Audience
- Content
- Presentation and Tone
- Future Research

Board discussions have highlighted the need to:

- Clearly communicate the nature of the research studies and the key findings.
- Avoid jargon.
- Caution users against misuse and misinterpretation.

Hence, the Board prefers to release the completed research study reports as a package on the Internet, accompanied by brief summaries of their methodology and key findings. Additionally, the Board would like this package (an online technical report) to include a statement on the status of the Board’s preparedness research and future plans.

Board staff have assembled draft summaries for the Board’s review (sent under separate cover) in order to develop this online technical report. The report is embargoed (pending Governing Board review). It is intended to be accessible to the research community as well as interested policymakers.

Potential discussion questions for COSDAM:

1) How well do the conclusions listed in the introduction to each section appropriately convey the key takeaways from the Board’s Program of Preparedness Research?

2) Overall, is the technical report clear in describing the purpose of the Board’s program of preparedness research, the research design for each study, and the respective study findings?

3) Has enough context been provided to ensure that the online report is comprehensible to interested policymakers? Does the provided context support accurate interpretations of the research?
The online technical report will include the following individual documents:

**Content Alignment**
- Assessment Content Comparison: Methodology for Alignment Studies
- Preliminary NAEP and SAT Content Comparison: Mathematics
- Preliminary NAEP and SAT Content Comparison: Reading
- NAEP and WorkKeys Content Comparison: Mathematics
- NAEP and WorkKeys Content Comparison: Reading
- NAEP and ACT Content Comparison: Reading and Mathematics
- NAEP and SAT Framework Comparison: Mathematics
- NAEP and SAT Content Comparison: Mathematics
- NAEP and SAT Framework Comparison: Reading
- NAEP and SAT Content Comparison: Reading
- NAEP and ACCUPLACER Framework Comparison: Mathematics
- NAEP and ACCUPLACER Content Comparison: Mathematics
- NAEP and ACCUPLACER Framework Comparison: Reading
- NAEP and ACCUPLACER Content Comparison: Reading

**Statistical Relationship**
- Statistical Linking of National Results from NAEP and SAT
- Longitudinal Statistical Relationships for Florida NAEP Examinees: First-Year College Performance Outcomes

**Judgmental Standard Setting**
- Identification of Exemplar Occupations: Report
- Identification of Exemplar Occupations: Appendix A
- Identification of Exemplar Occupations: Appendix B
- Paper: A Study of “Irrelevant” Items: Impact on Bookmark Placement and Implications for College and Career Readiness
- Paper: Preparing Job Trainers to Describe Knowledge, Skills, and Abilities Measured in an Academic Assessment
- Paper Appendix: Preparing Job Trainers to Describe Knowledge, Skills, and Abilities Measured in an Academic Assessment
- Paper: The Standard for Minimal Academic Preparedness in Mathematics to Enter a Job Training Program
- Paper: The Standard for Minimal Academic Preparedness in Reading to Enter a Job Training Program

**Survey**
- Survey on Postsecondary Course Placement Assessments: Technical Report
- Survey on Postsecondary Course Placement Assessments: Summary Report
- Survey on Postsecondary Course Placement Assessments: Data Tables

**Benchmarking**
- Benchmarking Study with Texas College Freshmen: Methodology Report
- Benchmarking Study with Texas College Freshmen: Project Feasibility Report
- Benchmarking Study with Texas College Freshmen: Appendix A
- Benchmarking Study with Texas College Freshmen: Appendix B
- Benchmarking Study with Texas College Freshmen: Appendix C
Attachment A-3

Proposed Preparedness Research Projects for 2013 NAEP

Continued research plans call for NAEP-SAT, NAEP-ACT, and NAEP-EXPLORE statistical linking studies, more research partnerships with states, analysis of course content prerequisites for job training programs and freshman college courses, and efforts to partner with experts in military occupational training. A summary of each proposed research study follows.

Potential discussion questions for COSDAM:

1) What other information would be helpful in shaping statements that can be used for reporting preparedness in NAEP reports?

2) To support the Board’s goal of reporting preparedness, which research studies should receive the greatest emphasis in planning for Phase 2 of the Board’s Preparedness Research Program?

In 2013, the Governing Board will partner again with the College Board, as it did in 2009, to conduct a statistical linking study at the national level between NAEP and the SAT in reading and mathematics. Through a procedure that protects student confidentiality, the SAT records of 12th grade NAEP test takers in 2013 will be matched, and through this match, the linking will be performed. A similar study at the national level is planned in partnership with ACT, Inc.

In addition, the state-level studies, begun in 2009 with Florida, will be expanded in 2013. Again using a procedure that protects student confidentiality, the postsecondary activities of NAEP 12th grade test takers in the state samples in partner states will be followed for up to five years using the state longitudinal data bases. Preliminary conversations have indicated the possibility of involving at least four states, FL, MA, MI, TN, in these studies. Others will be considered as time for completing the planning process and executing formal data sharing agreements permits. These studies will examine the relationship between 12th grade NAEP scores and GPA, placement into remedial versus credit-bearing courses, and scores on admissions and placement tests.

In 2013, linking studies between 8th grade NAEP in reading and mathematics and 8th grade EXPLORE, a test developed by ACT, Inc. that is linked to performance on the ACT, have been discussed with partners in two states, KY and TN. The objective is to determine the feasibility of identifying the point on the NAEP scales that indicate students are “on track” for being academically prepared for college and job training by 12th grade. As a foundation for the linking study, content alignment studies between 8th grade NAEP reading and mathematics and 8th grade EXPLORE would also be conducted.
The Governing Board is conducting a procurement (1) to design a comprehensive and multi-method evaluation of the grade 12 NAEP frameworks and item pools in both reading and mathematics as measures of academic preparedness for college and job training; and (2) based on the evaluation, to produce specific recommendations for changes that may be required to develop NAEP for 12th graders in reading and mathematics as valid measures of academic preparedness for placement in first year college courses without remediation in the subject areas and entry in job training programs that require at least three months of post-secondary training, but not a bachelor's degree in college.

Central to the validity of reporting preparedness of students on the NAEP grade 12 scale for reading and for mathematics is confirmation that the assessments actually measure the knowledge, skills, and abilities required for students to be academically prepared for college course work or for entry in job training programs. In this procurement, the Board seeks innovative, practicable design proposals for evaluations that will provide the foundation needed to make valid statements about academic preparedness.

Reporting on academic preparedness for college and job training is a challenging and important new direction for NAEP. Hence, the Governing Board is also conducting a procurement to seek proposals for research designs and studies that are feasible. The objective of the research is to conduct research that will advance the Governing Board’s efforts to identify locations on the 12th grade NAEP reading and mathematics scales that represent the knowledge and skills to qualify for training in various occupations.
As part of the ongoing updates to COSDAM, this document includes an overview of each study type.

Content alignment studies are a foundation for the trail of evidence needed for establishing the validity of preparedness reporting, and are, therefore, considered a high priority in the Governing Board’s Program of Preparedness Research. The alignment studies will inform the interpretations of preparedness research findings from statistical relationship studies and help to shape the statements that can be made about preparedness. Content alignment studies were recommended to evaluate the extent to which NAEP content overlaps with that of the other assessments to be used as indicators of preparedness in the research.

A design document was developed by Dr. Norman Webb for the NAEP preparedness research alignment studies, and this design was implemented for the studies of the 2009 NAEP with the SAT and ACUPLACER in reading and mathematics. This design, with minor modifications, has also been used for the alignment of the 2009 NAEP with WorkKeys tests in these subject areas.

Content alignment studies for the first phase of the Board’s Program of Preparedness Research have been completed for NAEP in reading and in mathematics with WorkKeys, the SAT, and ACCUPLACER. In addition, a content alignment study was designed and conducted by ACT for the ACT and NAEP in reading and mathematics before the content alignment design document was developed.

For statistical linking, there has been a study to relate SAT scores in reading and in mathematics to the national sample of NAEP scores for grade 12. The objective was to provide a statistical linking of SAT and NAEP scores for all students in the 2009 grade 12 NAEP who had taken the SAT by June 2009. ETS staff reported that the match rate of approximately 33% of NAEP scores to SAT scores compares favorably to the national SAT participation rate of approximately 36% of public school students. The final sample used for linking the NAEP reading and SAT critical reading included approximately 16,200 students. The correlation between the two reading scales was 0.74. For NAEP and SAT mathematics, the linking sample included approximately 15,300 students, and the correlation between the math scales was 0.91.

The correlation between NAEP and SAT reading was found to be lower than that for mathematics. Research into those relationships, as suggested by the Technical Advisors for 12th Grade Preparedness Research, included: (1) inclusion of SAT scores in the NAEP conditioning
model, (2) identification and removal of outliers, (3) evaluation of demographics of outliers, i.e., sensitivity analysis, (4) evaluation of alternative SAT scores (e.g. highest, most recent, composite driven), and (5) disaggregation of NAEP reading students’ scores based on block content of their assessment booklet.

For longitudinal data, a series of analyses were conducted to examine statistical relationships for Florida’s NAEP examinees. NAEP’s 2009 state-representative sample of Florida 12th graders was used to match NAEP scores for reading and mathematics to student scores on several tests collected by the Florida Department of Education (FLDOE). The data sharing agreement with FLDOE provides access to scores for the SAT, ACCUPLACER, and WorkKeys. Additionally, ACT, Inc. has given permission to the Florida Department of Education to share ACT scores with the Governing Board for purposes of conducting the grade 12 preparedness research. We also plan to obtain employment data and salary data for Florida examinees, but access to those data was not included under the current data sharing agreement. A plan to allow for electronic transfer of data was developed to keep secure the identity of students, consistent with the NAEP legislation, FLDOE requirements, and requirements of each assessment program.

Records for roughly half of the Florida grade 12 NAEP examinees in 2009 could be matched to an ACT score and half to an SAT score. This match rate is consistent with other data for Florida students. The match of WorkKeys scores to the total 2009 state NAEP sample of 12th graders was only about 6%. FLDOE reported that around 89,300 Florida 12th graders were enrolled in vocational-technical programs in school year 2008-09. The match of WorkKeys examinees to NAEP examinees was not sufficient to warrant additional analyses for the 2009 cycle. The state of Florida has only recently implemented the testing of high school students in vocational programs with the WorkKeys exam, and we anticipate that the number of examinees will increase in subsequent years.

A series of judgmental standard setting studies was planned to produce preparedness reference points on the NAEP scale for entry into job training programs and for placement in college credit-bearing courses. Within this category of studies, the Technical Panel for 12th Grade Preparedness Research placed highest priority on the judgmental studies related to preparedness for job training programs in 5-7 exemplar jobs. This priority is largely related to the paucity of national data available for statistical studies in these areas. Unlike most other studies of preparedness for post-secondary activities in college or job training programs, the Governing Board has not assumed that prepared for college and prepared for the workplace are the same. Rather, our studies are aimed at determining whether the level of performance on NAEP is approximately the same or significantly different for entry in job training programs for the occupations included in our research studies and placement in credit-bearing college courses that fulfill general education requirements for a bachelor’s degree.

In order to maximize the standardization of judgmental standard setting (JSS) studies within and across post-secondary areas, a design document was developed to specify the number of panelists, the eligibility criteria for panelists, the procedures for drafting and finalizing borderline performance descriptions, the methodology to be implemented, feedback to be provided, key
aspects to be evaluated, and reports to be produced. The methodology and basic procedures specified for the design of these studies were those implemented for the achievement levels-setting process for the 2006 grade 12 economics NAEP and for the 2009 science NAEP for grades 4, 8, and 12.

The five exemplar jobs approved by COSDAM for inclusion in these studies are as follows:
1. automotive master technicians
2. computer support specialists
3. heating, ventilation, and air conditioning technicians
4. licensed practical nurses
5. pharmacy technicians

A pair of replicate panels with 10 panelists each was convened for each subject and post-secondary area for a total of 24 operational panels.

A survey of two-year and four-year post-secondary institutions was conducted in Fall 2011 to gather information regarding (1) the placement tests used and (2) the cut scores on those tests in reading and mathematics below which need was indicated for remedial/developmental courses in reading and mathematics, and at or above which placement in credit-bearing entry level courses was indicated. The sample of accredited postsecondary education institutions was nationally representative. A weighted response rate of 81% was achieved.

Benchmarking studies in the preparedness research context are studies in which NAEP is administered to groups of interest, e.g., college freshmen enrolled in credit-bearing college level courses that fulfill general education requirements for a four-year degree without the need for remediation. Determining the average NAEP performance of this group would then provide a “benchmark” score that can be considered as one of the reference points on the NAEP scale. A benchmarking study in combination with reference points from other studies in the Program of Preparedness Research can assist the Board in determining the areas of the NAEP scale that indicate preparedness. A benchmarking study of Texas college freshmen was planned, and it had the support of the Texas Commissioner of Higher Education and the cooperation of nine Texas higher education institutions. A small scale pilot study to evaluate the feasibility of the study design was implemented.

The Governing Board and the National Center for Education Statistics (NCES) collaborated on the implementation of this small scale pilot study, which was carried out by Westat, the NAEP sampling and administration contractor to NCES. The data collection phase for the pilot ended on October 15, 2010. Of the eligible sample of 1,234 students, 255 actually attended a NAEP session, for an overall response rate of 20.7 percent. As announced at the November 2010 meeting of COSDAM, NCES, Westat, and Governing Board staff met to discuss alternatives. Board staff decided that we will not proceed to the operational phase of this study due to low participation rates and the lack of feasible alternatives to increase participation.
No additional benchmarking studies are planned for the 2009 NAEP preparedness research.

For additional background information, the following list presents a brief description of the assessments that the Technical Panel on 12th Grade Preparedness Research recommended for analysis in NAEP preparedness research. Many of these assessments are the primary focus of the proposed content alignment studies and statistical relationship studies. In each case, only the mathematics and reading portions of the assessments are the targets for analysis, although analyses with the composite scores may be conducted.

- **ACCUPLACER** – ACCUPLACER is a computer adaptive test used for college course placement decisions in two-year and four-year institutions. It is produced by the College Board and includes assessments of sentence skills, reading comprehension, arithmetic, elementary algebra, college level math, and written essays.

- **ACT** – The ACT assessment is a college admissions test used by colleges and universities to determine the level of knowledge and skills in applicant pools, including reading, English, and mathematics tests. ACT has *College Readiness Standards* that connect reading or mathematics knowledge and skills and probabilities of a college course grade of “C” or higher (75%) or “B” or higher (50%) with particular score ranges on the ACT assessment.

- **ACT WorkKeys** – WorkKeys is a workplace focused set of tests that assess knowledge and skills in communication (business writing, listening, reading for information, writing) as well as problem solving (applied technology, applied mathematics, locating information, observation). There is also an interpersonal skills section of WorkKeys.

- **COMPASS** – ACT Compass is a computer-adaptive college placement test. It is produced by ACT and includes assessments of Reading, Writing Skills, Writing Essay, Math, and English as a Second Language.

- **SAT** – The SAT reasoning test is a college admissions test produced by the College Board. It is used by colleges and universities to evaluate the knowledge and skills of applicant pools in critical reading, mathematics, and writing. The College Board has provided SAT score data to be used in research studies to establish a statistical relationship between the SAT and NAEP.
Update: NAEP-TIMSS Linking Study

In November 2009, the Governing Board unanimously adopted a resolution in support of studies to statistically link NAEP and international assessments in 2011, including the Trends in International Mathematics and Science Study (TIMSS). The Board noted that the timing of this assessment presented a unique opportunity to have U.S. students take both NAEP and one of the international assessments in the same grade and subject, enabling statistical linking of the two sets of results. Accordingly, the Governing Board added 8th grade science at the national and state levels to the NAEP schedule of assessments in 2011 and moved the state and national science assessments to a once every four year schedule in 2015 and thereafter, to provide opportunities for future linking studies with TIMSS.

The goal of conducting the NAEP-TIMSS linking study is to enable states to interpret their NAEP results in an international context, with the possibility of translating the state’s 8th grade NAEP scores in mathematics and science into TIMSS-equivalent scores.

At the November 30, 2012 COSDAM meeting, NCES will provide a closed session confidential briefing to report on current methodological issues, findings, timeline for completion of analyses, and reporting options. The presentation will also include a brief overview of the study goals and research design—a summary document is attached here on page 145 with some of these details. This briefing will be a timely opportunity for COSDAM to provide feedback and guidance on the methodological issues and the reporting options.
Plans for Linking NAEP and TIMSS

In an increasingly global economy, comparisons of student achievement in the United States to student achievement in other countries are of interest to the nation. The Trends in International Mathematics and Science Study (TIMSS), initiated in 1995, reports on the mathematics and science achievement of fourth- and eighth grade students in different countries and regions throughout the world. The National Center for Education Statistics (NCES) has launched a study to project TIMSS scores for the U.S. states that participated in the National Assessment of Educational Progress (NAEP). If the linking procedure is successful, the projected scores will enable the states to compare their performance with that of other countries.

NAEP measures student learning in the 50 states, several urban districts, and other U.S. jurisdictions in a way that permits comparisons over time to the nation and among the participating jurisdictions. TIMSS measures students’ mathematics and science learning in more than 60 countries including the United States. Unlike NAEP, TIMSS does not have an on-going state component.

This linking study targeted eighth-grade students. Mathematics and Science NAEP assessments were conducted in 2011 winter (January-March) and Mathematics and Science TIMSS were conducted in 2011 spring (April-June). In addition to the national level assessments, for the linking study, two representative national samples were tested on their knowledge of mathematics and science by taking both the NAEP and TIMSS assessments. One sample of eighth-graders took combined test booklets (i.e., including both NAEP and TIMSS items) in the winter of 2011 as part of NAEP. The other sample of eighth-graders took combined test booklets in the spring of 2011 as part of TIMSS. In addition, a set of states administered TIMSS 2011 to state representative samples to help evaluate the accuracy of the linking projections. Those validation states are: Alabama, California, Colorado, Connecticut, Florida, Indiana, Massachusetts, Minnesota, and North Carolina.

Three NCES contractors--AIR, ETS, and HumRRO--were involved in this linking effort. Their major roles are:

- AIR will apply a statistical moderation linking technique to project TIMSS scores for the US states;
- ETS will apply two approaches, calibration and projection, to project TIMSS scores for the U.S. states;
- HumRRO will evaluate the quality of these various approaches and recommend to NCES which approach to use to report the final data.

Preliminary data analyses are complete. NCES plans to produce two reports with results from the NAEP-TIMSS Linking Study. The first will be a general public-oriented report intended for a general audience that will summarize findings from the study and their implications for the validity of comparisons of NAEP and TIMSS scores. NCES plans to release this highlights report by spring 2013. The second will be a technical report that will provide more details on the analytic approaches. The technical report is scheduled to be released by summer 2013.
## AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Presenter(s)</th>
<th>Attachment</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30 – 9:50 am</td>
<td>Welcome and Introductions&lt;br&gt;Overview of Committee Work&lt;br&gt;<em>Andres Alonso, Committee Chair</em></td>
<td></td>
<td>Attachment A</td>
</tr>
<tr>
<td>9:50 – 10:00 am</td>
<td>Review of NAEP Release: NAEP Writing 2011&lt;br&gt;<em>Stephaan Harris, NAGB Staff</em>&lt;br&gt;<em>Amy Buckley, Reingold Communications</em></td>
<td></td>
<td>Attachment B</td>
</tr>
<tr>
<td>10:00 – 10:10 am</td>
<td>Projected Schedule for Future NAEP Reports&lt;br&gt;<em>Angela Glymph, NCES</em></td>
<td></td>
<td>Attachment C</td>
</tr>
<tr>
<td>10:10 – 10:30 am</td>
<td>Update on Mega-States and Other Focused Reports&lt;br&gt;<em>Ebony Walton, NCES</em></td>
<td></td>
<td>Attachment D</td>
</tr>
<tr>
<td>10:30 – 11:05 am</td>
<td>Planning for Parent Outreach Activities&lt;br&gt;<em>Stephaan Harris and Ray Fields, NAGB</em>&lt;br&gt;<em>Amy Buckley, Reingold Communications</em></td>
<td></td>
<td>Attachment E</td>
</tr>
<tr>
<td>11:05 – 11:30 am</td>
<td>Puerto Rico Assessment and Reporting in 2013&lt;br&gt;<em>Emmanuel Sikali, NCES</em></td>
<td></td>
<td>Attachment F</td>
</tr>
<tr>
<td>11:30 am – 12:00 pm</td>
<td>Implementation of Policy on Students with Disabilities and English-Language Learners&lt;br&gt;<em>Grady Wilburn, NCES</em></td>
<td></td>
<td>Attachment G</td>
</tr>
</tbody>
</table>
National Assessment Governing Board

Reporting, Release, and Dissemination of NAEP Results

Policy Statement

The Nation’s Report Card™ informs the public about the academic achievement of elementary and secondary students in the United States. Report cards communicate the findings of the National Assessment of Educational Progress (NAEP), the only continuing and nationally representative measure of achievement in various subjects over time. The Nation’s Report Card compares performance among states, urban districts, public and private schools, and student demographic groups.

Introduction

NAEP collects data through representative-sample surveys and reports fair and accurate information on academic achievement to the American public. By law (P.L. 107-110, as amended by P.L. 107-279), NAEP is administered by the Commissioner of the National Center for Education Statistics (NCES) under policy set by the National Assessment Governing Board (“the Governing Board”), a bipartisan, independent policymaking body.

According to the statute, the Governing Board shall exercise “independent judgment, free from inappropriate influences and special interests” and in the exercise of its responsibilities, “shall be independent of the Secretary and the other offices and officers of the Department [of Education].” Among the responsibilities specifically delegated to the Governing Board are: (1) “develop guidelines for reporting and disseminating [NAEP] results”; (2) “take appropriate actions needed to improve the form, content, use, and reporting of [NAEP] results”; and (3) “plan and execute the initial public release of [NAEP] reports.”

To carry out these responsibilities, the Governing Board hereby adopts policy principles and guidelines for the reporting, release, and dissemination of The Nation’s Report Card.
As outlined in the appendix, this policy defines *The Nation’s Report Card* as, and applies to, the initial reporting of NAEP results from national, state, and trial urban district assessments (TUDA), and to other special reports or studies authorized by the National Assessment Governing Board, including printed reports and the initial release Web site.

**Delineation of NAEP Reporting, Release, and Dissemination Responsibilities**

The NCES Commissioner, under Governing Board policy guidance, is responsible for administering the assessment, ensuring the technical soundness and accuracy of all released data, preparing NAEP reports, and presenting NAEP results.

In addition to setting policy, Governing Board is responsible for ensuring policy compliance of Governing Board-authorized NAEP reports, determining their respective dates of release, and planning and executing the initial public release of NAEP results.

**Part I: Report Preparation and Content**

**Policy Principles**


2. The primary audience for *The Nation’s Report Card* is the American public.

   a. All reports shall be written in language appropriate for an audience of the interested general public, the majority of whom are unlikely to have a technical understanding of education statistics or assessment.

3. *The Nation’s Report Card* shall report data objectively, accurately, clearly, and fairly, in accordance with NCES data quality standards. Results shall be insulated from ideological and other special interests.

   a. *The Nation’s Report Card* shall include straightforward presentations of data. Reports may suggest correlations, but should not conclude cause-and-effect relationships. Any interpretation of results must be strongly supported by NAEP data.

   b. *The Nation’s Report Card* and its Web site may include references and links to the National Assessment Governing Board Web site, NCES Web site, and the NAEP Validity Studies Panel. Non-NAEP materials and links to non-NAEP resources shall not be included in initial release documents, with the exception of relevant federal and state government information, such as NCES surveys and other district, state, national, or international testing programs.
c. To improve public understanding of results, *The Nation’s Report Card* should contain information about Governing Board-approved NAEP contextual variables and subject-specific background information—as outlined in the *Background Information Framework for the National Assessment of Educational Progress* (adopted by the National Assessment Governing Board, 8/1/03)—when available and reliable. Reports may also contain other contextual information from trustworthy sources outside of the NAEP program, such as expenditures per pupil, student/teacher ratios, and student enrollment.

4. In accordance with the law, *The Nation’s Report Card* shall include results for the nation; states and school districts, when collected in conjunction with specific NAEP programs, respectively; and school types, disaggregated by subgroup whenever reliable. Subgroup results shall be prominently positioned to facilitate public review but shall not be used to adjust findings.

   a. Disaggregated subgroup data should be accompanied by information about demographic changes in the student population assessed.
   b. Results for states and school districts may be presented in alphabetical or rank order, accompanied by appropriate language to make the public aware of any data comparison limitations.
   c. Data shall be publicly released on inclusion and accommodation rates for all NAEP samples, including national, state, district, and school type. Results for students with disabilities and English language learners shall be presented separately.

5. *The Nation’s Report Card* shall report results by Governing Board-adopted achievement levels, average scale scores, and percentile distributions. Trend information shall be an important part of reports unless comparable and reliable data are not available.

   a. Reports shall contain clear explanations of achievement levels, including item maps and sample test questions and answers to illustrate what students in each grade assessed should know and be able to do at each achievement level.

6. All NAEP data determined by the NCES Commissioner to be valid and reliable shall be made available on the World Wide Web at the time of initial public release, except for data from limited special purpose samples and pilot studies. A separate, dedicated Web site aimed at a broad public audience—*http://nationsreportcard.gov*—shall be utilized for initial public releases.

   a. All released NAEP data shall be subject to NCES quality control procedures to ensure accuracy and completeness.
   b. At least one block of released NAEP questions shall be posted on the World Wide Web for each subject and grade for which results have been collected.
   c. Concise information on test content, methodology, performance standards, and scoring shall be included in all NAEP reports. More extensive material on these topics should be readily accessible on the World Wide Web.
7. Results of special studies authorized by the Governing Board will be reported after careful review of information quality and statistical validity. These shall be treated as initial public releases of *The Nation's Report Card*, and shall be subject to NCES quality control procedures and Governing Board policies.

8. The Governing Board shall adopt general guidelines to inform the development of *The Nation’s Report Card* and its Web site, and may set additional specifications for particular reports.

9. The Governing Board shall review the format and content of initial releases, including Web pages, to ensure compliance with Governing Board policy.

   a. *The Nation’s Report Card* shall contain a description of the policymaking roles and responsibilities of the National Assessment Governing Board, including a list of current Governing Board members, their affiliations, and regional locations.

---

**Part II: Public Release of NAEP Results**

**Policy Principles**

1. Release activities shall be planned and executed by the National Assessment Governing Board. The Governing Board shall determine the release date, time, embargo policies, and manner of release for *The Nation’s Report Card*, as covered by this policy.

   a. After the Governing Board has approved the final draft of *The Nation’s Report Card*, including the pages that will be made available through the initial release Web site, the Chairman of the Reporting and Dissemination Committee, on behalf of the Governing Board, shall determine the date of the initial public release, in consultation with the Chairman and Executive Director of the National Assessment Governing Board and the NCES Commissioner.

   b. The initial release shall be completed within 30 days of approval of the final draft of *The Nation’s Report Card*. In setting that release date, attention will be paid to balancing the priorities of an expeditious release with provision for adequate planning time, given the scheduling circumstances of the various parties involved.

   c. Prior to the initial public release, NAEP results may be provided on an embargoed basis to federal, state, and TUDA-district officials and members of the press.

2. The Governing Board shall be responsible for organizing and conducting the release event and related activities.

   a. A release plan shall be adopted by the Governing Board for each report. Elements of the plan may include issuance of a press release, a press conference and/or Web-based announcement, distribution of summary
findings and graphics, time period for the initial public release phase of http://nationsreportcard.gov, and other related activities.

b. The official press release announcing NAEP results shall be issued by the Governing Board. Accompanying statements from the Governing Board’s Executive Director or Governing Board members may also be issued.

c. At the press conference or other event for release of NAEP results, the NCES Commissioner or his/her designee shall present major data findings, accompanied by a written statement. The National Assessment Governing Board shall select members to provide individual commentary on the meaning of results. In addition, the Governing Board may invite other officials or experts to comment on the significance of the results in accordance with the approved release plan.

d. At press conferences, questions from the audience shall be limited to accredited members of the media. At other public release events, the Governing Board shall determine who may attend and ask questions or comment.

3. The Nation’s Report Card shall seek to encourage wide public attention to NAEP results and clear understanding of their meaning and significance.

a. Video materials may be prepared to accompany the release. These shall be clearly identified as having been provided by the Governing Board or NCES of the U.S. Department of Education. The video materials may only contain sound bites, background footage, and other information for journalists to develop their own stories.

4. Release procedures shall underscore the credibility of The Nation’s Report Card and encourage the participation of schools, school districts, and states in NAEP.

a. NAEP data in statements distributed at The Nation’s Report Card initial public release events shall be checked for accuracy by NCES.

5. The Nation’s Report Card releases shall be clearly separated from any ideological or other special interests.

a. Activities related to the initial public release of The Nation's Report Card shall not be used to disseminate any materials unrelated to NAEP. No materials of any kind may be distributed at an initial release event without the prior approval of the Governing Board.

6. The National Assessment Governing Board will cooperate with the NCES Commissioner in the release of technical reports, working papers, and secondary analyses not covered by the policy.

7. The Governing Board will develop a reporting schedule each year for upcoming NAEP assessments based on data review and report production plans that are provided and updated by NCES.
**Part III: Dissemination and Outreach**

**Policy Principles**

1. Information from *The Nation’s Report Card* shall be disseminated through the media, the World Wide Web, and special publications and materials. Efforts shall be made to develop widespread public awareness of NAEP data and their meaning and of the value of *The Nation’s Report Card* to the nation and participating jurisdictions.

   a. NAEP results shall be available in both printed and electronic form, including on *The Nation's Report Card* Web site, at the scheduled time of release and in the permanent record.

   b. To build public awareness of *The Nation’s Report Card*, the home page of the initial release Web site shall remain on-line and include links to previous releases. This homepage shall link to respective pages found on the NAEP Web site.

2. To build understanding of *The Nation’s Report Card* and the data it reports, other information about NAEP may be disseminated at the time of the initial release and on a continuing basis.

   a. Informational materials accompanying results shall explain the mission and value of *The Nation’s Report Card* in clear and compelling terms.

3. *The Nation’s Report Card* and supplementary NAEP materials shall be made available through a wide network of education, business, labor, civic, and other interested groups and to policy makers and practitioners at all levels of education and government.

   a. *The Nation’s Report Card* shall be distributed promptly to governors and chief state school officers, as well as to superintendents of TUDA districts. The reports shall be posted on the World Wide Web immediately at the time of initial release, with printed copies available to the public upon request.

   b. Notification of upcoming releases shall be widely disseminated. Schools and school districts participating in NAEP samples shall be provided with information on how to access reports electronically and obtain printed copies upon release.

   c. NCES and Governing Board staff shall encourage national and state organizations that are interested in education to disseminate NAEP results to their members.

   d. The NCES Commissioner and staff, Governing Board members and staff, and NAEP State Coordinators are encouraged to increase awareness and understanding of NAEP among the public, educators, and government officials. They are encouraged to speak about the
NAEP program to a variety of audiences; at meetings and conferences of national, state, and local organizations; on radio and television; and to writers for magazines and newspapers and other members of the media.

e. Talking points on key data findings shall be developed for each release and distributed to Governing Board members.

4. A variety of materials shall be developed, appropriate to various audiences, to carry out NAEP dissemination. Key audiences for these materials shall include the interested general public, policymakers, teachers, administrators, and parents.

5. Detailed data on cognitive results, Governing Board-approved contextual variables, and subject-specific background information (as outlined in Part I, Policy Principle 3, Item C) shall be made readily available through the World Wide Web to all those wishing to analyze NAEP findings, subject to privacy restrictions. Additional restricted data shall be available for scholarly research, subject to NCES licensing procedures.

   a. The limitations on interpretations, conclusions, and recommendations in official NAEP reports (as outlined in Part I, Policy Principle 3) shall apply fully to any materials disseminated as part of the NAEP program by NCES and the Governing Board.

   b. Researchers receiving secondary analysis grants from NCES may analyze data and provide commentary. Their reports may be disseminated by NCES if they meet NCES standards.
Appendix

NAEP Initial Release Reporting Covered by this Policy

The Nation’s Report Card™

The primary means for the initial public release of NAEP results shall be a summary report in each subject, known as The Nation’s Report Card™ and intended for the interested general public. The reports shall be made available in both print and electronic (Web-based) form. These reports shall present key findings and composite and disaggregated results. The printed reports shall be relatively brief, and written in a clear, jargon-free style with charts, tables, and graphics that are understandable and attractive. Data tables may be included in an appendix, either bound into the report or printed separately. This format shall be used to report key results for the nation and the states and of NAEP Trial Urban District Assessments.

A separate, dedicated Web site for the initial release of NAEP results shall be focused on a broad public audience, including less sophisticated users of the technology. The URL – http://nationsreportcard.gov – should be readily located via Internet search engines. Key NAEP findings will be available, clearly organized and prioritized. World Wide Web pages shall provide key findings, including composite and disaggregated results, as well as access to more extensive data sets.

Individual State and School District Reports

Relatively brief reports of key results shall be prepared for individual states, as well as for TUDA-participating school districts. All reports shall contain composite and disaggregated data, and may include an appendix with data tables.

Special Studies and Reports

Special studies and reports authorized by the National Assessment Governing Board and based on NAEP data collections will focus on specific topics of public interest and educational significance. They are aimed at policymakers and interested members of the public. They may include newly released data as well as data previously released that are analyzed to address issues identified by the Governing Board.
NATIONAL ASSESSMENT GOVERNING BOARD
STRATEGIC COMMUNICATIONS PLAN

INTRODUCTION

The theme of this plan is getting beyond the scores and NAEP releases to expand outreach of the Governing Board and NAEP. The two main objectives are:

1. Enhance and elevate the NAEP brand as the gold star of academic assessment and thought leadership in advancing excellence in achievement reporting.

2. Strengthen the relevance and use of NAEP – The Nation’s Report Card – results and NAEP research and resources by existing and new audiences.

To achieve these communications objectives, there will be a new approach that includes the following areas of engagement:

1. Practice consistent, year-round outreach and engagement with stakeholders and audiences.

2. Enhance collaboration with NCES and other entities involved with NAEP.

3. Use multiple communications channels, including social media.

4. Mobilize stakeholders and partners.

The Governing Board defines its audience as the general public. Effective communication requires breaking down the audience into segments based on their level of interest in the Board’s work and education in general, how they might use the Board’s information, and their capacity and tools to influence and effect change. This segmentation will allow resources to be targeted and used most efficiently by delivering messaging and information that are most pertinent to each audience.

The target audiences are as follows:

- **General Public** – the broad grouping of individuals who would be inclined to be receptive to effective messaging and information about NAEP.

- **Education Policymakers** – federal, state, and local officials with responsibility for enacting legislation and policies affecting elementary and secondary education.
Higher Education – educators and administrators of postsecondary institutions, including two- and four-year colleges and trade schools.

Business Leaders – Public and private sector employers, including the military, which are interested in the knowledge and skills of entry-level workers.

Education and Workforce Stakeholder Groups – membership, advocacy and policy groups addressing education and workforce issues.

K-12 Teachers – professionals in public, private, or charter schools who teach K-12.

Parents – families of K-12 students in public, charter, and private schools.

Each of these audiences will require specific messaging and a well-defined “call to action,” such as:

- Learn more about NAEP and the Governing Board.
- Understand how NAEP and the work of the Governing Board may be relevant to issues that are important to you.
- Use NAEP and the Governing Board as a resource in your pursuits.

COMMUNICATION STRATEGIES

The Governing Board’s communications plan is built on six distinct but integrated strategies focused on the most effective ways to educate and engage these target audiences. The strategies are designed to leverage the reach and impact of messaging delivered through other organizations, the media, and the Internet. At the same time, they provide the flexibility needed to pursue timely opportunities. Specifically, these strategies will use:

I. Report Card Releases – Reinventing the release events to reach broad audiences with greater impact and use the releases as a catalyst for other Board communications efforts.

II. Stakeholder and Partnership Outreach – Identifying organizations with valuable contacts and communications vehicles for spreading the Governing Board’s messaging.

III. Traditional Media – Using targeted media relations with traditional print and media outlets that provide skilled and trusted educational reporting.

IV. Social Media – Identifying and participating in emerging electronic media that reach the Board’s targeted audiences and offer interactive, real-time discussion formats.
V. Website Development – Enhancing the site to ensure that target audiences can readily find it, use it as a resource for both learning about and informing education initiatives, and pursue web tactics to increase traffic and impact.

VI. NAEP Communications Alignment – NAGB and NCES working together to review branding, materials and outreach.

I. REPORT CARD RELEASE STRATEGY

Release of The Nation’s Report Card will be conducted as part of a comprehensive, integrated communication campaign with a series of planned activities designed to generate traditional news coverage, to disseminate information about the assessment to stakeholder groups, and to further position The Nation’s Report Card as the most trusted national yardstick of student achievement. This can be accomplished through the following:

- **Webinar-style NAEP Releases.** With declining attendance and higher costs of renting venues, the traditional press conference is not giving the Board the best return on its investment. We recommend online webinar releases whenever possible and appropriate for future releases. So panelists can participate via Web-Ex and graphics and presentations can be seen by the viewing public. However, for releases like TUDA and Grade 12 Reading and Mathematics that involve specific cities or states, we leave open the option of having the more traditional style of release in a city or state that would involve local leaders as guests and panelists and add a unique angle to the release.

- **Strategic Release Dates.** Choose Report Card release dates (within the dictates of Board policy and NCES timeline) that optimally use media cycles, coinciding events, and other opportunities to leverage attention so that the release is driven by a date not vice versa.

- **More Accessibility to Media and Other Stakeholders.** The Board can take important and innovative steps to expand Report Card outreach to media and others, by facilitating better access through methods such as:
  
  - Pursue meetings and deskside briefings with key education journalists to illuminate them on various data, trends, and related efforts.
  - Issue a post-event news release that updates the reactions to NAEP results, gathering some of the best quotes from superintendents, parents, and other stakeholders and using them in another round of outreach to relevant groups.
  - Conduct phone chats with journalists and stakeholders before and after the release to help shape and influence media stories on NAEP.
  - Pitch the participation of event panelists and the Board chair and executive director in online events, including web chats, online forums, or discussion room Q&As with major news organizations such as the Washington Post.
Utilizing Web Site and Social Media. The Board should harness its web site and social media opportunities to extend the life of each Report Card. Several ideas include:

- Obtain video and audio sound bites of Governing Board staff, members, and other panelists form each event to disseminate to media and post online.
- In advance of each release, create a “splash” page on the www.nagb.org to host all materials related to the event, including bios of panelists, facts from past and related releases, information about relevant Board task forces and commissions to build momentum for the event.
- Develop an integrated social media strategy that links to the splash page that will help create a following on social networking sites leading up to the launch.

II. Stakeholder and Partnership Outreach

As a highly respected, independent source of unique objective data, the Governing Board is an attractive partner for numerous organizations. Relationships are mutually beneficial: the Board gains the support of other respected organizations and another outlet for its message, while the partner’s stature and message are also enhanced. Partnership activities can range from simply establishing website links to publishing reports and newsletters; co-sponsoring workshops, events, and forums; creating awards programs; actively participating in partners’ initiatives and conferences; and disseminating NAEP resources to organizational constituents.

Potential Partner Types

- The Media
- Colleges and Universities
- Think Tanks
- Education Advocates
- Parent Groups
- Foundations
- Private Companies
- Minority Advocacy Groups
- Governmental Organizations
- Individuals
- Other Testing Entities

Recommended Partnership Activities

Implementing a partnership strategy involves several steps to review, vet and establish the optimal partnership. The following list suggests a handful of specific ideas for activities for the Governing Board to undertake with potential partners. It ranges from big events to daily interactions and demonstrates the cumulative power of partnership development.
This course of action will entail such initial tasks as developing a list of recommended partners and related database; conducting research on priority stakeholders in each audience category; creating a partnership scorecard that identifies the specific opportunity, approach, and outcome for each group; developing partnership outreach materials and other content; and conducting ongoing stakeholder monitoring to identify partnership opportunities.

- **Events**
  - Present NAEP and related issues at education conferences.
  - Join with a teacher group like Teach for America, Phi Delta Kappa, or the National Staff Development Council to hold workshops for teachers on how to use NAEP.
  - Increase partnership with NCES and NAEP State Coordinators and local education groups to host state conferences and/or workshops in states or TUDA districts.
  - Partner with national and local PTAs to hold workshops for parents.

- **Content**
  - Co-sponsor a series of monthly webinars, with a different NAEP-related topic.
  - Create electronic newsletters on Board and NAEP subject-specific topics, using NAEP data and other information.
  - Publish booklets or one-pagers on Board initiatives, task forces, or important topics.
  - Partner with a media outlet or a local university to do background reports on TUDA cities to put the TUDA data in richer context.

- **Other Outreach**
  - Co-sponsor sections on the websites of NAEP partners, such as the Council of Chief state School Officers, and establish linking agreements with each.
  - Create an association of school districts that commit to using NAEP as a resource, partnering with them on assessment matters and making resources available to school staff and parents on how NAEP works.
  - Work with the Hechinger Institute (a non-profit organization based at Columbia University that focuses on training education reporters and producing in-depth national and investigative journalism on education) to showcase NAEP as a resource for reporters.
  - Join with a teacher’s group to give an annual award to a district, school, or principal that demonstrates best use of NAEP to improve instruction.

### III. Traditional Media Strategy

The traditional print and broadcast media are important vehicles for public education. However, NAEP coverage in the media has been largely limited to Report Card releases. The extent and value of traditional media coverage can be increased through a number of tactics and tools. These might include media events, a Board directory and experts “tip sheet,” op-eds, a story bank, and improved website usability for the press. The Report Card releases will be used as a
catalyst for generating ongoing use of NAEP data in coverage of broader educational policy issues. Ideas include:

- **More Events.** Create additional media events to release new frameworks, for example, or respond to emerging issues, and not just rely on Report Cards to generate news.

- **Media Training.** Conduct media training for Board members so they are comfortable and prepared for interviews.

- **Experts Directory.** Develop an expert’s directory of Board members, alumni and staff available for interviews and speaking opportunities, as appropriate.

- **Op-eds.** Write and pitch op-eds to various newspapers, magazines, and online sites on NAEP-related topics and Board endeavors.

- **Develop Contacts.** Cultivate media contacts and resources by regularly keeping in touch, seizing opportunities to send occasional emails and making phone calls.

- **Advance Outreach.** Conduct media pre-calls to create initial effective media placements on Board releases, events, and ongoing work.

- **Interactive Website.** Create dynamic online press kits and updating the “what’s new” section with press releases and video releases to entice more media interest.

- **Multiple Platforms.** Reporters for mainstream media now routinely produce web stories, videos, audio Q and As, and blog entries for each assignment. Outreach efforts should acknowledge these areas and tailor story ideas to a number of formats, helping reporters repurpose the material for different platforms.

- **Story Bank.** The Board should create a bank of broader story ideas that came out of release events, reports, and publications, and pitch those to journalists.

- **Database Expansion.** Expand media lists to include influential bloggers, online journalists, and others outside of traditional mainstream media.

**IV. SOCIAL MEDIA STRATEGY**

The Governing Board can engage in social media effectively while honoring its mission and maintaining its position of independence. Tactics include the following.

- **Create Facebook and Twitter Accounts.** The Board should develop profile pages for Facebook and Twitter to allow it to quickly and easily communicate with others using a variety of social media tools, including blogs, videos, images, tags, lists of friends, forums, and messaging. Alerts and postings on Board happenings and resources – events, data, background variables, etc. – can easily be disseminated and daily or weekly account
updates keep the Board in the spotlight between releases. Also, Board members and staff with Facebook and Twitter accounts already can help promote Board activities.

- **Blogs by Board Members.** Board members can rotate in writing a blog for www.nagb.org, with postings prompted by test score trends, framework issues, news topics, and the like. Board members can share insights, pose questions, and provoke thoughtful discussion without overstepping their bounds. Ideally, the content would then be picked up by other bloggers who will send it to others, generating a viral effect.

- **Disseminate E-mail Newsletters.** The Board can develop a robust newsletter that includes content of interest to various audience groups, including teachers, associations, alumni, parents, and students who may not be aware of the Board and NAEP. It will help to forge connections and a sense of community among these audiences.

V. **WEBSITE STRATEGY**

To position the Governing Board as a leading voice and authority on the complex issues of academic assessment and advancing educational innovation and excellence, its website should be positioned to play a more prominent role in achieving its objectives. This requires a redesign that supports and promotes the various communications channels and content of the entire communications plan outlined above, including:

- **Website Design.** The overall design should support the key content areas the website is targeting and be organized for easy navigation by subject or audience.

- **Search Engine Optimization (SEO).** Reingold, the Board’s communications contractor, will work with the Board and its web contractor, Quotient, to ensure the website receives full credit from search engines for content as it is published. This will involve ensuring design, word usage, tags, and the like will be positioned to help www.nagb.org show up on searches, so that people looking up phrases like “national assessment” and “high school achievement” would find us as well.

- **Keyword Research.** This process will help the Board identify high-traffic subject areas and the associated keywords or search terms most frequently used to research them. It will help shape the organization and development of content in the “language” of the Board’s target audiences, using keywords and phrases they use when navigating search engines to find information and relevant content. Because nearly 90 percent of all clicks from search engine results pages originate on the first results page, it is critical to understand which words and phrases the Board can realistically compete for to achieve a first-page position and then ensure those keywords and phrases appear in the target page’s URL, tile, meta description, image alt text, video narration, and/or body text.

- **Content Development.** Once the above preliminary work is done, the site’s content that is interesting and relevant to the Board’s target audiences must be continuously developed, integrating your targeted keywords, posted in the appropriate areas of the site, and refreshed regularly.
- **Link-Building & Outreach.** The Board should develop an effective link-building campaign that includes initial research to identify a broad list of other relevant and authoritative websites, blogs, forums and other outlets based upon the www.nagb.org content and keyword strategy and approved by Board members and staff. Reingold can then approach these sites with requests that should identify a specific page on their website and connect that content/topic back to a specific page on the www.nagb.org website with complimentary content, information or resources. The strategy would increase Board exposure and improve SEO efforts.

**VI. NAEP COMMUNICATIONS ALIGNMENT**

In the campaign’s first six months, Reingold will help the Governing Board work with NCES and other internal stakeholders to develop the foundation for expanded outreach. This foundation will focus on specific tasks under the strategies for stakeholder outreach and partnerships, traditional media, social media, Report Card releases, and the website.

**Overarching Tasks**

- **Review Governing Board branding.** Reingold will help the Governing Board and NCES to review the NAEP brand platform, determining how well its messaging and graphic elements distinguish and elevate NAEP and communicate the roles of the Governing Board and NCES.

- **Establish working group with NCES.** The Governing Board will create a NAEP working group with NCES to examine the activities and outreach undertaken by each group to determine if optimization is possible through greater coordination and collaboration. The group also can review the effectiveness of all NAEP materials and the Report Card release process, provide feedback, and recommend improvements.

- **Collaborate.** The Board staff and Reingold will define release plan roles, discuss deadlines, and streamline approval processes for release materials with NCES in a timeframe that enables optimal messaging, materials and content development.

- **Synergy.** The Board and NCES will work to align outreach strategies in communications and the website. For example, if NCES and its contractors sponsor a NAEP booth at a convention, then the Board can look into offering a member or staffer to give a presentation related to NAEP. Also, the Board and NCES can link to each other’s sites more regularly on NAEP-related items so that each group is contributing to increased exposure for the other.
THE NATION’S REPORT CARD
WRITING 2011, GRADES 8 AND 12
SEPTEMBER 14, 2012

Overview
The public release of Writing 2011, Grades 8 and 12 took place on September 14, 2012, at 11 a.m. EDT as a webinar. For this release, there were a total of 251 webinar participants (internal staff and contractors were not counted). In fewer than two business days, 18 original articles appeared in 209 outlets. An additional 210 news websites ran “The Nation's Report Card Releases Results” news release. Stories appeared in publications and on websites based in 43 states, Washington D.C., the United Kingdom and Australia.

Release Event
Webinar panelists included:
- Arthur Applebee, Distinguished Professor of Education; Chair, Department of Educational Theory and Practice; Director, Center on English Learning and Achievement, University at Albany, State University of New York
- Jack Buckley, Commissioner, National Center for Education Statistics
- Beverly Chin, Director, English Teaching Program, Department of English, University of Montana, Missoula
- Susan Pimentel, Educational Consultant; Curriculum Specialist and Member, National Assessment Governing Board
- Mary Crovo, Deputy Executive Director, National Assessment Governing Board (moderator)

Webinar Event Attendees (251 webinar participants)
The Nation’s Report Card: Writing 2011, Grades 8 and 12
Selected Media Clips

Nation’s Report Card: Even with Spell-check, Just a Quarter of Students Proficient in Writing
Associated Press, Sept. 14, 2012 – Christine Armario

NAEP Shows Most Students Lack Writing Proficiency
Education Week, Sept. 14, 2012 – Nora Fleming

Writing Scores Could Preview New Standards' Effects
USA Today, Sept. 14, 2012 – Greg Toppo

Teenagers’ Texting Time Fails to Translate Into Sparkling Prose

Nation’s Report Card: Writing Test Shows Gender Gap
CNN, Sept. 17, 2012 – Donna Krache

Most U.S. Students Lack Writing Proficiency, National Assessment Of Educational Progress Finds
The Huffington Post, Sept. 14, 2012 – Alex Kuczynski-Brown
Associated Press

Nation's Report Card: Even with Spell-check, Just a Quarter of Students Proficient in Writing
By Christine Armario

Students who have access to computers at home and regularly use them for assignments are more likely to be strong writers, a national exam suggests. But it also says just a quarter of America’s eighth- and 12th-grade students have solid writing skills.

Twenty-seven percent of the students at each of those grade levels were able to write essays that were well developed, organized and had proper language and grammar — 24 percent were considered proficient, 3 percent advanced. The remainder showed just partial mastery of these skills.

“It is important to remember this is first-draft writing,” said Mary Crovo, deputy executive director of the National Assessment Governing Board, which administers the Nation’s Report Card tests. “They did have some time to edit, but it wasn’t extensive editing.”

Students who took the writing test in 2011 had an advantage that previous test takers did not: computers with spell-check and thesaurus. Previously, young people taking the National Assessment of Educational Progress writing test had to use pencil and paper; the switch was made in line with changes in technology and a need for today’s students to write across electronic formats.

Because this was the first version of the computerized test, the board cautioned against comparing the results to previous exams. In 2007, some 33 percent of eighth-grade students scored at the proficient level, which represents solid writing skills, as did 24 percent at grade 12.

Crovo said most students already use such technology as spell-check on a daily basis. Without those tools, she said, “It’s as if years ago we had given them a pencil to write the essay and took away the eraser.”

She said word processing tools alone wouldn’t result in significantly better writing scores if students didn’t have the core skills of being able to organize ideas and present them in a clear and grammatical fashion.

Still, students in both grades who used the thesaurus and the backspace key more frequently had higher scores than those who used them less often. Students who scored below the 25th percentile were less likely to have computers at home: 87 percent said they did, compared to 99 percent were in the top quarter.
The technology gap was hinted at in other statistics as well: The lowest scorers reported less daily computer use for school assignments, and 44 percent fewer said they always used a computer to make changes to papers or reports.

Mark Warschauer, an education professor at the University of California, Irvine, said research consistently shows the use of computers in the classroom improves writing performance. He said students end up writing more, getting more feedback from peers and teachers and publishing more, all of which keeps them motivated.

“It just improves every aspect of the writing process,” he said.

The latest test results make a strong argument for more use of technology in English language programs at school, Warschauer said, as home access is more uneven.

The results at both grade levels showed a continuing achievement gap between white, black, Hispanic and Asian students. At the eighth grade, Asian students had the highest average score, which was 33 points higher than black students on a 300-point scale. At the 12th grade, white students scored 27 points above black students.

There was also a gender gap, with girls scoring 20 points higher on average than boys in the eighth grade and 14 points higher in 12th grade. Those who qualified for free and reduced price lunch, a key indicator of poverty, had lower scores than those who did not; there was a 27 point difference between the two at the eighth grade.

For the 2011 exam, laptops were brought into public and private schools across the country and more than 50,000 students were tested to get a nationally representative sample. Students were required to write essays that explained, persuaded or conveyed an experience.

Kathleen Blake Yancey, a professor at Florida State University who served on the advisory panel for the test, said one factor to keep in mind is that research shows most students in the United States don’t compose at the keyboard.

“What they do is sort of type already written documents into the machine, much as we used to do with typewriters four decades ago,” she said.

Yancey said for this reason there was some concern about having students write on computers as opposed to by hand. Likewise, having the advantage of spell-check assumes students know how to use it. And in some schools and neighborhoods, computers are still not easily accessible.

“There are not so many students that actually learn to write composing at the keyboard,” she said. Yancey added that many kids who do have access to computers are not necessarily using them to write at school, but to take standardized tests and fill in bubbles.

“Digital technology is a technology,” she said. “Paper and pencil is a technology. If technology were the answer, that would be pretty simple.”
After decades of paper-and-pencil tests, the new results from the “nation’s report card” in writing come from a computer-based assessment for the first time, but only about one-quarter of the 8th and 12th graders performed at the proficient level or higher. And the proficiency rates were far lower for black and Hispanic students.

With the new National Assessment of Educational Progress in writing, students not only responded to questions and composed their essays on laptop computers, but also were evaluated on how frequently they used word-processing review tools like “spell check” and editing tools such as copying and cutting text. Some prompts also featured multimedia components.

According to the NAEP report, released today, the switch from paper and pencil to a computer-based test is tied to recognition of the role technology plays in a 21st-century student’s life. In 2009, a hands-on and computerized science NAEP was administered, and all new NAEP exams are slated to be computerized, including, for example, a 2014 technology and engineering assessment administered entirely on computers.

“This is a very exciting time for us,” said Mary Crovo, the executive director of the National Assessment Governing Board, which sets policy for NAEP, on a conference call with reporters. “[Technology] is becoming more the norm than the exception in our nation’s schools and certainly the way students communicate in college and the workplace.”

With the new format, which is evaluated on a revised NAEP writing framework, the latest results are not comparable to past exams, but future tests will use these results as a benchmark. The most recent paper-and-pencil tests were administered in 1998, 2002, and 2007.

On the new writing NAEP, given last year, the nationally representative sample of students—24,100 8th graders and 28,100 12th graders—were asked to respond to two 30-minute writing prompts that asked them to persuade, explain, or convey experiences. Results show the percentages of students in each grade reaching the “basic,” “proficient,” or “advanced” levels, which reflect how well they could communicate purposeful messages to specific audiences, such as a college-admissions committee.

At the 8th grade level, for example, one exercise called “Lost Island” asked students to imagine they had arrived on a remote island and listen to an audio file that included nature sounds and lines of a journal read aloud. Students then were required to write personal stories that chronicled an experience they would have had on the island, had they been there.
To reach “advanced” on the exam, students told well-organized stories with strong details, precise word choices, and varied sentences, according to the NAEP report. Students at the “basic” level would use some detail in their stories, but organization was “loose,” sentence structure unvaried, and word choice limited.

Teachers of students who took the new exam were surveyed on how frequently they assign schoolwork to be completed on computers. The report finds that those students who were required by teachers to use computers more often to write and edit assignments for school performed better on the test.

Overall, only 27 percent of students in both grades tested scored at or above the proficient level in 2011. The data also reveal some persistent achievement gaps. For instance, at the 12th grade level, 9 percent of black students and 12 percent of Latinos scored proficient or above, compared with 34 percent of white students.

Also, females outperformed males at both grade levels. In 8th grade, 37 percent of girls scored proficient or above, compared with 18 percent of boys. Such performance differences for various populations were similar to those seen with the paper-and-pencil tests, according to NAEP data.

David P. Driscoll, the chairman of the NAEP governing board, saw reason for concern in the new data.

“We need to focus on supporting students beyond the ‘basic’ levels so that they have a solid grasp of effective writing skills,” he said in a press release.

Access to Technology

Beverly Ann Chin, a professor of English at the University of Montana, in Missoula, said the report provides insights on how students use technology to write. She also highlighted the stronger outcomes for students who used computers regularly in class.

“These findings support the importance of integrating computers into writing instruction,” she said in a statement. “When teachers encourage students to use word-processing features on a regular basis, students learn how computers can facilitate their writing processes and improve their final product.”

Ms. Chin raised concerns about access to technology, noting survey data from the NAEP report suggesting that students from low-income families were less likely to be asked by their teachers to use computers to draft and review their writing.

“Students who are skilled in using technology tools in writing will be more successful in school, the workplace, and society,” she said.

A pilot test of the writing NAEP also was given to 4th grade students. Students at that grade level will be included in the regular administration of the exam moving forward.
Writing Scores Could Preview New Standards’ Effects
By Greg Toppo

Just one in four middle- and high-schoolers produced solid writing on a new, more rigorous federally administered exam, offering a glimpse of what schools nationwide may face as they move to a similarly tough set of writing standards over the next two years.

The findings, out Friday from the federal government's National Assessment of Educational Progress (NAEP), suggest that new standards in writing and other topics, due in 2014, could put pressure on teachers to raise kids' basic skills.

"This does telegraph what we might expect to see in those early 2014 assessments," said Elyse Eidman-Aadahl, director of national programs for the National Writing Project, a network of college-level instructors who train teachers nationwide.

The new NAEP results show that only 24% of students scored "proficient," representing what educators call "solid academic performance" at each grade level.

Today's results represent the first from a more rigorous test administered last year to 52,200 eighth- and 12th-graders by the U.S. Department of Education's National Center for Education Statistics. Part of a push to align what they learn with what colleges and workplaces require, the new framework resembles the Common Core, a series of new standards developed over the past several years and approved by 46 states. Eidman-Aadahl said the lackluster showing suggests that educators may soon realize how poorly many schools teach the topic. "It's going to take a lot of shift to get people teaching writing again," she said.

When they looked at how kids composed, examiners also found a curious phenomenon: Using government-issued laptops loaded with rudimentary word-processing software, students were asked to compose two longish pieces of writing in one hour. In the course of writing, records show, only about one in five students even touched the "cut," "copy," "paste" or "delete" buttons. Meanwhile, 100% hit the "backspace" key at least once.

The results suggest that the teenage "digital natives" backspaced their way through mistakes as their grandparents might. Either they weren't comfortable with basic editing functions or simply didn't bother. By contrast, 89% used the computer's spell-check function before they handed in their writing assignment.

The complete results are available online at: http://nationsreportcard.gov/writing_2011/
Bloomberg News Service

Teenagers’ Texting Time Fails to Translate Into Sparkling Prose
By John Hechinger

U.S. teenagers’ texting, tweeting and posting on Facebook hasn’t improved their writing, even when students have laptops with a spell-checking program.

Nearly three-quarters of the eighth- and 12th graders failed to achieve proficiency on a national writing test, according to a U.S. government report released today. For the first time, the exam let students use a computer, rather than pencil and paper.

Most students’ writing “falls far short of the well-organized, well-developed prose that connects with those they are trying to reach,” Susan Pimentel, a member of the U.S. Education Department board overseeing the test, said in a statement. That performance will hurt them in college, damaging their career prospects and earnings potentials.

The 2011 test, known as the Nation’s Report Card, adds to concern about American schoolchildren’s knowledge of math and science relative to other countries, particularly China, Japan and other Asian economic rivals. Lagging student performance has bedeviled U.S. presidents from Republican Ronald Reagan to Democrat Barack Obama.

The new version of the test, officially called the National Assessment of Educational Progress, offered students the tools of modern writing: a laptop with a word-processing program, including spell-checking, cutting and pasting and other editing functions, as well as a thesaurus.

More than 24,000 eighth graders and 28,000 12th graders took the exam. The report cards, which measure subjects such as math, reading, science and history, are the largest nationally representative of American student learning. Students have fallen short of national standards in other subjects, as well.

Writing Skills

On the writing exam, 24 percent of students were considered proficient in writing and 3 percent, advanced.

The 2011 results can’t be compared with the past pencil- and-paper exams. In 2007, the last time the government assessed writing, scores had increased from five years before, though most students also had poor writing skills.

On the latest report card, students who wrote more often at home did better on the test. So did those who made use of computerized tools during the exam to revise their work or find words on the thesaurus. By contrast, those who relied heavily on spell check scored lower.
Mirroring demographic results on other tests, Asian students outperformed other ethnic groups in eighth grade. In 12th grade, white, Asian and multiracial students performed comparably. Whites did better than blacks and Hispanics. Poor students lagged richer ones. Private and Catholic schools scored higher than public schools.

Girls beat boys by a higher margin than for any other subject. On questionnaires, girls said they wrote more and were more likely to call it a favorite activity.

Organization, Detail

The Education Department judged writing based on organization, level of detail and variety of sentence structure. The students wrote for 30 minutes. The results were evaluated as first drafts, rather than polished works. Students wrote narratives, including fiction, and essays.

In one eighth-grade assignment, students imagined they were stranded on an island.

A weak response featured the following sentence: “There is five guys and five girls, the girls will get to sleep inside the plane so they don’t get to cold or scared.”

One of the stronger passages built suspense, conjuring a threat from dinosaurs:

“We slowly trudged through the dense sand back to our boat, which was now in sight. But it seemed that time stopped and the next thing I saw was a gigantic foot on top of our ruined boat. A dinosaur’s face 20 feet above leered down at us and growled.”
When it comes to writing, girls are better than boys.

That’s a generalization, but it’s one that is supported by the latest writing test from the National Assessment of Educational Progress (NAEP), better known as the Nation’s Report Card.

The test, taken by 24,100 eighth-graders and 28,100 students in the 12th grade, was administered in early 2011. NAEP tests in different subjects have been given to students in the U.S. since 1969. This year, however, marked the first time that the writing test was computer-based. Students were able to take advantage of editing software and other writing tools, such as spell check and a thesaurus, as they crafted their writing samples.

Since this was the first large-scale writing assessment designed to be taken on a computer, the National Assessment Governing Board, which administers the NAEP, said that it could not make comparisons to previous “paper and pencil” writing tests.

Students were asked to perform writing tasks in three areas: To persuade, trying to change the reader’s point of view; to explain, trying to broaden a reader’s understanding of a topic; and to convey experience, trying to provide an account of a real or imaginary experience to a reader.

The NAEP writing test is a scaled test with a range of 0-300, and a mean score of 150. “Achievement levels” were set along that scale for the categories Below Basic, Basic, Proficient and Advanced.

Among eighth-graders, about 3% scored advanced, 24% scored proficient or above, 54% basic, and 20% below basic. (Because the numbers were rounded, they do not add up to 100%).

Among 12th-graders, about 3% scored advanced, 24% scored proficient or above, 52% basic and 21% below basic.

According to the board, performances varied by race, ethnicity, gender, school location and other factors, such as parents’ educational attainment. But the most notable achievement gap was between males and females in both eighth and 12th grades.

On average, female students in the eighth grade scored 160; their male counterparts scored 140.

On average, female students in the 12th grade scored 157; males scored 143.

Education analyst Susan Pimentel, one of the team presenting the test scores on Friday’s NAEP conference call, said that while this test cannot determine cause and effect, there are some clues as to why the gap exists. Students were surveyed to find out some additional information about them as they took the test. Among those surveyed, said Pimentel, 53% of girls agreed or strongly
agreed that “Writing is one of my favorite activities”, but only 35% of the boys felt that way. Since writing improves with practice, she said this is “an important variable to observe.”

According to the survey, 39% of 12th-graders said they write only one page of homework or less per week in English, which is also of concern as high school teachers focus on college readiness as one of the goals of the Common Core State Standards, said Pimentel.

The NAEP test also revealed that regardless of income, students who frequently use computers to draft and revise their writing performed better than those who regularly do not.

To improve on writing scores, the board encourages engaging boys in “meaningful” writing as part of the curriculum and providing all students with opportunities to use computers to write and edit whenever possible.
Most U.S. Students Lack Writing Proficiency, National Assessment Of Educational Progress Finds
By Alex Kuczynski-Brown

Only roughly one quarter of eighth and 12th graders are proficient in writing, according to results from the National Assessment of Educational Progress' first-ever computer-based writing assessment. The new framework represents a move away from the traditional paper-and-pencil format that has dominated the testing scene for nearly four decades.

NAEP's exams are considered the gold standard measurement of student achievement. In May, results showed that about a third of eighth graders who took its science exam were proficient, a statistic National Science Teachers Association's interim director Gerry Wheeler slammed as "unacceptable." Similarly, only 32 percent of students performed at the proficient level on NAEP's math exam in 2007, ranking the U.S. 32nd out of 65 countries that were tested on the 2009 Programme for International Student Assessment (PISA), NAEP's international equivalent. This trend also appears to hold true for writing, though the format may have changed.

Drawing from a sample of 24,100 eighth graders and 28,100 12th graders representing both public and private schools, the 2011 writing assessment asked students to complete two 30-minute tasks, each of which was designed to measure one of three communicative purposes: to persuade, explain or convey experience. The prompts were presented in multimedia formats that included video or audio segments, newspaper articles, real-world data and other materials around which students could formulate a response. They recorded their answers on a laptop that featured commonly used word-processing tools such as spell check and a thesaurus.

"[Those who developed the framework] felt it was definitely time that we start assessing our students using computers," Dr. Mary Crovo, deputy executive director of the National Assessment Governing Board, said in a statement. "This is becoming more the norm than the exception in our nation’s schools, and it is certainly the way that students write and communicate in higher education and in the workplace. So we feel very strongly that this is a solid assessment for 21st century skills."

Results showed 24 percent of students at both grade levels scored at the proficient level on the writing assessment, while 54 percent of eighth graders and 52 percent of 12th graders met the benchmark for "basic." Around 20 percent of both grades performed below basic, while only 3 percent scored at the advanced level.

Among eighth graders, Asians outperformed other racial/ethnic groups, averaging a score of 165 on a 300-point scale. A mean of 150 was set for both grades. At the 12th-grade level, however, white students, Asian students and students of two or more races performed comparably. In both grades, African American and Hispanic students had lower average scores than the other races.
In addition to assessing students’ writing ability, the new computer-based format of the exam allowed test administrators to collect extensive information on 24 separate student “actions,” including keystrokes, backspacing, deletions and their use of spell-checking programs. Results found that at both grade levels, students who used the backspace key and thesaurus tool more frequently scored higher than those who did not routinely engage in these practices. Furthermore, English language learners were less likely to use the thesaurus tool than non-English language learners.

Dr. Jack Buckley, commissioner of the National Center for Education Statistics, said in a press call that the standards of proficiency were tailored to reflect the computer-based nature of the assessment, and that students’ writing was evaluated holistically -- taking into account development of ideas, organization and language facility and conventions.

Thus, while the spell check tool might have provided students with an advantage they did not have when taking the old paper-and-pencil tests, spelling was only evaluated under the category of “use of conventions,” and to the degree that it might interfere with what the student was saying.

“The raters who are scoring the students’ results were asked to consider these as first drafts. They don’t expect to see a polished final report; they’re expected to see first-draft quality,” Buckley said, later pointing out that the word processor tool is not going to result in significantly better writing if the student is not already fluent in expressing his or her ideas.

While the new computerized framework makes it difficult to directly compare results to the past, Buckley acknowledged, “there was not a lot of difference in levels of proficiency” from 2007, when the most immediate prior writing assessment was administered.

On the 2007 pencil-and-paper tests, 35 percent of eighth graders and 25 percent of 12th graders scored at or above proficient -- on par with 2011’s results, at least for 12th grade.

Additionally, female students in both grades scored higher than their male counterparts on the 2011 writing assessment -- a pattern that is consistent with previous results, according to Buckley.

Crovo, the deputy executive director of the National Assessment Governing Board, said that the NAEP hopes to add fourth graders to the sample in the near future.

Said Crovo, “We’re hopeful this new 2011 computer-based assessment can serve as a baseline for looking at trends over time.”
# Upcoming NAEP Reports as of November 2012

<table>
<thead>
<tr>
<th>Report</th>
<th>Expected Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial NAEP Releases</strong></td>
<td></td>
</tr>
<tr>
<td>2009 and 2011 Reading Vocabulary</td>
<td>December 2012</td>
</tr>
<tr>
<td>2012 Economics</td>
<td>April 2013</td>
</tr>
<tr>
<td>2012 Long-Term Trend</td>
<td>June 2013</td>
</tr>
<tr>
<td>2013 Mathematics Grades 4, 8, 12</td>
<td>October 2013</td>
</tr>
<tr>
<td>2013 Reading Grades 4, 8, 12</td>
<td>October 2013</td>
</tr>
<tr>
<td>2013 Mathematics TUDA Grades 4, 8</td>
<td>December 2013</td>
</tr>
<tr>
<td>2013 Reading TUDA Grades 4, 8</td>
<td>December 2013</td>
</tr>
<tr>
<td><strong>Other NAEP Reports</strong></td>
<td></td>
</tr>
<tr>
<td>2005 HSTS Math Curriculum Study</td>
<td>January 2013</td>
</tr>
<tr>
<td>Linking NAEP and TIMSS 2011 Mathematics and</td>
<td>March 2013</td>
</tr>
<tr>
<td>Science Results for the 8th Grade</td>
<td></td>
</tr>
<tr>
<td><strong>NAGB Reports</strong></td>
<td></td>
</tr>
<tr>
<td>Mega-States Report: Grades 4, 8, and 12</td>
<td>February 2013</td>
</tr>
<tr>
<td><strong>Other Related Reports from NCES</strong></td>
<td></td>
</tr>
<tr>
<td>Applying Simulation Methods to National TIMSS Data to Estimate State-Level Average Scores</td>
<td>October 2012</td>
</tr>
<tr>
<td>Revenues and Expenditures for Public Elementary and Secondary Education: School Year 2009-2010</td>
<td>October 2012</td>
</tr>
<tr>
<td>Dual Credit and Exam-Based Courses in U.S. Public High Schools: 2010-2011</td>
<td>November 2012</td>
</tr>
<tr>
<td>Debt Burden Among Non-Completers</td>
<td>December 2012</td>
</tr>
<tr>
<td>Title</td>
<td>Date</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Highlights from PIRLS 2011: Reading Achievement of 4th Grade Students in an International Context</td>
<td>December 2012</td>
</tr>
<tr>
<td>Highlights From TIMSS: 2011 Mathematics and Science Achievement of U.S. 4th and 8th Grade Students in an International Context</td>
<td>December 2012</td>
</tr>
</tbody>
</table>
NCES Assessment Data Release Timeline

LEGEND
- NAEP Report Cards
- NAEP Studies
- TIMSS
- PIRLS
- PISA
Releases in 2012

- 2009 and 2011 Reading Vocabulary
- 2011 TIMSS: Grades 4 and 8 (National only)
- 2011 PIRLS: Grade 4 (National only)

Releases in 2013

- Linking NAEP and TIMSS 2011 Mathematics and Science Results for the 8th Grade
- Mega States Report: Grades 4, 8, and 12
- 2012 Economics Report Card: Grade 12 (National only)
- 2012 Long-term Trend (LTT) Reading & Math: Ages 9, 13, and 17 (National only)
- 2013 Reading Report Card: Grades 4 and 8
- 2013 Reading Report Card: Trail Urban Districts (TUDA: Grades 4 and 8
- 2013 Mathematics Report Card: Grades 4 and 8
- 2013 Mathematics Report Card: Trial Urban Districts (TUDA): Grades 4 and 8

Assessment Data Collection Schedule 2013

- Reading: Grades 4, 8, 12
- Math: Grades 4, 8, 12
- Writing: Grade 4 (National only)
NCES is in the process of developing several reports of interest to the Governing Board.

A Focus on NAEP on gender gaps will examine the differences in performance between 4th and 8th grade boys and girls in NAEP mathematics, reading and science using 2011 data. It is currently in development and is expected to be released in the spring of 2013.

A profile of Black male students is expected to be released in the Fall/Winter of 2013. Focusing on 8th grade males, this report will use performance on NAEP mathematics, reading, and science as well as background questions. The report will also include information from other Department of Education surveys to help contextualize the data presented.

NCES is also planning reports on charter and private school performance, which will include 2013 results. Both reports will focus on 4th and 8th grade mathematics and reading performance. The reports will include comparisons with public school results. The two reports are expected to be released in 2014.

Prepared November 2012
Mega-States: An Analysis of Student Performance in the Five Most Heavily Populated States in the Nation

This report will provide NAEP results for the five Mega-States—the most populous states in the nation. The Mega-States are California, Florida, Illinois, New York, and Texas. Close to 40 percent of the nation’s students attend schools in these states. Additionally, eight of the ten most heavily populated cites are located in the Mega-States and they represent distinct regions of the country. They also have the highest number of English language learners (ELL) in the nation. Given the scope of these school systems and the challenges they face, outcomes in these states inform and influence decision makers regionally and nationally. An NCES report that tabulates, organizes, and discusses these specific results provides a needed service to the educational community.

CONTENT

The NAEP 2011 Mega-States report will discuss NAEP results for fourth- and eighth-graders in reading, mathematics, science, and writing in California, Florida, Illinois, New York, and Texas. In particular, the report will show how score changes in these states compare to score changes nationally. It will look at score gains made in these states overall and by major reporting groups (e.g., race/ethnicity, National School Lunch Program eligibility status, disability status, and English Language Learner status). These comparisons will be presented graphically within the report and online. The report also offers graphical representations of the percentage of students in selected student groups at or above Proficient. The report will present these data for the most recent assessment year, while online users may view how the data points have changed over time. The web report will present special data visualization tools for users to investigate how student performances in these Mega-States compare with the nation and each other and how their individual performances have changed over time. There will be tables and graphs displaying trends in scores and achievement-level results. Additionally, there will be a concise display of the performance for all Mega-States in one interactive figure.

PUBLICATION PLANS

The report will be issued in both printed and electronic formats. The printed report will contain the main findings, comparisons, and trends. Links will be embedded in the electronic form of the report to more detailed findings on the NAEP web sites. This companion Web report will complement this presentation with interactive displays and state ranking tools, and will also include contextual variables and additional demographic data.

Governing Board Review: 9/20/2012

Projected Release: February/March 2013
NAEP 2009 Mega-States Report

STATES INCLUDED
California, Texas, Florida, New York, and Illinois

RATIONALE
These five mega-states, each enrolling over 2 million students in public school, are of great importance to the United States as a whole because they contain the nation’s largest cities, ports, and industrial and financial centers. Together they account for almost 40 percent all U.S. public school students, and share common factors of diversity and size.

FORMAT, SUBJECTS, AND GRADES
The report will be designed primarily for the Internet, and will be accompanied by a short highlights document to be available both in print and on-line. It will present NAEP data across the curriculum at grades 4 and 8 for reading, mathematics, science, and writing. In addition, data for grade 12 reading and math will be presented for Florida and Illinois, the two mega-states that participated in the NAEP 12th grade pilot assessment.

Almost all the data included will be available on the NAEP Data Explorer. The mega-states report will repackage this material to permit much more accessible comparisons and analysis.

YEARS AND METRICS FOR REPORTING
Data will be reported for the most recent year assessed—2009 for reading, math, and science; 2007 for writing—with trends back to the first year in which the state participated in NAEP. For 8th grade math that would go back to 1990. Data will be presented in terms of both scale scores and achievement levels. Changes will be highlighted from the initial assessment year, the most recent previous assessment, and from 2003, the first year in which all states were required to participate in NAEP.

TIMELINE, FUNDING, AND RELEASE
The report will be issued in November-December 2010 following the initial release of national, state, and urban district results for all grades and subjects in the 2009 National Assessment. The report will be prepared by the National Center for Education Statistics and paid for from regular NAEP appropriations. The release will be conducted by the National Assessment Governing Board.
GROUPS FOR WHICH DATA WILL BE PRESENTED

- Statewide overall
- Racial/ethnic—white, black, Hispanic, and Asian
- Male-female
- Socio-economic status—poverty (as measured by eligibility for federal school lunch program) and level of parent education
- Districts in the Trial Urban District Assessment (TUDA) compared with achievement in the balance of each state
- English language learners and students with disabilities

ADDITIONAL DATA ELEMENTS

- Demographic profile of each state, showing selected subgroups as percentage of total student population with comparison to base year and 2003.
- Composite of student achievement across grades and subjects, based on percentage of students above Proficient and/or national percentile rankings for average scores.
- International benchmarks when available through linking studies.
- Charter and regular public school schools.
- Comparisons between NAEP and state standards, placed on the NAEP scale.
- Achievement gaps between black, white, Asian, and Hispanic students. Difference between 10th and 90th percentiles statewide.
- Contextual factors of public policy interest, such as class size, per pupil spending, algebra in 8th grade, teacher experience and major or minor in discipline taught, availability of classroom computers and Internet, and frequent school changing (years the student has been in school).
Introduction – Parent Outreach Planning Proposal

At its May 2012 meeting, the Governing Board approved recommendations by the Ad Hoc Committee on NAEP Parent Engagement to increase outreach activities to inform parent leaders and parent groups about NAEP. From this point on, the Reporting and Dissemination Committee will play a key role in reviewing and recommending activities and strategies designed to inform parent leaders about NAEP and how it can be a useful resource for the nations’ parents. The goal is to develop ways the Board and others can use NAEP data and resources to increase awareness among parents about the urgency to improve overall student achievement and reduce achievement gaps by race, ethnicity, and income.

The following document, “Parent Outreach Planning Proposal,” outlines recommendations from Board staff and Reingold, the Board’s communications contractor, on potential outreach activities the Board can pursue. At its November 30 meeting, the Reporting and Dissemination Committee will provide feedback on the parent outreach recommendations and input on additional ideas the Board should consider. This information will be collected to create an overall strategy that will be formally approved by the committee and then the full Board, potentially at the March 2013 Board meeting. Below are several discussion questions suggested by Board staff to facilitate the conversation on parent outreach.

**Discussion Questions**

How can NAEP be relevant to parents, especially as there are no data available at the student or school level?

What messages and information would we most want parent leaders and organizations to hear and know about NAEP and the Governing Board?

What kinds of education-related resources would be useful to parents that NAEP can provide?

Besides organizations like PTA, which groups or communities would benefit from parent-centered outreach on NAEP?

What activities and strategies might best resonate with parents, parent leaders, and parent groups?

Which parent outreach strategies should receive a high priority during the first six months of implementation?
Parent Outreach Planning Proposal

OVERVIEW

The National Assessment Governing Board recognizes that parents have a vital interest in the quality of our students’ education and can be an influential force to effect change. The Board understands that parents have a fundamental personal interest in the education of their children; the challenge therefore is to communicate messages that also instill a concern for increasing the achievement of all children. Parents are therefore a key audience for the valuable information NAEP provides.

To alert parents to the crises we face in K-12 education and to encourage them to use NAEP resources to educate themselves and advocate for improvements in education, the Governing Board formed an Ad Hoc Committee on NAEP Parent Engagement in March 2011. The final report of that Committee was issued March 2, 2012, and presents recommendations for how the Governing Board can meet the overriding goal of effectively communicating to parents the urgent need to improve student achievement and close achievement gaps. These recommendations, approved unanimously by the Board on May 19, 2012, include:

- **Audience:** Specify the target audience: national, state, and local parent leaders and parent organizations.
- **Partnerships and Outreach:** Establish relationships with recognized parent and community-based organizations.
- **Materials:** Develop presentations and materials targeted to parents for use by Governing Board members and others.
- **Website and Online Outreach:** Develop parent pages on the Governing Board and NAEP websites.

What follows is a set of recommended approaches for Board consideration. It is important to note that the Governing Board defines its primary parent audience as parent leaders and influencers at the national and local level. It is not practical to assume that all parents can be reached effectively, nor is it sufficient to communicate only with parents of children participating in NAEP.

The Governing Board and the National Center for Education Statistics (NCES) both seek to inform parent leaders about NAEP and encourage them to use its numerous resources. The organizations should coordinate outreach activities to minimize duplication of effort or potential outreach gaps. This outreach proposal takes into account activities undertaken to date by the Governing Board and NCES to engage parent leaders with NAEP, and suggests additional strategies and tactics, across a variety of communications channels, to assist the Governing Board in effectively implementing the recommendations of the Ad Hoc Committee.

GOALS FOR OUTREACH

The Governing Board’s parent outreach efforts should clearly convey how the Board believes parent leaders can use NAEP. Initiatives should inspire parents to take the following actions:
Learn about NAEP and understand the data and resources available.  
Access and use NAEP tools to better understand achievement trends and drivers.  
Have discussions and ask questions about improving student achievement and narrowing achievement gaps.

OUTREACH EFFORTS TO DATE

As background to our recommendations, it is important to note the outreach activities the Governing Board and NCES have pursued thus far to reach the parent audience.

The Governing Board has extensively tested materials and messages with parent leaders over the last year. These leaders have made it clear there is a demand for NAEP data and resources that are presented in a parent-friendly manner. Feedback suggests there is still work to be done to create materials and website resources to better meet parent leaders’ needs.

Audience. The Governing Board has targeted its outreach efforts to influential parent organizations and parent leaders that are in a position to have an impact on education policy, and can act as conduits to their potentially broad national networks of parents. Parallel efforts by NCES have focused on general parent audiences, as well as specifically on parents whose children have been selected to take NAEP.

Partnerships and Outreach. The Governing Board’s initial efforts have been promising. The Board has successfully established relationships with nationally recognized parent-focused organizations including the National PTA and Public Education Network (PEN).

Both organizations helped to recruit parent leaders for meetings in February 2012 to review Governing Board materials, and both have begun to share NAEP information with members through channels including newsletters and social media. National PTA invited Ad Hoc Committee Chair Tonya Miles to give presentations at its legislative conference in March 2012 and annual conference in June 2012.

The Board also received enthusiastic participation and substantive input from Washington, D.C.-area parent leaders from organizations including Learning First Alliance, The Parent Institute, Parents Across America, Parent Educational Advocacy Training Center, and Parent Advocacy Coalition for Educational Rights at outreach events in August of 2011 and 2012.

Last, the Board is planning an education summit for parents in early 2013 in Washington, D.C., and available across the nation via live-streaming Internet video, with the potential for live TV and radio coverage. The objective of the summit is to convey the urgency of improving student achievement in the United States for all children and the urgency of reducing achievement gaps among student subgroups. The audience of 150 to 300 would consist primarily of parent and community leaders, parent organizations, and leaders in education, business, civil rights, the religious community, and legislative policy.

Materials. The Governing Board has developed working outreach materials including a PowerPoint presentation and state profile one-pagers, which were presented at workshops held in February 2012 for the national organizations named above. The materials were well received; participants indicated that a demand exists for parent-focused materials and requested access to
additional NAEP data and resources. The materials were subsequently refined, in response to participant feedback.

NCES has developed a brochure for the general parent audience, “What Every Parent Should Know About NAEP,” which debuted at the National PTA annual conference in June 2012. It is available in both English- and Spanish-language versions and can be downloaded from NCES’ website, and will be available at the meeting. Details on major materials are as follows:

**Materials in Process:**

- **Parent PowerPoint.** The draft PowerPoint was developed to use at Governing Board events, conferences, and other stakeholder activities. It is also suitable for use by parent leader groups with their constituencies. The presentation features the core messages for parent leaders (discussed in detail below), illustrating how NAEP materials can help parent leaders ask the right questions. It comes with a tested script and includes basic information on NAEP online tools, frameworks, achievement levels, the assessment landscape, noteworthy economic indicators, and report card results. This presentation and variations have been made with parent leaders and received very positive feedback. Available here: “Improving Achievement and Closing Gaps: What Parent Leaders Can Do.” Reingold has submitted to the Board options for ways that parent leaders can customize the presentation with their local or state data.

- **State and District Profiles.** Reingold worked with the Governing Board to develop a snapshot template for possible print use and download from the Board website. These are intended to be parent-friendly versions of the NAEP state and district profiles, with a focus on *Proficient*-level achievement information and key background variable findings. They also include brief explanations of what the data show, including trend lines. This document was shared with parent leaders and feedback received.

- **Mapping State Standards One-Pager.** This piece was developed to call attention to the differences among state standards, the role NAEP plays in facilitating state standards comparisons, and the discrepancies between NAEP’s *Proficient* achievement level and the states’ various levels for “proficiency.” The piece also includes interpretations for reading the graphs. This document has not yet been tested with the parent leader audience.

**Website and Online Outreach.** In July 2012, the Governing Board launched its redesigned website, including a section targeted specifically to the parent audience with information on NAEP and the Governing Board, news for parents, and links to various tools parents can use.

At its parent outreach event in August, the Board presented the new pages to parent leaders and facilitated a discussion to capture their feedback. Attendees expressed a high level of interest in the Board’s parent activities, and gave numerous suggestions for ways to make the site more user-friendly for parents. Key feedback included:

- The pages are dense and potentially intimidating to parents. There is a lot of text with few visuals and no clear hierarchy of content.
The pages should convey more clearly what actions a parent visiting the site should take and give a better understanding of how they can use NAEP resources. This call to action could include specific questions parents can ask of education leaders or examples of how parents have used NAEP data.

The pages should be more direct in conveying NAEP’s relevance to parents. They should emphasize how NAEP sheds light on education at the local level, how subjects on which students are tested are relevant to life skills, and how NAEP is distinct from state and local assessments.

The content should be available in other languages, especially Spanish.

Parent leaders would be eager to use NAEP resources available from the site—for example, materials that parents can customize with their local information.

Reingold has also submitted to the Governing Board an audit of the new website, which includes additional recommendations for ways to revise the parent pages.

NCES has also developed pages for parent audiences, available at nationsreportcard.gov/parents and nces.ed.gov/nationsreportcard/parents/, which focus on, respectively, general parent audiences and parents of students taking NAEP.

**Parent Messages**

Prior to developing further materials and conducting outreach, the Governing Board should confirm its core messages for parent leaders. Messages should reflect NAEP data to make the case that there is an urgent need to improve student achievement and close achievement gaps. Once parent leaders understand that urgent need, the messages must include suggestions for how they can use NAEP resources as a foundation to ask questions and discuss ways to have productive conversations with state and local educators and policymakers that are focused on improving student achievement and closing achievement gaps.

Featured for consideration below are proposed primary messages (in bold) and secondary messages (bulleted) for the Board to consider. These messages would be used across various outreach materials and activities.

**Our nation faces a crisis in K-12 education.**

- There is an urgent need to improve our students’ achievement and close persisting gaps between underserved students and their peers.
- In the core subjects of reading and mathematics, less than 40 percent of 4th, 8th, and 12th graders meet the *Proficient* level on NAEP.
- Today’s students are the drivers of tomorrow’s economy. Low student achievement will affect the competitiveness of our future workforce and our nation’s prosperity.
- U.S. student achievement has fallen significantly behind that of students in the highest-performing countries.
- Demographics are changing. Our nation cannot afford for minority groups to achieve at levels below their peers.
Our nation is dedicated to the principle of equality. It is morally unacceptable to allow wide gaps on the basis of race or income level to persist between groups of students.

NAEP is a truth teller, shining a light on how our nation’s students are performing.

- As the only nationally representative measure of student achievement, NAEP provides a unique diagnosis of the condition and progress of education at the national, state, and urban district levels.
- NAEP doesn’t report on individual students, but rather provides the big picture.
- Unlike state tests, which vary in their standards from state to state, NAEP is a common measure across all states. It also allows a comparison of student achievement across urban districts and student subgroups.
- The Governing Board identifies the Proficient level of achievement as the benchmark. Proficient designates “mastery over challenging subject matter,” and helps us understand what is “good enough” in terms of student achievement.
- NAEP collects background information from students, teachers, and administrators that helps identify both in-school and at-home factors related to achievement.

Parent leaders can use NAEP as a tool to engage other parents in helping improve student achievement and close achievement gaps.

- Parents are the primary advocates for their children’s education, and they can also serve as an influential force advocating for improved education for all students.
- Parent leaders can use NAEP data and tools to identify where student achievement stands, as well as areas that deserve further attention to improve achievement.
- NAEP resources can help parent leaders have productive conversations with peers and state and local educators and policymakers to promote student achievement and close gaps.
PROPOSED OUTREACH RECOMMENDATIONS

Based on the established goals, messages, and existing activities described above, Reingold proposes the following outreach strategies, presented in the context of the Ad Hoc Committee’s recommendations. We have commented on the level of effort required and anticipated impact to assist the Governing Board in setting priorities in its outreach approach.

Audience

In its report, the Ad Hoc Committee has identified the target audience for parent outreach as follows: Groups of active parents and parent organizations who see the connection between system performance and the potential for impact on individual students. These include local and state leaders, often members of recognized parent and community organizations, who regularly work with the leaders of education systems, examine data, and ask fundamental questions to support and foster improved achievement and the closing of achievement gaps.

RECOMMENDATION: Having defined its audience, the Governing Board should:

- Review its stakeholder database to ensure that all relevant groups have been captured.
- Develop a relationship map that connects Board members, alumni, and other NAEP champions with the target parent leader audience.
- Develop a list of 50 key parent leaders on which to focus initial outreach efforts. These contacts can help to reach as many parents as possible by using their networks to efficiently disseminate the Board’s messages.

Partnerships and Outreach

RECOMMENDATION: The Governing Board should continue to identify key parent influencers and approach them about opportunities to partner, present, or share in activities, including co-hosting in-person or online events.

- **Target education journalists or publications and pitch parent-focused articles or newsletters.** The Board’s media database includes parent-focused journalists who receive alerts regarding report card releases, but there has not yet been a coordinated effort to reach these journalists with targeted messages. Reingold suggests the following:
  - Further refining a parent-specific media database.
  - Developing specific messages to send to contacts between releases, and encouraging them to write articles for the parent audience.
  - Providing content to education and parent reporters or bloggers to post on their social media pages or in their newsletters.

- **Co-sponsor panels, forums, or workshops with local and national parent-focused groups.** Such events could be held after a report card release to break down the results specifically for parent leaders. The Board could also hold an online national conversation on assessments—a town hall meeting, or “NAEP Day”—during which experts can provide comments and be available to answer assessment-related questions. Further, the
Board could partner with civil rights groups that have significant parent involvement for sessions on using NAEP tools to gain insights on achievement gaps. Many groups, such as the Alliance for Excellent Education, put on webinars with education leaders. The Board has already participated in several of these, and could seek additional opportunities to participate in similar webinars with a parent group leader like National PTA to discuss the importance of parent involvement in education, or what the latest NAEP results mean for parents.

- **Partner with prominent organizations to develop parent-focused op-eds.** Board members or parent groups that use NAEP could write op-eds on a timely education topic that highlights NAEP results in the context of various local and regional education issues.

- **Distribute materials to community groups and schools.** Community facilities including parks, recreation and community centers, churches, and libraries are potential places for parents to find NAEP information. The Governing Board can work with NAEP state coordinators and parent groups to distribute materials at these points. There may also be opportunities at schools, such as during back-to-school events, book fairs, and parent-teacher conferences, where NAEP materials for parents can be shared. Additionally, administrators and school counselors could be invited to workshops or webinars on how to use NAEP resources to inform and empower parents.

- **Speak at education-related conferences.** Most large parent, education, and civil rights organizations have annual conferences. Representatives of the Board who are practiced in speaking to parents could present on the ways parent leaders can use NAEP. At a larger education conference such as Education Nation, the Governing Board could team up with a national, state, or local recognized parent leader and co-present on a topic, such as using data for improving achievement.

- **Work with NAEP state coordinators to connect with parent leaders.** NAEP state coordinators provide invaluable opportunities to reach parent organizations at the state level. The Board should collaborate with NCES on ways to consider having state coordinators assist in raising awareness of NAEP among parent leaders, establish relationships, connect parent leaders with NAEP resources, and explain how parent leaders in their states use NAEP.

### Materials

**RECOMMENDATION:** Existing and new parent materials will be integral to the outreach effort. Relevant materials should be customized for particular states and urban districts. Many of the materials listed below can serve as important assets for outreach activities listed in this proposal.

**Proposed New Materials:**

- **High Standards One-Pager.** Reingold has created a concept version of this piece, which includes a summary of the overall student achievement crisis and the need to narrow
achievement gaps in the United States. The one-pager also includes a call to action for parent leaders, with priority tools and information they can use to get involved. The document would feature:

- An overview of the urgency of addressing low achievement and achievement gaps
- Definitions of the achievement levels, communicating NAEP’s *Proficient* level as the benchmark to aim for
- Sample questions mapped to achievement levels as practical examples through which parent leaders can better understand what *Proficient* means
- Information on economic and social implications of low achievement
- Callouts such as parent testimonials, questions parents can ask, and compelling background variable statistics

### Parent Leader Testimonials

The Governing Board should gather testimonials from parent leaders who have used NAEP as a resource and motivator for change. This could be a video to show on the Web or at conferences, or a PDF file for print distribution. The tools would include:

- Examples of how parent leaders can use the frameworks, report card data, NAEP tools, and other resources
- Quotes that convey “I use NAEP to…”
- Parent leader-created checklists with steps others can take to get parents involved in improving education for all children

### Background Variable One-Pager and Quiz

This tool would include a one-pager with information on collected background variable data and how parent leaders can use these data. An interactive quiz for parent leaders—its goal to educate parents in ways to improve student achievement—could test their knowledge of ways to improve academic performance, focused on compelling findings from a cross section of report cards. The tool would include:

- Information on the collection of the background data, the questionnaires, and the types of information that are made available through them
- A results page or section with report card findings, for example, the 2011 NAEP Reading finding about students who read outside of school at least once a week

In addition, we strongly recommend the renaming of “background variables.” While a suitable term for educators, we don’t believe it will connect with parent leaders. We offer “achievement drivers” as a starting point for consideration.

### Parent Leader Discussion Guide

Parent leaders could use this tool when speaking to a teacher, administrator, or policymaker to learn how their school, school system, or state compares with others nationwide, and to learn what is being done to increase academic rigor and achievement for all students. It should include general talking points to shape the conversation in a way that fosters collaboration, yields valuable information, and identifies next steps.
• **Promotional Materials.** A variety of materials could be produced and distributed at conferences and other events to create awareness of NAEP and provide potential partners and advocates with takeaways to better remember key messages. These include USB drives, pencils, bumper stickers, erasers, or posters; all could contain a one-line statement for parent leaders on the value of using NAEP, with links to the appropriate websites.

### Website and Online Outreach

**RECOMMENDATION:** As described above, parent leaders at the Board’s August outreach event provided a variety of thoughts and recommendations for improving the parent Web pages. Reingold also performed an audit of the website, looking specifically at the parent pages from that audience’s perspective, and provided further recommendations on revising the pages’ structure, design, and prioritization of content to better reach and engage parents. The Board should also pursue supplementary online and social media outreach efforts to reinforce the website and reach parents through the channels they use online.

• **Refine the parent Web pages per the feedback from parent leaders and the recommendations from Reingold’s website report.** The parent landing page should be the primary portal for parents and parent leaders seeking information and resources. The Governing Board can redesign or restructure the website to visually prioritize the information it wants parent leaders to access, and to eliminate redundant or unrelated content. The Governing Board should also expand its suite of materials available to parents, as outlined above in the “Materials” section, and provide user-friendly access and customization from the website. All recommendations—including both content and design—are included in the outreach meeting notes and website services findings report.

• **Develop a quarterly newsletter for parent leaders.** A newsletter would help the Board stay top of mind with parent leaders by informing them of Board news and events, such as report card release data, updates on upcoming assessments, and highlights of other Board initiatives. Although some periods may have more news content than others, newsletters can be brief and direct, and encourage parent leaders to visit the website for more information.

• **Perform search engine optimization (SEO) to capitalize on search terms parent leaders use.** Reingold can help determine priority keywords the Board can use to optimize its parent pages for search engines. By creating or refining website content based on language that research shows parents use, the Governing Board can use SEO to help raise the website’s ranking in search engine results, increasing the chances parents will find and use the website content and resources.

• **Share social media content with targeted parent groups.** Develop a list of 25 priority parent leader groups and provide them a monthly calendar with the latest news from the Governing Board, compelling NAEP stats from recent releases, and other updates, for their use when developing their editorial calendars.
Seed topics on discussion portals where parent leaders share ways for parents to get involved in education. The National Coalition for Parent Involvement in Education is a clearinghouse with links on its website to numerous parent organizations, including the Governing Board and NCES. The Board could co-sponsor a parent portal on such a site and work with the forum or site managers to promote topics, questions, or conversations on some of the many other sites where parent leaders share information.

Develop and disseminate data infographics. Infographics are being used more frequently to display statistics and data, and would provide a great opportunity to communicate report card findings visually. The Governing Board can work with NCES to package report card results into compelling infographics geared toward parent leaders. Examples of what this might look like include The Nation’s Report Card: Writing 2011 infographic and “Six Years and a Thousand Students.”

Create an assessment resource directory for parent leaders on the Governing Board website. As the Governing Board serves as a thought leader on assessments, it is natural that the Board would offer a list of resources for parent leaders seeking information on NAEP, student education data, and other educational sources.

Develop a blog on the Governing Board website. Board members, report card release panelists, and others can provide perspectives that relate NAEP to various topics of interest to parent leaders. Hosting a blog on the Governing Board’s site would keep the site timely and drive website traffic if promoted through outlets such as media and social media. This could also be included in a newsletter.

Develop an interactive NAEP data map. Maps are visually attractive and approachable, and can serve as a valuable entry point to complex data. A NAEP map would be hosted on the Governing Board website and allow users to click on a given state to access NAEP highlights by subject and grade level, featuring the most compelling information from the state profiles. The map could also include district data, and would allow parent leaders to click through to bring them deeper into NAEP tools. Examples of similar concepts are here: dataqualitycampaign.org/; washingtonpost.com/wp-srv/special/nation/census/2010/; achieve.org/states; completecollege.org/state_data/.
Whereas, the National Assessment Governing Board is implementing an initiative to make a difference in fostering the improvement of student achievement in the United States and of closing achievement gaps by race, ethnicity, and income levels using NAEP data and resources; and

Whereas, the National Assessment Governing Board established the Ad Hoc Committee on NAEP Parent Engagement in March 2011 to

“present recommendations…the Governing Board and representatives of the NAEP program can take directly, and/or support the efforts of others to increase parent awareness about the urgency to improve the levels of student achievement in the U.S. and the urgency to reduce the size of achievement gaps by race, ethnicity, and income levels, using NAEP data and resources”; and

Whereas, the Ad Hoc Committee on NAEP Parent Engagement presented its recommendations to the National Assessment Governing Board on March 2, 2012; and

Whereas, the Ad Hoc Committee on NAEP Parent Engagement recommended that the National Assessment Governing Board

- Specify National, State, and Local Parent Leaders and Parent Organizations as the Target Audience
- Establish Relationships with Recognized Parent and Community-based Organizations
- Develop Presentations and Materials Targeted to Parents for Use by Governing Board Members and Others
- Develop Parent Pages on the Governing Board and NAEP Websites
- Conduct a Parent Education Summit in Late Summer/Early Fall 2012; and

Whereas, adoption of the Ad Hoc Committee recommendations will be valuable, feasible, and consistent with the Governing Board’s authority to "develop guidelines for reporting and disseminating results” and “…improve the form, content, use, and reporting of [NAEP] results…”; and

Whereas, implementation of the Ad Hoc Committee recommendations will require staff and financial resources and oversight by one or more standing committees of the National Assessment Governing Board;

Therefore, the National Assessment Governing Board hereby

1. adopts the recommendations of the Ad Hoc Committee on NAEP Parent Engagement presented on March 2, 2012;

2. approves the use of appropriate staff and financial resources to implement the recommendations; and

3. authorizes the assignment of oversight of these activities to Governing Board standing committees.
Ad Hoc Committee on NAEP Parent Engagement

Reaching Parents with NAEP Resources

March 2, 2012

Committee Members
Tonya Miles, Chair
Louis M. Fabrizio
Shannon Garrison
Doris R. Hicks
Hector Ibarra
Henry Kranendonk
Warren T. Smith
Blair Taylor

Staff
Ray Fields
Ad Hoc Committee on NAEP Parent Engagement

Overview of Recommendations

1. Specify the Target Audience: National, State, and Local Parent Leaders and Parent Organizations
2. Establish Relationships with Recognized Parent and Community-based Organizations
3. Develop Presentations and Materials Targeted to Parents for Use by Governing Board Members and Others
4. Develop Parent Pages on the Governing Board and NAEP Websites
5. Conduct a Parent Education Summit in Late Summer/Early Fall 2012

Committee Activity Timeline

November 2010 Recognize Need to Address NAEP Parent Engagement
March 2011 Approve Mission Statement and Establish Ad Hoc Committee on NAEP Parent Engagement
April 2011 First Ad Hoc Committee Teleconference
May 2011 First Committee Meeting
August 2011 Second Committee Meeting
October 2011 Second Teleconference
December 2011 Third Committee Meeting
February 2012 Third Teleconference
March 2012 Final Committee Meeting; Present Recommendations to the Board
Foreword

The National Assessment Governing Board, in overseeing the National Assessment of Educational Progress (NAEP or the Nation’s Report Card), is carrying out an initiative to raise public awareness about the status of student achievement in the United States.

The Governing Board believes that the low levels of student achievement and the persistent, large achievement gaps between student demographic subgroups are cause for alarm—for individuals, for families, for communities, and for the nation’s future.

Although the release of NAEP reports brings periodic public attention to this problem, this attention is not sustained for very long.

Consequently, the Governing Board is implementing an initiative to convey the urgency of improving achievement for all students and of closing achievement gaps between student subgroups by race, ethnicity and income levels, using NAEP data and resources.¹

One part of this initiative is aimed at reaching parents. In March 2011, the Governing Board established the Ad Hoc Committee on NAEP Parent Engagement, composed of Board members. The Ad Hoc Committee’s assignment was to study ways to reach parents with NAEP data and resources and to present the Committee’s recommendations to the Governing Board by March 2012.

The members of the Ad Hoc Committee have worked diligently over the past year and are pleased to present our report and recommendations on the following pages.

We would like to express appreciation for the important contributions of the National Center for Education Statistics in supporting the Ad Hoc Committee’s work and in embracing the objective of reaching more parents with NAEP data and resources. We also thank the Governing Board’s CCSSO² Policy Task Force members for their valuable comments and suggestions.

Tonya Miles
Chair
Ad Hoc Committee on NAEP Parent Engagement

¹ The authority for this initiative is found under the Governing Board’s duties in the NAEP legislation, Public Law 107-279. Specifically, Section 302(c)(1) authorizes the Board to “take appropriate actions needed to improve the form, content, use, and reporting of results” and “plan and execute the initial public release of National Assessment of Educational Progress reports.”

² The acronym CCSSO stands for Council of Chief State School Officers.
Introduction
The National Assessment Governing Board, recognizing that NAEP report releases were not conveying a sense of urgency, began an initiative in May 2010 to see what the Board could do to “make a difference” in fostering concern and action about the need to improve achievement and reduce achievement gaps, using NAEP data and resources. Toward this goal, the Governing Board established the Ad Hoc Committee on NAEP Parent Engagement. The Committee’s task was to develop recommendations on ways to reach parents with NAEP information. The purpose of this report is to document the work of the Ad Hoc Committee and present its recommendations.

Background
U.S. Secretary of Education Arne Duncan addressed the Governing Board on November 19, 2010. He focused on the urgent need to improve student achievement and reduce achievement gaps among student subgroups. He has said publicly that “our nation will pay the price socially and economically” if we fail to act with determination and dispatch and stressed to the Board that “we have to continue to awaken our country to the huge consequences” of inaction.

Secretary Duncan emphasized the important role of parents in improving student achievement. He told the story of President Obama meeting with the President of South Korea, Lee Myung-bak. President Obama asked him about education issues in South Korea. President Lee said his biggest challenge is that parents in South Korea are very assertive in demanding a good education from their schools and great effort from their children. He emphasized that this includes parents of all income levels.

Implicit in this story is the fact that South Korean students, as well as others in the world, outperform U.S. students in mathematics and science on TIMSS. Today’s students are tomorrow’s workers and leaders. It follows that failing to improve U.S. student achievement could have disastrous effects on the nation’s future work force and global competitiveness, and that parents have an important role to play in promoting improved student achievement.

Secretary Duncan continued by saying “I wish my biggest problem, my biggest challenge, was parents knocking down my door saying, ‘Get better faster!’” He said that there are good examples in the U.S. of parent initiatives that impact student achievement. But Secretary Duncan wanted to “scale up” parent engagement programs that “are really showing the ability to drive student achievement.”

---

3 The acronym TIMSS stands for the Trends in International Mathematics and Science Study.
The Secretary’s remarks and the Board’s initiative to make a difference served as the backdrop to Board member Tonya Miles asking what can the Board do to make NAEP data available to parents and guardians about student achievement, especially about the urgency of addressing achievement gaps by race, ethnicity, and income levels.

The question—“What can the Board do?”—is pertinent and important. Parents have a significant stake in the quality of their local schools and, most immediately, in their own children’s achievement.

Governing Board Chair David Driscoll recognized the opportunity and value of reaching parents with NAEP data. Therefore, at the conclusion of the November 2010 Governing Board meeting, he asked Ms. Miles, and she agreed, to lead a Board initiative to increase parent awareness about and access to NAEP data. The goal was to bring attention to the unacceptably low levels of student achievement in the U.S. and the disgraceful size of the achievement gaps.

The Mission
At the March 2011 Board meeting, the Executive Committee approved the mission statement for and established the Ad Hoc Committee on NAEP Parent Engagement (Appendix A). The Committee would be composed of Board members and chaired by Ms. Miles. The Ad Hoc Committee’s task was to present recommendations to the Governing Board by March 2012. The recommendations would describe steps and strategies the Governing Board and representatives of the NAEP program can take directly, and/or support the efforts of others to increase parent awareness about the urgency to improve the levels of student achievement in the U.S. and the urgency to reduce the size of achievement gaps by race, ethnicity, and income levels, using NAEP data and resources.

The mission statement indicated that the recommendations were to be clear about the limits on the Board’s role under the law. This was to ensure the Committee considered all appropriate options without exceeding the Board’s authority.

The recommendations were to help reach parents in feasible, innovative, and meaningful ways, across all income levels, and whether residing in urban, rural, or suburban areas. Finally, the recommendations were to include strategies to make NAEP parent engagement an ongoing part of the work of the Board and the NAEP program.

---

4 The term “parents” as used throughout this report is intended to refer to parents and guardians of school children.
Committee Activities
Members of the Ad Hoc Committee have met four times during the May 2011, August 2011, December 2011, and March 2012 Board meetings. The agendas for these meetings are in Appendix B.

The Ad Hoc Committee also has met three times in between Board meetings via conference calls:
- April 15, 2011 - reviewed the Committee’s mission statement and a timeline for completing their work (Appendix C)
- October 12, 2011 - focused on formulating the Committee’s preliminary recommendations for discussion at the December 2011 Board meeting (Appendix D)
- February 8, 2012 - reviewed the Committee’s initial draft report

In addition, Chair Tonya Miles and Ray Fields conducted meetings with leaders of three nationally recognized parent-related organizations. The purpose was to brief them on the Board’s initiative to reach parents with NAEP data, to receive their input and feedback, and to determine their interest in supporting this initiative. The three organizations are the National PTA, the Public Education Network, and the Center on School, Family, and Community Partnerships at Johns Hopkins University.

In connection with the August 2011 meeting in Washington, D.C., the Board conducted an outreach event with parent leaders and national and local parent organizations. The discussion with meeting participants, led by Ms. Miles and Governing Board Chair Driscoll, resulted in valuable feedback and input on the Board’s parent initiative. A summary of the discussion at this parent outreach meeting is in Appendix E.

Concluding Comment
Parents are the primary advocates for the quality of their children’s education. Having solid information about education achievement improves their ability to advocate and ask the right questions. NAEP can be one potentially valuable source of such information. Therefore, it is appropriate to seek ways to reach parents with NAEP data and resources. Some progress already has been made, in a small way, as will be seen in the activities and relationships described below. The recommendations that follow are offered as a set of feasible next steps, all within the Governing Board’s authority. All have the potential to reach parents in meaningful ways. Recommendations that the Governing Board decides to adopt should be assigned to appropriate Board committees and staff for implementation.
Recommendations

1. Specify the Target Audience: National, State, and Local Parent Leaders and Parent Organizations

The target audience needs to be defined. Approximately 55 million students are enrolled in public and private K-12 schools in the U.S. It is not feasible to reach the parents of all these children with NAEP data, nor is it within NAEP or the Governing Board’s scope to do so.

Further, the achievement of their own children is the most pressing and immediate interest of parents. Because NAEP does not provide individual student results, this interest of parents is not served by NAEP.

A unique aspect of NAEP is its ability to report patterns of overall and subgroup student performance within and across education systems. These patterns may reflect education system strengths and weaknesses that can affect the achievement of individual students. The NAEP data for the states and 21 urban districts provide ample evidence of differences in achievement across comparable groups at points in time and differences in gains in achievement over time. The NAEP data also document persistent and unacceptable achievement gaps between groups. This NAEP information does have potential interest for parents.

Also of potential interest to parents is how their education systems compare internationally. The linking studies the Board has endorsed, beginning in 2011, between NAEP and the international assessments (TIMSS and PIRLS) will provide a way to compare student achievement at the state level in the U.S. with achievement in other nations.

State and local education policymakers use NAEP data to ask fundamental questions about the levels of student achievement in schools under their authority. For example, Tennessee Commissioner of Education Kevin Huffman discussed how he uses NAEP at a November 2011 meeting in Nashville on NAEP 12th grade academic preparedness. Commissioner Huffman said that he analyzes student subgroup results in his state (e.g., students on free and reduced lunch) in comparison to other states. Raising questions about how subgroup performance compares across jurisdictions can help highlight where state or local policies may or may not be working.

Asking thoughtful questions about the implications of NAEP results can be a positive way for parents to begin a productive conversation with state education leaders seeking to improve

5 The 21 participants in the NAEP Trial Urban District Assessment Program are: Albuquerque, Atlanta, Austin, Baltimore City, Boston, Charlotte, Chicago, Cleveland, Dallas, Detroit, Fresno, Hillsborough County, Houston, Los Angeles, Louisville, Ky. (Jefferson County), Miami (Dade County), Milwaukee, New York City, Philadelphia, San Diego, and Washington, DC.

6 The acronym PIRLS stands for the Progress in International Reading Literacy Study.
achievement and close achievement gaps. Of course, while NAEP can be used as a source of information to help parents identify important questions to ask about the status of student achievement locally, the answers about what to do must be made by state and local officials with authority for the schools.

The Ad Hoc Committee believes there are groups of active parents and parent organizations who see the connection between system performance and the potential for impact on individual students. These include local and state leaders, often members of recognized parent and community organizations, who regularly work with the leaders of education systems, examine data, and ask fundamental questions to support and foster improved achievement and the closing of achievement gaps. These parent leaders and parent organizations should be the initial target audience for NAEP data and resources.

More specifically, because NAEP provides data for each of the 50 states and 21 urban districts, the initial target audience should be state and local parent leaders and parent organizations associated with these jurisdictions.

### 2. Establish Relationships with Recognized Parent and Community-based Organizations

To reach the target audience with NAEP data, it is important to work collaboratively with existing parent and community-based organizations. Many of these organizations have state affiliates and/or affiliates associated with local school districts. These organizations have direct access to parent and community leaders through email networks, social media, newsletters, and websites. These mechanisms are potentially effective, viable avenues for the dissemination of NAEP data and resources. In addition, these organizations often conduct national and state conferences, which could afford opportunities for presenting NAEP data and resources.

The Ad Hoc Committee has initiated conversations with the National PTA (NPTA), with positive results (see Appendix F). For example, the NPTA has begun announcing NAEP release events through its email networks and social media. In addition, Tonya Miles has been invited to make a presentation on March 7, 2012 at the NPTA Legislative Conference and on June 21, 2012 at the NPTA Annual Conference. Further, the NPTA assisted in recruiting parents for a meeting on February 16, 2012 to help review the NAEP presentation and materials for parents described in Recommendation 3.

Likewise, collaborative activity has occurred with the Public Education Network (PEN). Cornelia Orr, Governing Board Executive Director, made a presentation on NAEP and 12th grade academic preparedness at the PEN annual conference in November 2011. PEN also helped recruit experts from among its member organizations for a one-day meeting held on February 14,
2012 to provide input and feedback on the NAEP presentation and materials for parents
described in Recommendation 3. PEN already transmits information about NAEP data and
NAEP releases to its members and newsletter subscribers.

The Governing Board should continue to develop the relationships with the NPTA and PEN, and
develop similar collaborative relationships with other organizations.

3. Develop Presentations and Materials Targeted to Parents for Use by Board Members
   and Others

Recognizing that the scope and depth of NAEP data and resources can be overwhelming, the
Governing Board is working to develop a model PowerPoint presentation and associated
materials for parents. Consistent with the information needs of the target audience in
Recommendation 1, the presentation and materials can be customized for particular states and
urban districts. The materials will include easy-to-understand charts and graphs and avoid the
use of technical terms and jargon. In addition to explaining what NAEP is, the presentation will
highlight NAEP data regarding the levels of achievement and the gaps between subgroups in
ways that convey urgency.

The presentation and materials should be designed to help the audience understand the role of
NAEP in the context of state and local assessments. Sample test items can be used to illustrate
what content NAEP measures and how it is measured; consideration can be given to how this
information about NAEP may complement state assessments. As noted in Recommendation 2,
conducting input and feedback meetings with parent leaders and representatives of parent
organizations is important to ensure that the level of detail and amount of information is
appropriately tailored for the target audience.

The intent is for these resources to be available for use by Governing Board members invited to
make presentations to the public and by interested parent and community-based organizations in
making presentations specific to their locale.

The National Center for Education Statistics (NCES) is currently developing a general
publication for parents. This publication will inform parents about what NAEP is, how it fits into
the education landscape, and options to learn or do more. This publication will be debuted at the
NPTA conference in June and displayed at the NAEP booth at the conference.
4. Develop Parent Pages on the Governing Board and NAEP Websites

Currently, the Governing Board website has no pages aimed at parents as the target audience. The NAEP website, managed by the National Center for Education Statistics (NCES) does have pages for parents whose child has been selected to take NAEP, but not for parents in general.

The Ad Hoc Committee invited NCES to examine what it can do to make NAEP information on the website more accessible to parents. As a first important step, NCES made the “parent” navigation button more prominent on the NAEP website landing page. NCES is exploring additional changes to make the NAEP data more accessible to parents. As they develop the parent publication mentioned in Recommendation 3, NCES will update the NAEP web pages to ensure consistency. This will help expand the NAEP website audience from just parents of students selected for the NAEP sample to all interested parents.

The Ad Hoc Committee asked the Board’s communications and website contractors, Reingold, Inc. and Quotient, to develop page mockups for parent pages on the Governing Board website (Appendix G). These should be further developed and incorporated as components of the Governing Board’s website redesign, which is currently underway. The model PowerPoint presentation and materials in Recommendation 3 should be available for easy downloading from the Governing Board website.

In addition, the Governing Board should seek ways to leverage mass communications (e.g., TV, radio, public service announcements, and social media) to reach parents with NAEP data and resources.

5. Conduct a Parent Education Summit in Late Summer/Early Fall 2012

The Ad Hoc Committee proposes a one-day parent summit on education for the late summer or early fall of 2012. The summit would be conducted in Washington, D.C. and available across the nation via live-streaming internet video, with the potential for live TV and radio coverage via C-SPAN.

The objective of the summit would be to convey the urgency of improving student achievement in the United States for all children and the urgency of reducing achievement gaps between student subgroups.

In addition to Governing Board members, the audience of 150-300 would consist primarily of parent and community leaders, parent organizations, and leaders in education, business, civil rights, the religious community, and legislative policy.
To help convey the non-partisan, universal interest in achieving the summit objective, as well as to focus on its importance for the nation’s future, First Lady Michelle Obama and former First Lady Laura Bush would be invited to share the podium in delivering the keynote address.

A distinguished journalist or media representative, acknowledged for intellect and freedom from bias, would be invited to moderate and provide a concluding summary.

A respected education advocate, with a strong reputation for compelling presentations on student achievement would be invited to present the NAEP data as evidence of the need to address the summit objective.

Individual and panel presentations would be made to address the national imperative for achieving the summit objective, from a wide range of perspectives which, taken together, would provide a compelling, unassailable argument for the urgent need to take action.

For example (not listed in priority order):

- Religious leaders would provide the moral perspective
- Economists would provide the national economic perspective
- Civil rights leaders would provide the equity perspective
- Military leaders would address the national security imperative
- Business leaders would address the human capital and employment imperative
- Scholars from nationally recognized policy institutions and foundations, representing a diverse range of philosophical orientations, would provide societal perspectives
- Demographers would address the implications from the perspective of a changing population
- Parent leaders would address the imperative for families and students
- Educators would describe actions that are needed to improve academic achievement overall and close achievement gaps
Plans for NAEP in Puerto Rico in 2013

In 2011, NAEP conducted the mathematics KaSA (Knowledge and Skills Appropriate) study at grades 4 and 8. The purpose of the study was to increase the measurement precision in the estimation of student ability at the lower end of the NAEP scale, while still administering an assessment that is consistent with the NAEP mathematics framework.

The KaSA study included the administration of the 2011 operational assessment augmented by four KaSA blocks that were comprised of items targeted to the ability of lower performing students. These KaSA blocks were administered to students in both Puerto Rico and a relatively small national sample in the mainland U.S., in one of two book configurations: (1) books with two KaSA blocks, or (2) books with one KaSA block and one operational block. In addition, some students in the Puerto Rico sample were assessed using books with two operational blocks. This study design enabled comparisons of performance on the different kinds of items and blocks (KaSA or operational). The KaSA items and blocks functioned well and were, on average, more accessible than operational items.

In order to determine if the results from 2011 could be replicated in future years, the 2011 study will be repeated in 2013. As such, the same KaSA blocks from 2011 will be administered to students in both Puerto Rico and a relatively small national sample in the mainland U.S., using the same study design. We will determine whether the encouraging results of 2011 can be replicated in 2013, which will establish the level of confidence and validity desired to report trend results on the NAEP scale for Puerto Rico and use the KaSA items for operational purposes in the mainland U.S.
May 6, 2010

The Honorable Luis Fortuño
Governor of Puerto Rico
La Fortaleza
P.O. Box 9020082
San Juan, PR 00902-0082

Dear Governor Fortuño:

I am writing to keep you informed of the extensive research conducted and plans that are underway in regard to the testing of students in Puerto Rico on the National Assessment of Educational Progress (NAEP).

Public schools in Puerto Rico were included in the National Assessment for the first time under the No Child Left Behind law enacted in 2002. Because NAEP Reading is an assessment of reading in English the Commonwealth of Puerto Rico was granted a waiver from this exam. However, the mathematics assessment was translated into Spanish, the language of instruction in Puerto Rico, and has been administered three times to representative samples of students at grades 4 and 8 in 2003, 2005, and 2007.

After 2003, because of the low percent of questions answered correctly, NAEP introduced several changes in procedure in an effort to improve the reliability of Puerto Rico results. However, it became apparent after the third administration that the problems had not been resolved. Students in Puerto Rico continued to score at the low end of the NAEP scale where there are too few items to report change reliably. While the problem is most severe in Puerto Rico, similar issues have emerged for some jurisdictions on the mainland (e.g., some urban school districts).

To address these issues, the National Center for Education Statistics (NCES) convened an expert panel and conducted a major program of research in 2009. The goal of the studies was to evaluate the appropriateness of NAEP math items and their translation into Spanish and to understand how mathematics is taught and learned in Puerto Rico. On February 22, 2010 NCES shared the findings as well as planned future assessment activities with the Secretary of Education of Puerto Rico, who was very supportive of these efforts.
In large part based on the studies, NCES has developed a new set of NAEP questions, designed to improve measurement precision for lower-performing students both in Puerto Rico and on the U.S. mainland. In 2011 there will be a tryout of these questions. If that proves successful, we will incorporate them into NAEP operations in 2013, and resume NAEP reporting on Puerto Rico that year.

For your information, I am enclosing a written summary of the studies. If you have questions or need further clarification, please contact Cornelia Orr, Executive Director of the Governing Board, or Peggy Carr, Associate Commissioner for Assessment of the National Center for Education Statistics.

We believe it is important, as required by law, that the National Assessment provides reliable data on student achievement and achievement trends in Puerto Rico schools. This has proved challenging over past few years. We are hopeful now that the challenges will be resolved by producing a more accurate assessment over the full range of student performance. This will fulfill the mission of NAEP and benefit all those who turn to the National Assessment for sound, independent information both in Puerto Rico and throughout the United States.

Sincerely,

David P. Driscoll
Chairman

Enclosure

Similar letters sent to:
Secretary of Education Carlos Chardan
Chairman and Ranking Minority Member, U.S. Senate HELP Committee
Chairman and Ranking Minority Member, U.S. House Education and Labor Committee
Hon. Pedro Pierluisi
A Summary of Six Special Studies on the
National Assessment of Educational Progress
Mathematics Assessment in Puerto Rico

NAEP Education Statistics Service Institute
Howard T. Everson

NAEP ESSI

Prepared for
U.S. Department of Education
National Center for Education Statistics
December 2009
ABSTRACT
During the past year, the National Center for Education Statistics (NCES) created a Puerto Rico Working Group (PRWG) to design and oversee a series of special studies to investigate the performance of public school students in Puerto Rico on the National Assessment of Educational Progress (NAEP) Mathematics Assessment. Members of the PRWG included representatives from NCES, Alliance contractors, and NAEP-ESSI.
This report presents a summary of six separate studies conducted by the PRWG in 2009. These include: (1) studying the fit of the mathematics items for Puerto Rican students to examine both student and item characteristics that might have led to item misfit in the NAEP Mathematics assessment; (2) examining the Spanish translation of the NAEP mathematics items and the potential for translation errors; (3) convening a group of mathematics teachers in Puerto Rico to study the potential sources of difficulty in NAEP mathematics items; (4) conducting interviews and a “cognitive lab” study to determine how well students in Puerto Rico understood the NAEP mathematics items, and the describe the solution strategies they employed when tested; (5) assessing the degree of alignment between the NAEP mathematics framework and the mathematics content standards and assessments used in Puerto Rico’s state-wide accountability tests; and (6) investigating the feasibility of creating specially designed blocks of mathematics test items for use in Puerto Rico during the next round of NAEP testing. The findings from these six studies are synthesized and a set of recommendations aimed at strengthening the validity of the NAEP Mathematics assessment for use in Puerto Rico is presented for consideration by NCES.
INTRODUCTION

In the pages that follow we present a summary of six separate studies commissioned in 2009 by the U.S. Department of Education’s National Center for Education Statistics (NCES) to shed light on the poor performance of public school students in Puerto Rico on the National Assessment of Educational Progress (NAEP) Mathematics Assessment.

For nearly forty years, the NAEP has been gathering information on American elementary and secondary students’ academic achievement across a variety of subjects. The NAEP assessments serve as the Nation’s Report Card and, consequently, play a central role in gauging the educational achievement of American students—providing a highly regarded benchmark of student achievement.

In an effort to gauge the mathematics achievement of elementary students in Puerto Rico, the NCES administered Spanish-language versions of the NAEP Mathematics assessment administered to public school students in Puerto Rico at grades 4 and 8 in 2003, 2005, and again in 2007. At each grade level, and on each assessment occasion, public school students in Puerto Rico scored much lower than students in other jurisdictions. Additionally, compared to other states and jurisdictions, higher levels of missing data and fewer correct responses were observed at the item level in Puerto Rico. Moreover, the discrepancy between observed (empirical) and expected (model-based) responses was large, indicating that many of the mathematics items did not fit the assumptions guiding the NAEP Mathematics scale development. These troubling results are documented in a series of reports published by NCES, and are available online at http://nationsreportcard.gov/puertorico.

To better understand the factors contributing to the performance of Puerto Rican students, the National Center for Education Statistics, in 2008, convened a *NAEP Puerto Rico Working Group* (PRWG) to design and conduct research that would, ultimately, inform the longer range goal of designing a NAEP mathematics assessment to measure accurately the achievement of public school students in Puerto Rico. This work resulted in a set of six separate investigations that included: (1) studying the fit of the mathematics items for Puerto Rican students to unveil student and item characteristics that might have led to item misfit in the NAEP Mathematics assessment; (2) examining the translation of the mathematics used in NAEP items and the potential for translation errors; (3) convening a group of mathematics teachers in Puerto Rico to study the potential sources of difficulty in NAEP mathematics items; (4) conducting interviews and a “cognitive lab” study to determine how well students in Puerto Rico understood the NAEP mathematics items and the solution strategies they employed when tested; (5) assessing the degree of alignment between the NAEP mathematics framework and the mathematics content standards and assessments used in
Puerto Rico’s state-wide accountability tests; and (6) investigating the feasibility of creating specially designed blocks of mathematics test items for use in Puerto Rico during the next round of NAEP testing.

In the following section we describe each of these studies and the findings from each are presented and discussed. (The full reports of these studies, including methodological and technical details, are posted on the NAEP IMS at https://ims.naepims.org/collaboration/puertorico/reports.) These summaries are then followed by a synthesis of these findings that aim at extending our understanding of Puerto Rican students’ pattern of poor performance on the NAEP Mathematics assessments. We conclude with a set of recommendations for steps that NCES might take to strengthen the validity of the NAEP Mathematics assessment in Puerto Rico.

**SUMMARY OF STUDIES**

*In-Depth Analysis of the Puerto Rican NAEP Data (2009)*. Wu, M., von Davier, M., Kulick, E., Davis, S., Pitoniak, M. (Educational Testing Service) and Beaton, A. (Boston College).

This study has been referred to as “the person/item-fit study of NAEP mathematics in Puerto Rico.”

As we noted earlier, a consistent pattern of results from the NAEP Mathematics assessments in Puerto Rico in 2003, 2005, and 2007 at both grades 4 and 8 showed that student performance in Puerto Rico did not fit the assumptions of the statistical (psychometric) model used to calibrate scores for students on the NAEP. This study attempted to identify both the item and student factors that may have contributed to model misfit. The research team reasoned that a clearer understanding of the factors that affect item and person psychometric model fit in Puerto Rico would possibly benefit future item development activities, as well as better inform interpretations of performance results.

*Design and Methodology.* This study extended the previous research on the NAEP mathematics model fit in Puerto Rico by examining the correlation of standardized fit residuals for the items, as well as comparing means and standard deviations of these residuals across groups of students. The standardized residuals, which were based on the Beaton Fit-Index, were calculated using data from the 2007 Mathematics assessment. Using factor analysis, the residuals were then examined for patterns across items, considering characteristics of the items (such as subscale, difficulty, verbal load, etc.) to identify patterns of misfit. Means and standard deviations (of the residuals) for different groups of students were compared using ANOVA techniques to identify characteristics that represent Puerto Rican students whose performance did or did not fit the NAEP model.
Measures of students’ gender, motivation and need for learning accommodation, as well as school location and type, were used as covariates to explore their possible relationships to the emerging item-fit patterns. Although the focus of this study was on Puerto Rico, the interpretations were contrasted with results from other states tested by NAEP. The research team chose NAEP item-fit data from Missouri because it was more or less representative of the mathematics ability distributions of many states across the United States.

Results and Findings. In general, the magnitudes of residual correlations were rather small (ranging from 0.10 to 0.27 across all blocks of test items). These small residuals provided little evidence about significant patterns of item misfit. While the residuals in Puerto Rico were larger in magnitude than those from Missouri, there were no obvious patterns in the correlations of residuals. Indeed, the correlations were quite similar between Puerto Rico and Missouri. No dominant component was found in any of the block-based Principle Component analysis results. Thus the model fit, or lack thereof, in Puerto Rico could not be attributed to a certain type or cluster of mathematics items. The results of the ANOVAs suggested that the effect of gender on model fit was only marginally significant. All the other variables included in the ANOVAs were not significant. A summary table of fit statistics by item classifications was shared with the cognitive laboratory team to help inform their item selection process.

Conclusions and Recommendations. The results of the study provided little evidence that any of the factors investigated were related to the magnitude of misfit, item characteristics, or examinee background variables. While these findings do not fully explain the model behavior in Puerto Rico, they are informative from a test development point of view—the mathematics items showed consistency, and the instrument was “fair” for students with low mathematical abilities. However, because no clear associations emerged between the examined item and student attributes and the fit patterns, the authors were unable to make recommendations for possible adjustments to improve the model fit.

Solano-Flores, G. and Chia, M. Y., University of Colorado at Boulder.
This study, known within the PRWG as the “translation study”, examined the relationship between the characteristics of the Spanish translation of the Mathematics used in the 2007 Grade 4 and Grade 8 National Assessment of Educational Progress (NAEP) in Puerto Rico and the performance of the Puerto Rican students on the assessment. Three research questions guided this study:
What kinds of translation errors could be identified in the items of the NAEP mathematics assessment administered in 2007 in Puerto Rico?

To what extent do those translation errors account for the low performance of Puerto Rican students in that assessment?

What lessons can be learned from reviewing translation error in this assessment that could then be used to improve the translation process of future Spanish translations of NAEP mathematics items used in Puerto Rico?

The theoretical foundation for the study comes from the theory of test translation error developed by Solano-Flores and his colleagues (Solano-Flores, Backhoff, & Contreras-Niño, 2009). From this perspective, errors of translation (in this case we are referring to the translation of NAEP mathematics items) are not necessarily the result of poor quality translations, but rather may stem from “the fact that languages encode meaning differently and have different sets of features like grammar differences and tolerance to ambiguity” (see p. 1, Solano-Flores & Chia, 2009).

**Design and Methodology.** A multidisciplinary team of nine native Spanish-speakers used a translation error coding system to examine the Spanish translation of a sample of 69 4th grade and 71 8th grade NAEP mathematics items used in 2007 in Puerto Rico. These items were drawn proportionately from each of the content areas tested by NAEP, and included proportional numbers of items with highest, intermediate, and lowest values of the Beaton fit index mentioned earlier in the study by Wu et al., 2009. The multidisciplinary team included two mathematicians, two 4th grade mathematics teachers, two 8th grade mathematics teachers, one mathematics curriculum specialist, one translator, one sociolinguist and a psychometric specialist.

The panel members coded the items according to nine primary translation error dimensions grouped into three major categories: design, language, and content. An additional category, “origin,” included errors that were present in the original version of the NAEP mathematics item. Participants first reviewed each item in Spanish only, and then compared the English and Spanish versions. After each participant coded the items alone, the research staff facilitated a group discussion and ultimately items were coded dichotomously at the category level through group consensus.

**Results and Findings.** Twenty-five NAEP mathematics items were classified as “objectionable”, and the remaining 115 were viewed as acceptable translations by the panel. A
number of translation errors were identified, including imprecise translation of terms, use of technical terms not used in Puerto Rico, confusing and unnatural syntax, use of unfamiliar contextual information, and distortions in graphic material that might lead to misinterpretations of graphs. The results from the correlation analyses suggested that translation errors do not appear to be an important factor with respect to the low performance of Puerto Rican students in the 2007 NAEP Mathematics assessment.

Conclusions and Recommendations. While various types of translation errors were observed, they did not appear to be responsible for low student performance. Translation errors occurred with about the same frequency among the 4th and 8th grade items; however, because of the variation in the academic language and the linguistic demands posed by certain item formats, translation errors may be more likely in some content areas than in others.

The authors suggest that NCES translation contractors be required to use a review procedure as the one described in this study as part of the routine process of test translation. These procedures include the approach used to code translation errors, as well as the use of multidisciplinary teams of translation reviewers. Finally, it was recommended that NCES conduct studies that examine the enacted curriculum in Puerto Rico. The authors argue that knowing what is taught in the classrooms in Puerto Rico is critical to properly interpreting the information on curriculum representation as manifested in the NAEP mathematics assessment.


This study, “the blind item review study”, was designed to investigate whether fourth- and eighth-grade NAEP mathematics items used in Puerto Rico in 2007 included elements of invalid moderators of difficulty for the target population. An invalid moderator of difficulty was defined as an item characteristic that affects the students’ ability to demonstrate their true competence, and is irrelevant to the construct being measured. Conceptually, it is similar to the notion of construct irrelevant variance. In theory, invalid moderators of item difficulty occur when an item contains needlessly complex language, uses unfamiliar graphs, charts or tables, or contains unfamiliar terms, words or phrases. The contextual features of an item can also moderate item difficulty, depending on the learning experiences of the target student population.

**Design and Methodology.** Thirty five Puerto Rican teachers (20 at grade four and at 15 grade eight) rated both NAEP mathematics items and items drawn from local sources (i.e., from textbooks and tests used in Puerto Rico) without knowing the sources of those items. The teachers used a specially
designed questionnaire, known as the Item Review Tool, and rated each test item using 11 dimensions or characteristics representing types of invalid moderators of item difficulty (e.g., unfamiliar terms, needlessly complex language, unfamiliar graphics, etc.) In addition, a third pool of items (accessible block items) was also rated, which included grade four NAEP pilot items designed to reduce construct irrelevant aspects. NAEP and local items were compared on teacher ratings. Each booklet included roughly 40 items and each item was rated by seven raters. Teachers were also asked to comment on the items they rated unfavorably.

At grade four, the mean ratings of NAEP items with best- and worst-fit statistics were significantly different on three dimensions when contrasted with the mean ratings of the local items: familiarity of the visuals (such as graphs, tables, etc.) (p<.05), complexity of the language (p<.05) and how demanding the calculations were (p<.10). For the most part, all the items, regardless of source, were rated favorably by the teachers. However, local items received perfect ratings more often from the teachers. On the other hand, when the mean ratings of the accessible NAEP items were compared to those of the local items, there were no significant differences except for the summative statement (I would use this item to assess my students) in the Item Review Tool. For this statement, the teachers rated the accessible items more favorably (at p<.10) compared to the local items.

At grade eight, the teachers rated the local items more favorably (p<.05) compared to both types of NAEP items included in the study (those with best and those with worst fit statistics) on five dimensions: familiarity of mathematical terms, complexity of the visuals (such as graphs, tables, etc.), how demanding the calculations were, how heavy the reading involved in the item was, and the familiarity of the context in the item.

Conclusions and Recommendations. There were a number of important dimensions on which the NAEP and local Puerto Rican mathematics items differed. In discussions with the teachers, a number of problematic features of the NAEP mathematics items, particularly at Grade 8, were identified. These included difficult and complex visuals, difficult abstractions and content knowledge, and minor translation issues. A majority of the teacher comments and recommended modifications to the ‘problematic’ NAEP items were aimed at making them less abstract, more concrete, and generally easier for their students. Future item development, translation and adaptation activities should take such dimensions into account. Moreover, the pilot accessible NAEP items fared well when contrasted with the local mathematics items. Hence, this study provides one piece of evidence that suggests that the use of accessible NAEP items in Puerto Rico ought to be considered in the future.

While a number of the other investigations in this portfolio of studies focused on item translation/adaptation, or item fit, the purpose of this study was to examine in greater detail students’ understanding of the cognitive items in the NAEP mathematics assessment. The motivation for the study centered on the belief that student interviews using the “cognitive lab” or “think aloud” methodology (Ericcson and Simon, 1999) would allow us to better understand students’ comprehension of NAEP mathematics items, including: whether the student understood the Spanish version of the item as intended; or whether the student encountered any unfamiliar words, contexts, or stimulus materials (e.g., graphics, etc.); and how the student approached and solved the mathematics items. The interviews did not focus on whether the student answered the items correctly, although this information was collected, but rather on acquiring information that could potentially inform item development and translation/adaptation activities for future items. With this as the goal, the ETS research team conducted a series of one-on-one interviews in the Spring of 2009 with grade 4 and 8 students from public schools in Puerto Rico.

Design and Methodology. The ETS research team collaborated with ASPIRA Inc. of Puerto Rico (http://www.aspirapr.org/) to facilitate the field activity, including recruiting interviewers. The cognitive laboratory project was directly coordinated and overseen by the Executive Director of ASPIRA of Puerto Rico. ASPIRA had a number of responsibilities, including: hiring interviewers, participating in interviewer training, coordinating logistics in Puerto Rico, including working with the PRDE to identify the school sample, obtaining school and parent permission, administering the cognitive laboratory interviews, recording student responses, compiling student data and returning the data files to ETS, and participating in feedback session with ETS. Thirty-five items selected from grades 4 and 8 of the 2007 assessment were assembled into seven 5-item mini-tests. The items were representative of the different content areas and item types that appear on NAEP. For grade 4, about 40% of the items were from the accessible booklet study. Mathematics items with various types of stimuli (i.e., geometric figures, graphs, etc.) were also included. Also, the statistical item fit information obtained from the study described earlier, In-Depth Analysis of the Puerto Rican NAEP Data, was considered during the item selection process so that both items that did and did not fit the NAEP model were included in the study.

A generic interview protocol was developed which focused on two general questions: 1) Does the student’s understanding of the Spanish version of the item match the intent of the item in the original
version in English?; and 2) has the student studied the particular topic assessed by the item in his or her math classes? A total of 151 students from seventeen school participated in the study (76 grade 4 and 75 grade 8 students). All of the materials were provided in Spanish and the entire interviews were conducted in Spanish. The analysis of the cognitive laboratory data examined the information regarding the students’ ability to read aloud the question, crossed with specific variables and questions of interest.

Results and Findings. A total of 372 responses to items (“instances”) were coded from students at grade 4 and a total of 368 responses were coded from students at grade 8. Many of the items were read aloud correctly by the students (42% at grade 4 and 56% at grade 8). Approximately 9% of the instances in each grade had difficulty reading one word. Students had difficulty reading the items in 12% of the instances at grade 4, but only 3% of the instances at grade 8. In addition, in 23% of the grade 4 instances and 24% of the grade 8 instances, students misread or misinterpreted either a number, fraction, or symbol. Overall, about half of the item responses were read correctly. In addition, almost one-quarter of the item responses included a misinterpretation or misreading of a number, fraction, or symbol. At grade 4, poorer reading ability was associated with short constructed-response items types, and misinterpretation and misreading of numbers, fractions, or symbols was associated with multiple-choice and extended constructed-response items. Therefore, it seems that students’ ability to understand the question is hampered by both their lack of reading ability and by their lack of knowledge of mathematical terms and symbols.

When asked to “think aloud” the process they used to answer the questions, students generally did not use the appropriate mathematical procedures. Rather, students would guess at an answer, particularly for multiple choice items, or estimate a response. In addition, students would often seem confident of their answer, but could not explain how they obtained the answer. The general patterns mentioned above may be tied to regional differences, particularly in the Metro area at grade 4 and the North area at grade 8. In addition, at grade 4, student responses associated with schools under an Improvement Plan under NCLB also showed the same general pattern. Reading the question correctly was associated with familiarity with the question or topic at grade 8, previously studying the concept in school at grade 4, and a correct response at both grades. In addition, while the findings are mixed across content areas, it is interesting to note that very few problems were seen for geometry items, at either grade. Of specific interest are the three concepts that were found to be problematic for students in Puerto Rico: fractions, temperature/degrees, and ordinal numbers.

Conclusions and Recommendations. It is important to keep in mind that in this study there was no control group, thus we are left wondering whether we would find comparable patterns among
students from other low-performing NAEP jurisdictions. The study results are, nonetheless, provocative. They indicate, for example, that student knowledge, particularly reading ability and familiarity with mathematical terms, is likely contributing to Puerto Rico student low-performance in NAEP.


The goal of this study was to provide information and statistical data to permit a determination of the degree of alignment between and among the NAEP Mathematics Framework for 2009 for grades 4 and 8, and the 2007 and 2009 mathematics standards and assessments that are associated with the Pruebas Puertorriqueñas de Aprovechamiento Académico (the PPAA) in Puerto Rico. Design and Methodology. A central design challenge for this project stemmed from the fact that the Puerto Rico mathematics frameworks and assessments changed substantially between 2005 and 2009. The relevant modifications to the mathematics standards and assessments were summarized in documents published by the Puerto Rico Department of Education (PRDE), including Estándares de Excelencia, Programa de Matematicas (2000), Estándares de Contenido y Expectativas de Grado (2007), and Pruebas Puertorriqueñas de Aprovechamiento Académico (PPAA, 2009, see http://de.gobierno.pr).

In general, the research design followed the approach outlined by Norman Webb (1997, 2005, 2007), and used the terms “standards” and “objectives” to describe the levels of expectations for what students should “know and be able to do” at specific grade levels.

The Webb alignment approach used in this study asks subject matter experts (i.e., reviewers or panelists) to code each test item and content standard or objectives along the dimensions of categorical concurrence, depth-of-knowledge consistency, and range of knowledge correspondence between standards and assessments.

A three-day alignment panel study was held in San Juan, Puerto Rico from July 28-30, 2009. Sixteen panel members, consisting of teachers, curriculum specialists, and mathematics education professors, drawn from elementary, secondary and post-secondary faculty in Puerto Rico, were recruited to review the 2009 NAEP mathematics content standards for grades 4 and 8, as well as the 2007 and 2009 mathematics content standards and assessments associated with the Pruebas Puertorriqueñas de Aprovechamiento Académico, the PPAA. The panelists examined the alignment between PPAA items and the NAEP framework (60 items for 2007, and 54 for 2009; in each grade). The research team was helped by the Director of Mathematics Curriculum in the Puerto Rico Department of Education, who made direct comparisons between the two sets of mathematics standards (the...
Estandares de Excelencia and the Estándares de Contenido y Expectativas de Grado) and the current 2009 NAEP Mathematics Framework to assess their degree of alignment. The degree of agreement between and among the PPAA content standards, the NAEP content standards, and the assessment items, was derived from the panelists’ classifications whose responses were averaged among the panelists within grade-level. In general, the panel members were instructed to identify any one assessment item as corresponding to up to three objectives—one primary objective and up to two secondary objectives. Each review session was then followed by a short period of debriefing in which panel members were afforded the opportunity to share comments and suggestions for improving the alignment of the standards and assessment items.

Results and Findings. According to the panelists, four of five NAEP mathematics content standards were covered by the sample of 2007 4th grade PPAA mathematics items. Surprisingly, the algebra content standard was not covered sufficiently enough. In addition, the DOK levels for both the measurement and geometry items were considered “weak.” This sample of mathematics items had an acceptable balance of representation, but the range-of-knowledge criterion was not met. Overall, the alignment with NAEP appears insufficient.

The 2007 grade 8 PPAA items provided adequate categorical coverage of the NAEP standards. The DOK criterion for the geometry standard was considered “weak”, and the range of knowledge criterion was “weak” for both the geometry and algebra content standards in NAEP. The alignment of the 8th grade PPAA items from 2007 with NAEP is marginal.

The picture was considerably better for the 2009 PPAA items at both the 4th and 8th grades. At the 4th grade, for example, the categorical concurrence, DOK, and balance of representation criteria were all met. The item pool needs bolstering, in the reviewers’ estimation, when it comes to the range-of-knowledge covered by the items—particularly with respect to number and operations, measurement, and geometry. A similar pattern emerged for the 8th grade PPAA mathematics items. The categorical concurrence, DOK, and balance of representation criteria were all met. This set of mathematics items, however, had less than acceptable range-of-knowledge for all the NAEP content standards, with the exception of algebra.

Working with one of the other panel members, the Director of Mathematics in the Puerto Rican Department of Education reviewed the Estándares de Excelencia (2000) and concluded that they did not adequately cover the 4th or 8th grade NAEP content standards—they were viewed as too broad and not aligned with the grade-level content.
standards and objectives of NAEP. The reviewers, however, reported that the current mathematics standards in place in Puerto Rico, the *Estándares de Contenido y Expectativas de Grado* (2007), were well aligned with NAEP’s content standards and objectives, and met the criteria for full alignment.

**Conclusions and Recommendations.** Overall, across all the alignment activities, the majority of panel members indicated that the PPAA test items did, indeed, cover the most important topics expected by the NAEP standards. There was a sense among the panelists that the 2007 PPAA items, in contrast to the 2009 PPAA items, covered fewer NAEP standards and objectives. The one consistent finding that emerged was that the current mathematics standards and assessments in Puerto Rico, though not perfectly so, were aligned better with the current NAEP Mathematics Framework than the previous set of standards used in Puerto Rico, which were in place at the time of the most recent NAEP mathematics assessment.

The single most consistent finding that emerged during the three-day meeting was that the current mathematics content standards and assessments in Puerto Rico, those based on the *Estándares de Contenido y Expectativas de Grado* (2007), were better aligned with the current NAEP mathematics content standards than the previous set. Thus, it is reasonable to assume that the demonstrable lack of alignment between the older set of PPAA mathematics content standards and the content assessed by NAEP in 2003, 2005 and again in 2007 may have contributed to the relative poor performance of public school students in the Puerto Rico.


This study used a two-stage approach to investigate and evaluate the feasibility and suitability of creating specially designed blocks of NAEP mathematics items for use in Puerto Rico that would function similarly if used in a larger national sample. If successful, this block assembly method would enable NCES to obtain meaningful and reliable results in Puerto Rico which would be comparable to those from the national sample.

**Design and Methodology.** The tasks delineated for Stage One included (1) identifying items from the 2007 operational assessments at grades 4 and 8 that performed similarly in both the Puerto Rico and National samples; (2) evaluating this collection of items to determine if they are sufficiently distributed across the framework to permit the assembly of two or more prototype blocks of NAEP mathematics items at the 4th and 8th grade; and (3) assembling the blocks using the items identified in step 1, above.
Stage Two required (4) Simulating performance on the blocks assembled in Stage 1 for both the Puerto Rico and the National Public samples; (5) performing item analysis on the simulated Puerto Rico and National Public samples to evaluate classical-test-theory-based psychometric characteristics of the pseudo blocks; (6) assembling a panel of experts to compare the group of items identified in Stage 1 with the remaining items from the 2007 assessment. The panel would look for patterns, based on various criteria that might emerge from comparing the two groups of items. The criteria for selecting the items included (a) students performing above the chance level on the items; (b) similar student performance in Puerto Rico and in the National samples; (c) scaled scores considered only if within the range attained by the majority of Puerto Rican students.

Finally, the resulting item pools were evaluated on the basis of meeting the major content areas covered by NAEP, balanced item types, sufficient mathematical complexity, and coverage of content subtopics.

Results and Findings. The items in the pool are not distributed across the content areas at either grade 4 or grade 8 in the proportion needed to assemble two blocks (i.e., at least one test book).

In summary, for each of the two grade levels, the distribution of the items in the pool by item type is close, but not identical to, the distribution specified in the NAEP Mathematics framework. On the other hand, for each grade level, the distribution of the items in the pool by complexity has a higher proportion of low complexity items than the operational assessment. And finally, at each grade, the distribution of items across the subtopics within the major content areas is insufficient to meet block assembly targets.

Conclusions and Recommendations. The study illustrated that it would only be possible to assemble one block at each grade level if the assumptions about similar performance in the Puerto Rico and national samples and about framework coverage were relaxed. The resulting blocks, we hasten to add, would not be representative NAEP mathematics blocks as currently configured.

Keeping in mind that, at most, one block can be assembled at each grade level, it does not seem to be worthwhile to pursue the Stage 2 tasks outlined in the earlier proposal. The results of the study suggest that it is not possible to assemble a subset of NAEP items for Puerto Rico that perform similarly in Puerto Rico and the Nation and that adequately represent the NAEP Mathematics framework.
SYNTHESIS OF FINDINGS

It is now time to ask what have we learned from this collection of studies? What do the six studies collectively tell us? Do they, individually or collectively, uncover the reasons for the pattern of startlingly poor performance on the NAEP Mathematics assessment by public school students in Puerto Rico? Was it because of the psychometric quality of the NAEP assessment? Was something lost in the translation of the mathematics items into Spanish? Were the NAEP mathematics items too unfamiliar or linguistically complex for these students? Or perhaps, the students were not taught the mathematics tested by the NAEP. Given the pool of NAEP mathematics items, we asked if we could build an assessment for this sample of students that would be more informative and more useful. These questions, and many others, animated the efforts of the researchers whose work is summarized in this report.

From a psychometric perspective, the NAEP results in Puerto Rico reveal large amounts of missing data (e.g., omitted and not reached items), fewer correct responses than expected in every content area, and a surprising mismatch between expected and actual student performance on the mathematics items when compared to students performance in other states (Baxter, Ahmed, Sikali, Waits, Sloan, and Salvucci, 2007). The “person/item-fit study” reported earlier by Wu et al. (2009) found no clear relationship between the degree of item misfit and characteristics of the examinees. The research team of Solano-Flores and Chia (2009) shifted focus and examined the quality of the translations of the mathematics items. This study used state-of-art linguistic perspectives, and examined a variety of sources of translation error. Though they surfaced a number of issues related to the quality of the translations (e.g., imprecision in the translation of some terms, confusing syntax, etc.), their results indicated that such errors were unrelated to students performance on the 2007 NAEP mathematics assessment. In a related vein, Dogan and Rivas (2009) used school teachers in Puerto Rico to review and rate samples of mathematics items from NAEP and elsewhere to investigate if the NAEP items contain language and terms that would be unfamiliar to students in Puerto Rico, or contained strange and uncommon terms or graphics. Perhaps these item characteristics would account for the performance differences? In general, the teachers looked favorably on all the items, regardless of whether they were from NAEP or from assessments used locally in Puerto Rico. There were, however, some dimensions of the NAEP items that were identified as problematic by the teachers, including, for example, complicated visuals, minor translation issues (some of these were also noted by Solano-Flores & Chia), and contextual abstractness. Although the study by Dogan and Rivas produced no eureka effect, their work did shed light on ways to improve the NAEP mathematics items in the future.
These three studies, though generally informative, did not point us in a clear direction. The internal characteristics of the assessment—their fit to the psychometric model, their translations, their uniqueness—do not tell us much about the nature of the problem. So we are left to wonder if the students or teachers can tell us anything more that would help. Everson, Rivas, and Rodriguez (2009) convened a panel of 16 mathematics teachers from Puerto Rico and asked them to tell us about the degree of alignment between what is tested by NAEP and the mathematics curriculum that has been enacted in the public schools. The teachers told Everson et al. that, in their view, the alignment had improved over the past three or four years, as changes were introduced to the mathematics curriculum in Puerto Rico. The content standards in place between 2000 and 2006 was not well aligned with NAEP and, according to the teachers, may partially explain the depressed performance levels of Puerto Rican students on the NAEP in 2003 through 2007.

To continue our story, we turn to the cognitive laboratory study conducted by Drescher et al. (2009). In this study the researchers were interested in the mental processes (the cognitions) used by students in Puerto Rico as they attempted to solve NAEP mathematics items. Using these “think aloud” methods, they wanted to better understand how these students understood or comprehended the mathematics items. Do the students in Puerto Rico think differently about mathematics, and approach problem solving in ways that are not unlike students in other states? Could this explain why they performed so poorly? The findings here are provocative. While thinking aloud it was apparent that many of the students did not employ appropriate mathematics procedures, and many often guessed when they could not solve a problem. The research team also noted that overall levels of reading ability, and the accompanying unfamiliarity with standard mathematical terminology, may also be interfering with the students’ performance on the NAEP mathematics assessment. Again, these findings appear to tell only part of the story.

The remaining study in the sextet asked whether it was feasible to create a specially designed block of NAEP mathematics items for use in Puerto Rico. Perhaps such an assessment would shed more light on students’ achievement and be used comparatively in other national samples? A study by Dion et al, (2009) at the Educational Testing Service suggested, unfortunately, that this idea was not practical. There are simply too few items that functioned appropriately in earlier administrations of NAEP in Puerto Rico to cover adequately the NAEP assessment framework. Thus, the researchers concluded that it was not possible to create a subset (a block) of NAEP mathematics items for use in Puerto Rico that would also perform similarly when administered to a sample of students from other states. So where do these studies leave us? We turn, next, to the initial recommendations from the Puerto Rico Working Group (the PRWG).
**Initial Working Group Recommendations**

After careful consideration of the collective findings from these six studies, the PRWG offered a number of field-based activities that could be conducted in Puerto Rico in 2011 to produce more meaningful data which could be reported on the NAEP scale. The activities include administering NAEP mathematics assessment in private schools in Puerto Rico, designing blocks of mathematics items that target knowledge and skills appropriate (KaSA) for use in Puerto Rico and elsewhere to improve measurement precision at the lower end of the ability continuum, and enhancing outreach activities in Puerto Rico to help boost student motivation and participation in NAEP. Each of these is described more fully below.

*Administering NAEP in Private Schools.* As is typical in NAEP, the student sample for Puerto Rico has been drawn only from public schools. This NAEP policy has been criticized in Puerto Rico because the population of students attending private schools is perceived as quite different in Puerto Rico, probably more so than in the mainland U.S. For example, 11% of mainland U.S. students attend private schools, while 18% attend private schools in Puerto Rico. Currently, NAEP has no information on the performance or behavioral characteristics of private school students in Puerto Rico. Therefore, the PRWG initially recommended that private schools be included in the sample for the 2011 NAEP administration. However, because of various policy concerns, NCES has decided not to exercise this recommendation at this point.

*Creating KaSA Blocks to administer in Puerto Rico and the Mainland.* In an effort to learn more about what students (including Puerto Rican students) performing at the lower end of the ability distribution know and can do, mathematics items that target the lower-end of the framework could be designed. We recommend, in this instance, that “KaSA blocks” be developed for use in both Puerto Rico and the mainland. In this scenario, these blocks of items would be administered to evaluate whether they do, in fact, provide improved estimates of students’ mathematical proficiency (i.e., smaller standard errors and improved model fit), relative to the regular, more traditional blocks of NAEP mathematics items.

Forming these so-called “KaSA blocks” would require the development of a blueprint that would identify objectives in the current NAEP Mathematics framework that could be assessed with less difficult items, as well as identifying clearly the characteristics of these easier items. The set of items, we suspect, could be developed in ways that are consistent with the NAEP Mathematics framework. However, the framework guidelines for testing time by item type would likely have to be relaxed for these new item blocks. The guidelines used to create the items for the NAEP Validity Study Panel’s *Accessible Booklet Study* may be a good starting place for this development effort. (See the
December 3, 2009 memo from the PRWG for more details on this particular design recommendation.

Outreach Activities. Currently, NAEP has a very small presence in Puerto Rico. With the 2011 assessments around the corner, and the introduction of a new NAEP State Coordinator, we believe there is an opportunity to increase awareness about NAEP in Puerto Rico. Limited outreach materials exist for Puerto Rican students, and preliminary data suggest that many students in Puerto Rico have poor test taking skills. Therefore, the PRWG recommends improving these materials and making sure that the Puerto Rico NAEP State Coordinator receives the same support that other state coordinators enjoy.

Additional Background Questions. If NAEP is interested in learning additional information regarding the students, teachers, or school climate, a few background questions could be added to the questionnaires for 2011. These questions could be based on information learned from the six studies conducted earlier this year and provide additional insight into the educational situation and performance in Puerto Rico.

In closing, we restate that the principal aim of this paper was to review and summarize the studies conducted by the PRWG. We also note that these studies are part of a much larger body of research bearing on the validity of NAEP. The studies reviewed here add to this knowledge base and help provide NCES and NAEP with a stronger footing as it moves forward with efforts to assess what students in Puerto Rico know and can do.
REFERENCES


APPENDIX A

Members of the Puerto Rico Working Group

The members of the Puerto Rico Working Group include:

Chris Averett, Westat
Carole Bower, Westat
Gina Broxterman, National Center for Education Statistics
Gloria Dion, Educational Testing Service
Enis Dogan, NAEP- ESSI, American Institutes for Research
Amy Dresher, Educational Testing Service
Howard Everson, NAEP- ESSI, American Institutes for Research/ City University of NY
Andrea Fercsey, Educational Testing Service
Kim Gattis, NAEP- ESSI, American Institutes for Research
Arnold Goldstein, National Center for Education Statistics
Ed Kulick, Educational Testing Service
Sylvia Ledesma, Educational Testing Service
Manuel Padilla, Westat
Mary Pitoniak, Educational Testing Service
Emmanuel Sikali, National Center for Education Statistics
William Tirre, National Center for Education Statistics
Dianne Walsh, Westat
Meng Wu, Educational Testing Service
Executive Summary

The NAEP mathematics assessment was administered to public school students in Puerto Rico for the first time in 2003. Although NAEP had previously administered some of the assessment in Spanish to students who required accommodations, this was the first time an entire NAEP administration was in a language other than English. The NAEP mathematics assessment was administered again to public school students in both fourth- and eighth-grades in Puerto Rico in 2005. This report presents the results of the NAEP mathematics assessment for Puerto Rico for 2003 and 2005. Because of modifications to the 2005 Puerto Rico administration, results from 2003 should not be compared to results from 2005. Although parallel changes were not made in the nation in 2005, within year comparisons between Puerto Rico and the nation are valid.

2003 Findings

• On average, fourth- and eighth-grade students in Puerto Rico scored lower than public school students in the nation.

Figure A.
Average NAEP mathematics scores for public school students in Puerto Rico and the nation in 2003

<table>
<thead>
<tr>
<th>Grade 4</th>
<th>Puerto Rico</th>
<th>Nation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale score</td>
<td>179</td>
<td>234*</td>
</tr>
<tr>
<td>Grade 8</td>
<td>Puerto Rico</td>
<td>Nation</td>
</tr>
<tr>
<td>Scale score</td>
<td>212</td>
<td>276*</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from students in Puerto Rico.


• At fourth-grade, 9 percent of students in Puerto Rico and 76 percent of students in the nation scored at or above Basic. At eighth-grade, 4 percent of students in Puerto Rico and 67 percent of students in the nation scored at or above Basic.

• Fourth-grade female students in Puerto Rico scored significantly higher than male students in the geometry and spatial sense content area.

2005 Findings

• Overall, fourth- and eighth-grade students in Puerto Rico scored lower, on average, than public school students in the nation.

Figure B.
Average NAEP mathematics scores for public school students in Puerto Rico and the nation in 2005

<table>
<thead>
<tr>
<th>Grade 4</th>
<th>Puerto Rico</th>
<th>Nation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale score</td>
<td>183</td>
<td>237*</td>
</tr>
<tr>
<td>Grade 8</td>
<td>Puerto Rico</td>
<td>Nation</td>
</tr>
<tr>
<td>Scale score</td>
<td>218</td>
<td>278*</td>
</tr>
</tbody>
</table>

* Significantly different (p < .05) from students in Puerto Rico.


• Twelve percent of students in Puerto Rico and 79 percent of students in the nation scored at or above Basic in grade 4. Six percent of students in Puerto Rico and 68 percent of students in the nation scored at or above Basic in grade 8.

• Eighth-grade female students in Puerto Rico scored significantly higher than male students in the data analysis and probability content area.

About this report

Throughout this report, results for Puerto Rico are compared to results for public school students in the nation because in Puerto Rico only public school students participated in the 2003 and 2005 NAEP mathematics assessments. The national sample does not include Puerto Rico at this time, although the intent is to include Puerto Rico as part of the national sample in future NAEP administrations.
Executive Summary

In 2007, public school students in Puerto Rico at grades 4 and 8 participated in a Spanish-language version of the National Assessment of Educational Progress (NAEP) in mathematics. A representative sample of approximately 2,800 students from 100 public schools was assessed at each grade.

This report contains performance results on NAEP mathematics questions for public school students in Puerto Rico and the nation. Results are presented as the average scores for the correct answers (see box below)—expressed as decimals ranging from 0.00 to 1.00—for all the questions in the NAEP mathematics assessment and for questions in each of the five mathematics content areas (as shown in figures A and B).

At grade 4
- The average of the question scores for students in Puerto Rico was lower than the score for students in the nation overall and within each content area.
- There was no statistically significant difference in performance between male and female students in Puerto Rico overall and in each content area.

At grade 8
- The overall average of the question scores for students in Puerto Rico was lower than the score for students in the nation. Results were similar for each content area.
- While there was no significant difference between the performance of male and female students in Puerto Rico overall, male students had a higher score than female students in the measurement content area, and female students had a higher score than their male peers in the data analysis and probability content area.

Interpreting Results for Puerto Rico

Question scores are calculated as the percentages of correct responses for multiple-choice questions and for constructed-response questions that are scored either correct or incorrect. For constructed-response questions that allow for partial credit, the question score is the sum of the percentage of students receiving full credit and a fraction of the percentage receiving partial credit. Individual question scores are then averaged together to report an average question score for the entire mathematics assessment or for each of the five content areas.

Because of technical concerns regarding the placement of the 2007 results for Puerto Rico on the NAEP mathematics scale, performance results could not be reported as average scale scores for Puerto Rico in this report, and students’ performance in 2007 could not be compared to performance in previous assessments.

When comparing the results for students in Puerto Rico to students in the nation, it is important to consider some of the differences in demographics. For example, between 76 and 78 percent of fourth- and eighth-graders in Puerto Rico attended public schools compared to 91 percent in the nation. All of the public school students in Puerto Rico were eligible for the National School Lunch Program compared to between 41 and 46 percent of fourth- and eighth-graders in the nation.
Puerto Rico Attains Low NAEP Scores

By Kathleen Kennedy Manzo

Students in Puerto Rico’s public schools are faring poorly in mathematics compared with their peers in the 50 states or even large urban districts, according to results of the National Assessment of Educational Progress released today.

Those results from the commonwealth were reported in a complicated format because of concerns about the validity of the scores the 4th and 8th graders received on the 500-point scale normally used in NAEP reports. Instead of scores, or achievement levels such as “basic” or “proficient,” the results are reported as “the overall average of the question scores,” meaning the percentage of correct responses on multiple-choice questions and those requiring short answers, some of which could receive partial credit.

“The reliability of our estimate for the Puerto Rico score was pretty unstable, in that it had a wide margin of error of confidence around the point estimate,” said Peggy Carr, the associate commissioner of the National Center for Education Statistics, the arm of the U.S. Department of Education that oversees the test. “They were much less reliable in terms of our comfort level with particular estimates, and they were in fact very low.”

Students on the island commonwealth struggled to answer most of the questions on the test. Puerto Rico’s 4th graders received an average 26 percent of possible points, compared with 55 percent for students across the United States. Eighth graders earned an average 25 percent of possible points, compared with 51 percent for students from a representative national sample.

Because of those reporting problems, which Ms. Carr said do not reflect problems with the quality of the assessment, Puerto Rico will not be included in the 2009 math assessment. The NCES will conduct further studies to ensure that future results are reported in a format consistent with those of other tests, she said.

On the math assessment, administered in early 2007, students were tested on number properties and operations, measurement, geometry, data analysis and probability, and algebra. Some 2,800 students at each grade level took the test.

Unique Status

Even relative to some of the nation’s struggling cities, the student population in Puerto Rico is unique among test-takers on the assessment, experts say. All the students in the commonwealth are eligible for the federal free- and reduced-price lunch program. Although Spanish and English are Puerto Rico’s official languages, the former is dominant.

The latest results cannot be compared against those given in Puerto Rico in 2003 or 2005 because of the different method of reporting them. On the 2005 test, whose results were released just last year, 12 percent of 4th graders and 6 percent of 8th graders in Puerto Rico’s public schools scored at or above the “basic” level. So few students scored at the “proficient” or “advanced” level that the percentages rounded to zero.
Those low scores have angered some education officials in Puerto Rico, who argued in a letter to the federal Department of Education last month that the translation of NAEP in math, which has been given to Puerto Rican students in Spanish, as well as cultural differences not taken into account on test items, might be dragging down students’ scores. In the Nov. 12 letter, the commonwealth’s secretary of education, Rafael Aragunde-Torres, asked that U.S. officials allow Puerto Rico to be “permanently exempted” from participating in the test.

But Luis G. Fortuno, the governor-elect, countered in his own letter that the commonwealth should continue taking part. He is expected to appoint a new education secretary after he takes office. ("Puerto Rican Officials Feud Over NAEP Participation," Nov. 21, 2008.)

Puerto Rican students first took the math NAEP in 2003 and took it again in 2005. It was the first time NAEP had been administered to an entire jurisdiction in Spanish for students taught primarily in that language. Their scores were so low, and there was such a mismatch between expected and actual student performance, that federal officials had difficulty interpreting them, resulting in a delay of the release of test results. ("Puerto Rico Falls 'Below Basic' on Math NAEP," April 4, 2007.)

Assistant Editor Sean Cavanagh contributed to this report.
IMPLEMENTATION OF GOVERNING BOARD POLICY ON INCLUSION IN 2013

In 2010, the Governing Board adopted the policy, *NAEP Testing and Reporting on Students with Disabilities and English Language Learners*. This policy requested changes in both the process of data collection and reporting for these two student groups. These changes will be implemented as part of the 2013 assessments. Certain aspects of the policy were implemented in 2011 and 2012. The 2011 NAEP report cards included a section highlighting those jurisdictions that did not meet the NAGB inclusion goals of 95% of all students and 85% of identified students with disabilities and English language learners. In 2012, NCES field tested the new SD and ELL decision trees. The reports from the field suggested that operationally the new decision trees would work and they could be used for the 2013 administration.

One aspect of the policy, however, was found to result in unintended consequences, and cannot be implemented in 2013. This issue concerns the re-classifying of students whose accommodations NAEP does not allow as “refusals” rather than “excluded.” While there are methodological complications that prevent NCES from reclassifying students who cannot participate because their IEP-specified accommodations are not allowed on NAEP, data will be collected in the 2013 assessments to 1) explore alternative ways to report such exclusions and 2) to document how implementation of this aspect of the policy would have affected NAEP results. NCES will discuss such plans at the November Board meeting.
National Assessment Governing Board

Policy Statement on NAEP Testing and Reporting on Students with Disabilities and English Language Learners

INTRODUCTION

To serve as the Nation’s Report Card, the National Assessment of Educational Progress (NAEP) must produce valid, comparable data on the academic achievement of American students. Public confidence in NAEP results must be high. But in recent years it has been threatened by continuing, substantial variations in exclusion rates for students with disabilities (SD) and English language learners (ELL) among the states and urban districts taking part.

Student participation in NAEP is voluntary, and the assessment is prohibited by law from providing results for individual children or schools. But NAEP’s national, state, and district results are closely scrutinized, and the National Assessment Governing Board (NAGB) believes NAEP must act affirmatively to ensure that the samples reported are truly representative and that public confidence is maintained.

To ensure that NAEP is fully representative, a very high proportion of the students selected must participate in its samples, including students with disabilities and English language learners. Exclusion of such students must be minimized; they should be counted in the Nation’s Report Card. Accommodations should be offered to make the assessment accessible, but these changes from standard test administration procedures should not alter the knowledge and skills being assessed.

The following policies and guidelines are based on recommendations by expert panels convened by the Governing Board to propose uniform national rules for NAEP testing of SD and ELL students. The Board has also taken into consideration the views expressed in a wide range of public comment and in detailed analyses provided by the National Center for Education Statistics, which is responsible for conducting the assessment under the policy guidance of the Board. The policies are presented not as statistically-derived standards but as policy guidelines intended to maximize student participation, minimize the potential for bias, promote fair comparisons, and maintain trends. They signify the Board’s strong belief that NAEP must retain public confidence that it is fair and fully-representative of the jurisdictions and groups on which the assessment reports.
1. As many students as possible should be encouraged to participate in the National Assessment. Accommodations should be offered, if necessary, to enable students with disabilities and English language learners to participate, but should not alter the constructs assessed, as defined in assessment frameworks approved by the National Assessment Governing Board.

2. To attain comparable inclusion rates across states and districts, special efforts should be made to inform and solicit the cooperation of state and local officials, including school personnel who decide upon the participation of individual students.

3. The proportion of all students excluded from any NAEP sample should not exceed 5 percent. Samples falling below this goal shall be prominently designated in reports as not attaining the desired inclusion rate of 95 percent.

4. Among students classified as either ELL or SD a goal of 85 percent inclusion shall be established. National, state, and district samples falling below this goal shall be identified in NAEP reporting.

5. In assessment frameworks adopted by the Board, the constructs to be tested should be carefully defined, and allowable accommodations should be identified.

6. All items and directions in NAEP assessments should be clearly written and free of linguistic complexity irrelevant to the constructs assessed.

7. Enhanced efforts should be made to provide a short clear description of the purpose and value of NAEP and of full student participation in the assessment. These materials should be aimed at school personnel, state officials, and the general public, including the parents of students with disabilities and English language learners. The materials should emphasize that NAEP provides important information on academic progress and that all groups of students should be counted in the Nation’s Report Card. The materials should state clearly that NAEP gives no results for individual students or schools, and can have no impact on student status, grades, or placement decisions.

8. Before each state and district-level assessment NAEP program representatives should meet with testing directors and officials concerned with SD and ELL students to explain NAEP inclusion rules. The concerns of state and local decision makers should be discussed.
IMPLEMENTATION GUIDELINES

For Students with Disabilities

1. Students with disabilities should participate in the National Assessment with or without allowable accommodations, as needed. Allowable accommodations are any changes from standard test administration procedures, needed to provide fair access by students with disabilities that do not alter the constructs being measured and produce valid results. In cases where non-standard procedures are permitted on state tests but not allowed on NAEP, students will be urged to take NAEP without them, but these students may use other allowable accommodations that they need.

2. The decision tree for participation of students with disabilities in NAEP shall be as follows:

<table>
<thead>
<tr>
<th>NAEP Decision Tree for Students with Disabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BACKGROUND CONTEXT</strong></td>
</tr>
<tr>
<td>1. NAEP is designed to measure constructs carefully defined in assessment frameworks adopted by the National Assessment Governing Board.</td>
</tr>
<tr>
<td>2. NAEP provides a list of appropriate accommodations and non-allowed modifications in each subject. An appropriate accommodation changes the way NAEP is normally administered to enable a student to take the test but does not alter the construct being measured. An inappropriate modification changes the way NAEP is normally administered but does alter the construct being measured.</td>
</tr>
<tr>
<td><strong>STEPS OF THE DECISION TREE</strong></td>
</tr>
<tr>
<td>3. In deciding how a student will participate in NAEP:</td>
</tr>
<tr>
<td>a. If the student has an Individualized Education Program (IEP) or Section 504 plan and is tested without accommodation, then he or she takes NAEP without accommodation.</td>
</tr>
<tr>
<td>b. If the student’s IEP or 504 plan specifies an accommodation permitted by NAEP, then the student takes NAEP with that accommodation.</td>
</tr>
<tr>
<td>c. If the student’s IEP or 504 plan specifies an accommodation or modification that is not allowed on NAEP, then the student is encouraged to take NAEP without that accommodation or modification.</td>
</tr>
</tbody>
</table>
3. Students should be considered for exclusion from NAEP only if they have previously been identified in an Individualized Education Program (IEP) as having the most significant cognitive disabilities, and are assessed by the state on an alternate assessment based on alternate achievement standards (AA-AAS). All students tested by the state on an alternate assessment with modified achievement standards (AA-MAS) should be included in the National Assessment.

4. Students refusing to take the assessment because a particular accommodation is not allowed should not be classified as exclusions but placed in the category of refusals under NAEP data analysis procedures.

5. NAEP should report separately on students with Individualized Education Programs (IEPs) and those with Section 504 plans, but (except to maintain trend) should only count the students with IEPs as students with disabilities. All 504 students should participate in NAEP.

At present the National Assessment reports on students with disabilities by combining results for those with an individualized education program (who receive special education services under the Individuals with Disabilities Education Act [IDEA]) and students with Section 504 plans under the Rehabilitation Act of 1973 (a much smaller group with disabilities who are not receiving services under IDEA but may be allowed test accommodations).* Under the Elementary and Secondary Education Act, only those with an IEP are counted as students with disabilities in reporting state test results. NAEP should be consistent with this practice. However, to preserve trend, results for both categories should be combined for several more assessment years, but over time NAEP should report as students with disabilities only those who have an IEP.

6. Only students with an IEP or Section 504 plan are eligible for accommodations on NAEP. States are urged to adopt policies providing that such documents should address participation in the National Assessment.

**For English Language Learners**

1. All English language learners selected for the NAEP sample who have been in United States schools for one year or more should be included in the National Assessment. Those in U.S. schools for less than one year should take the assessment if it is available in the student’s primary language.

   One year or more shall be defined as one full academic year before the year of the assessment.

---

* NOTE: The regulation implementing Section 504 defines a person with a disability as one who has a physical or mental impairment which substantially limits one or more major life activities, has a record of such an impairment, or is regarded as having such an impairment. 34 C.F.R. § 104.3(j)(1).
2. Accommodations should be offered that maximize meaningful participation, are responsive to the student’s level of English proficiency, and maintain the constructs in the NAEP framework. A list of allowable accommodations should be prepared by NAEP and furnished to participating schools. Such accommodations may be provided only to students who are not native speakers of English and are currently classified by their schools as English language learners or limited English proficient (LEP).

3. Bilingual versions of NAEP in Spanish and English should be prepared in all subjects, other than reading and writing, to the extent deemed feasible by the National Center for Education Statistics. The assessments of reading and writing should continue to be in English only, as provided for in the NAEP frameworks for these subjects.

4. Staff at each school should select from among appropriate ELL-responsive accommodations allowed by NAEP, including bilingual booklets, those that best meet the linguistic needs of each student. Decisions should be made by a qualified professional familiar with the student, using objective indicators of English proficiency (such as the English language proficiency assessments [ELPA] required by federal law), in accordance with guidance provided by NAEP and subject to review by the NAEP assessment coordinator.

5. Schools may provide word-to-word bilingual dictionaries (without definitions) between English and the student’s primary language, except for NAEP reading and writing, which are assessments in English only.

6. NAEP results for ELL students should be disaggregated and reported by detailed information on students’ level of English language proficiency, using the best available standardized assessment data. As soon as possible, NAEP should develop its own brief test of English language proficiency to bring consistency to reporting nationwide.

7. Data should be collected, disaggregated, and reported for former English language learners who have been reclassified as English proficient and exited from the ELL category. This should include data on the number of years since students exited ELL services or were reclassified.

8. English language learners who are also classified as students with disabilities should first be given linguistically-appropriate accommodations before determining which additional accommodations may be needed to address any disabilities they may have.
RESEARCH AND DEVELOPMENT

The Governing Board supports an aggressive schedule of research and development in the following areas:

1. The use of plain language and the principles of universal design, including a plain language review of new test items consistent with adopted frameworks.

2. Adaptive testing, either computer-based or paper-and-pencil. Such testing should provide more precise and accurate information than is available at present on low-performing and high-performing groups of students, and may include items appropriate for ELLs at low or intermediate levels of English proficiency. Data produced by such targeted testing should be placed on the common NAEP scale. Students assessed under any new procedures should be able to demonstrate fully their knowledge and skills on a range of material specified in NAEP frameworks.

3. A brief, easily-administered test of English language proficiency to be used for determining whether students should receive a translation, adaptive testing, or other accommodations because of limited English proficiency.

4. The validity and impact of commonly used testing accommodations, such as extended time and small group administration.

5. The identification, measurement, and reporting on academic achievement of students with the most significant cognitive disabilities. This should be done in order to make recommendations on how such students could be included in NAEP in the future.

6. A study of outlier states and districts with notably high or low exclusion rates for either SD or ELL students to identify the characteristics of state policies, the approach of decision makers, and other criteria associated with different inclusion levels.

The Governing Board requests NCES to prepare a research agenda on the topics above. A status report on this research should be presented at the November 2010 meeting of the Board.
Technical Advisory Panel on Uniform National Rules
for NAEP Testing of Students with Disabilities

Report to the National Assessment Governing Board

July 22, 2009

Chair: Alexa Posny

Members: Louis Danielson, George Engelhard, Miriam Freedman, Claire Greer, Robert Linn, Debra Paulson, and Martha Thurlow
The panel believes the National Assessment of Educational Progress (NAEP) is an important tool for understanding academic achievement among students with disabilities. To ensure that NAEP samples are fully representative and to maintain the comparability of state and district NAEP results, the panel recommends that NAEP:

1. Encourage as many students as possible to participate in NAEP, and provide for the use of allowable accommodations that are necessary to enable students with disabilities to participate.

2. Clarify and expand NAEP’s guidance to schools, encouraging maximum participation of students with disabilities so at least 95% of those drawn for the NAEP sample participate.

3. Report separately on students who have individualized education programs (IEPs) and those with Section 504 plans, but (except to maintain trend) only count the students with IEPs as students with disabilities.

4. Provide incentives for schools to include students with disabilities, including additional outreach and public reporting of participation rates below 95% of students with disabilities.

5. Support research efforts to develop targeted testing for students at both the top and bottom levels of achievement, with sound procedures to identify students to receive targeted test booklets on the basis of their performance on some standard indicator of achievement.

6. Encourage and review research on the identification and progress of students who have a significant cognitive disability but in the short term do not test this 1% of students on NAEP.

7. Assess the English language proficiency of students with disabilities who are English language learners and are drawn for the NAEP sample and provide linguistically appropriate accommodations for those who need them before determining whether additional accommodations may be needed to address any disabilities those students may have.
Although NAEP can establish rules for students to be tested in the same way, individual students participate in NAEP on a voluntary basis, and it is their schools that normally make the decision about whether a student drawn for the NAEP sample participates or not. Therefore, the cooperation of schools and parents is essential to ensure that NAEP samples in every jurisdiction are fully representative and that test results are comparable among the states and districts assessed. The recommendations in this report are intended to be of practical use in determining NAEP testing procedures and in working with states and districts to continue the assessment’s tradition of producing comparable results and useful information.
Technical Advisory Panel on Uniform National Rules for NAEP Testing of Students with Disabilities

Report to National Assessment Governing Board

July 22, 2009

Chair: Alexa Posny

Members: Louis Danielson, George Engelhard, Miriam Freedman, Claire Greer, Robert Linn, Debra Paulson, and Martha Thurlow

BACKGROUND AND INTRODUCTION

The National Assessment of Educational Progress (NAEP) was established in 1969 to measure the academic achievement of a representative sample of elementary and secondary students in the United States. It is sometimes called the Nation’s Report Card. Subsequently, the assessment was expanded to provide representative-sample results for states and large urban school districts.

NAEP is designed to produce valid, comparable data on large groups of students. It is prohibited by law from providing results for individual children or schools. Scores are not intended and (because no student takes the entire test) cannot be calculated for individual students. Because NAEP measures change over time, it can provide participating states and districts with reliable, independent information about the success of their efforts to improve education. It is an important common measure of student performance.

Recently, concern has arisen about the wide variation among states and districts in the rates at which students with disabilities participate in NAEP. Confusion can arise when in some states almost all students with disabilities who are selected for the NAEP sample take the test, and in others many do not. Some advocates for students with disabilities believe that having good information on the achievement of the full population of students with disabilities is a critical tool in improving services for them. The purpose of this report is both to increase the uniformity of NAEP participation rates among states and districts and to make participation rates high and participation procedures uniform.

Specifically, the National Assessment Governing Board (NAGB) convened a technical advisory panel to recommend a uniform set of rules for testing students with a disability on NAEP. The eight-member group held an all-day meeting in Washington, DC, on April 23, 2009, for initial briefings and discussion. The panel conducted four conference calls and exchanged numerous drafts and e-mails between May and July.

The Governing Board charged the panel to make recommendations that:
• provide that students with similar disabilities be tested on NAEP the same way, regardless of where they live;
• maximize student access and meaningful participation;
• ensure that the constructs on NAEP frameworks be measured and that all students may be placed on the same scale;
• permit only accommodations that maintain the validity, reliability, and comparability of NAEP results; and
• are feasible, logistically and financially, and without detrimental consequences.

RECOMMENDATIONS

1. **Encourage as many students as possible to participate in NAEP, and provide for the use of allowable accommodations that are necessary to enable students with disabilities to participate.**

The panel recommends that all students with disabilities participate in NAEP with appropriate accommodations that they need, which are approved by NAEP. The panel understands that some students will not be allowed to use on NAEP some of the accommodations or modifications that are permitted on tests administered by the state or district.

The panel defines an appropriate accommodation as:

i. a change to the way NAEP is normally administered, and
ii. a change that does not alter the construct being measured, and
iii. a change that is needed to enable a student to take the test.

If a proposed accommodation alters the construct being measured, the panel considers it a modification. The panel defines a modification as:

i. a change to the way NAEP is normally administered, and
ii. a change that does alter the construct being measured.

The panel recommends against the use of any change that would alter the construct NAEP is designed to measure, as defined by the NAEP frameworks.

The panel understands that the Governing Board defines the construct underlying the NAEP reading test as “an active and complex process that involves understanding written text.” Because the Governing Board defines this construct to include the ability to decode written text, the panel reaffirms the current NAEP practice of not allowing “read aloud” as an accommodation on the reading test.

The panel understands that the Governing Board defines the construct underlying the NAEP mathematics test as involving five elements, one of which is “Number Properties and Operations (including computation…).” Because this construct includes computation, the panel reaffirms current NAEP practice of not allowing the use of calculators on those parts of the NAEP math test that assess computation.
2. Clarify and expand NAEP’s guidance to schools, encouraging maximum participation of students with disabilities.

As stated previously, the panel recognizes that the testing rules NAEP adopts will not yield comparable state and local results if jurisdictions vary in their participation practices. The panel therefore recommends changes to the guidance given school personnel in deciding whether students drawn for the NAEP sample are to be tested. The panel recommends advising schools on the purpose and nature of NAEP and the desirability of high participation rates, and setting the clear expectation that at least 95% of all students with disabilities drawn for the NAEP sample are expected to take the test.

In a departure from past guidance, the panel recommends state and local decision makers begin with the expectation that almost all students with disabilities will take the test, and then make decisions regarding the accommodations that individual students will be allowed to have. Specifically, the panel recommends this revised Decision Tree be provided to schools:

### NAEP Decision Tree for Students with Disabilities

**BACKGROUND CONTEXT**

1. NAEP is designed to measure constructs carefully defined by frameworks adopted by the Governing Board. Those frameworks include a definition of reading as “an active and complex process that involves understanding written text,” (including the ability to decode text) and include in its definition of mathematics five elements, one of which is “Number Properties and Operations (including computation…).”

2. NAEP provides a list of accommodations that are and are not allowed in reading, mathematics, and other subjects. [See Column B of appendix for accommodations allowed and not allowed on NAEP.]

**STEPS OF THE DECISION TREE**

3. In deciding how this student will participate in NAEP:
   
   a. If the student has an IEP or 504 plan and is tested without accommodation, then he or she takes NAEP without accommodation.

   b. If the student’s IEP or 504 plan specifies an accommodation permitted by NAEP, then the student takes NAEP with that accommodation.

   c. If the student’s IEP or 504 plan specifies an accommodation or modification not allowed on NAEP, then the student takes NAEP without that accommodation or modification.
Students should be excluded from participating in NAEP **only** if they have previously been identified in an IEP as having a significant cognitive disability, and are assessed by the state on an alternate assessment based on alternate achievement standards (AA-AAS). Students should be included if tested on an alternate test with what is called modified achievement standards (AA-MAS).

The panel recommends that guidance to school decision-makers include:

i) a short, clear account of the purpose and value of NAEP, why the inclusion of virtually all selected students is needed to provide representative samples, and the steps to determine how a selected student should participate, and

ii) the target for the percentage of students appropriately to be excluded from participating in NAEP would be 1% of the sample.

The panel also recommends that a broader effort at public information be undertaken to explain the value of NAEP and of securing high participation rates in the assessment.

3. Report separately on NAEP results for IEP and 504 students.

The panel recommends that NAEP report results for both IEP and 504 student groups, but report them separately, and calculate state scores for students with disabilities using IEP results only. At present the National Assessment reports on students with disabilities by combining the results for students with an individualized education program (who receive special education services under the Individuals with Disabilities Education Act [IDEA]) and those with Section 504 plans under the Rehabilitation Act of 1973 (a much smaller group who are not special education students but may be allowed test accommodations).

Under the Elementary and Secondary Education Act, only students with an IEP are counted as students with disabilities in reporting state test results. NAEP should be consistent with this practice. However, the panel recognizes the usefulness of maintaining NAEP trends, and therefore recommends reporting both sets of data and combining results for IEP and 504 students only to preserve the trend line. The panel recommends over time defining students with disabilities for NAEP as only those who have an IEP. All 504 students should participate in NAEP.

4. Provide incentives for schools to include students with disabilities.

The panel recommends that NAEP make enhanced efforts to provide a short clear description of the purpose and value of NAEP and of full student participation in the assessment. These materials should be aimed at school personnel, state officials and the general public, including the parents of students with disabilities.

The panel recommends that upon release of each new set of NAEP results, information indicating the states and districts with more or less than 95% participation rates of students with disabilities with IEPs be among the information bullets highlighted for the
public and the press. All students with 504 plans are expected to participate. Participation rates should be reported both as a percentage of the total sample and as a percentage of the students identified with disabilities within the sample.

The panel further recommends undertaking special studies to look at any outlier states, with unusually high or low exclusion rates, and to continue work previously done for NCES to probe whether there is a cut point beyond which exclusion rates appear suspect.

Some members of the panel noted that there is significant variation among the states in the rate at which they identify students with disabilities for IEPs. While on average states identify about 12-13% of their students as having a disability and needing special education services, some states identify only 9% of their students, and others identify twice that percentage. The differences result mostly from state and local policy rather than the incidence of disability itself. Generally, jurisdictions with high identification rates include more students with mild disabilities. Those with low identification rates include only the more severe, which would make it more difficult to achieve 95% SD participation even though, overall, more of their students may be taking the assessment.

As an alternative to the 95% participation guideline for students with disabilities, some members of the panel recommend that NAEP study the possibility of developing a uniform SD participation guideline based on a percentage of the total student population, regardless of the percent identified as SD. If more than the selected percentage were excluded on the basis of disability, that would be noted in NAEP reports as indicating that the sample was not fully representative. For example, a maximum of 0.6% of the total sample not tested, or 99.4% participating, would correspond to a SD participation rate of 95% where 12% of the sample is identified as having a disability.

5. Support research efforts to develop targeted testing for all students at both the top and bottom levels of achievement, with sound procedures to identify students to receive targeted test booklets on the basis of their performance on some standard indicator of achievement.

The panel recommends that research and development efforts be pursued for NAEP to test all students, not only students with disabilities, at the top and bottom levels of achievement on targeted booklets with a high concentration of difficult or easy items that can be placed on the existing NAEP scale.

Currently all students are tested by NAEP with two 25-minute blocks of items covering a broad range of difficulty, some easy, some difficult, many in the middle. Any student might be randomly assigned any of the various booklets covering the complete range of difficulty for the grade and subject in which he or she is being tested.

The National Center for Education Statistics (NCES) is now developing booklets with a concentration of existing easy items that could be targeted for low-performing students. The panel recommends building upon this research effort, if successful, to create targeted tests at both the top and bottom of the achievement spectrum. High-performing students, those doing work well above grade level, would encounter more challenging items that
allow them to demonstrate knowledge at the advanced level. Likewise, low-performing students would encounter more items that allow them to demonstrate knowledge at the below basic level. This would allow NAEP to measure and report more accurately and in greater detail the knowledge and skills of those students scoring below basic and those scoring advanced. At both ends of the continuum, standard errors would be reduced, and better information would be available about student performance and improvements over time. If needed, additional easy and difficult items should be developed that test NAEP constructs on the existing NAEP scale.

The panel recommends that NAGB attend closely to NCES’ on-going research in this area, and base future decisions on this work and similar research by others. If targeted testing becomes part of future NAEP operations, this information should be described carefully for state and local decision makers. Efforts should be made to explain how these innovations enable students with disabilities who are studying at below basic levels and those who are studying at advanced/above grade levels to engage with NAEP at all points of the continuum of achievement.

The panel recommends that NAEP find an objective and psychometrically sound method to identify which students take any targeted tests that are developed. It recommends consideration of the following possibilities:

- **a) a universal 2-stage process, the system proposed by R. Darrell Bock, in which all students receive a comprehensive block first (a locator test), and then receive either a booklet with a concentration of easy items, a test with a concentration of difficult items, or the usual full-range test in the second block, depending upon their performance on the initial locator test.**

  While this option was the preference of many panel members, it entails major issues of test administration that need to be taken into account before the technique would become feasible.

- **b) a specially constructed new NAEP screener.**

  This would entail new development work.

- **c) student performance near the top or bottom percentile rank of the state’s previously administered state assessment.**

  While several panel members were hesitant to use results of varying state assessments, existing research shows that even the widely different tests used by states produce scores that correlate well enough with NAEP to be useful in identifying top and bottom performers who would be assigned high or low blocks of items.
d) a new or different method that may emerge, which is psychometrically sound and easy to administer.

The panel wants to see the adoption of a method that is fair, feasible, objective and effective, but recognizes that considerable technical development would be required before targeted testing can become a regular part of NAEP.

The panel recommends that the assignment of a targeted test to a student be based on how the student performs on some standard indicator of achievement (such as a test), and NOT upon a student’s label, such as having a disability or being in advanced placement classes. The panel intends that the availability of the easy form of the test assure participating schools that low-performing students, including students with disabilities, are able to participate without altering NAEP standards. Likewise, high-performing students could be challenged on items in the assessment at the greater level of difficulty.

6. Encourage and review research on the identification and progress of students who have a significant cognitive disability but in the short term do not test this 1% of students on NAEP.

The Panel recommends that NAGB form a panel of experts and stakeholders to review research and best current practices for identifying, measuring and reporting the progress of students who have a significant cognitive disability, and to make recommendations to NAGB for how emerging findings can and should be applied to NAEP in the future so such students could be included in NAEP.

The panel believes that NAEP should encourage the appropriate assessment of all children, but recommends that for the near future students with a severe cognitive disability—about 1% of the student population—be excluded from NAEP. The exclusion of these students should not be considered in determining whether a jurisdiction meets participation rate guidelines.

7. Assess the English language proficiency of students with disabilities drawn for the NAEP sample and provide NAEP-approved, linguistically appropriate accommodations for them before determining whether additional accommodations may be needed to address any disabilities these students may have.

Some students drawn for the NAEP sample will be both English language learners and students with disabilities. For these students it is important first to determine the level of their English proficiency, and the accommodations allowed for them on NAEP. If these students have also been identified as having a disability and are eligible to receive special education services, they should receive whatever accommodations are allowed by NAEP that they need to participate in the NAEP assessment.
APPENDIX A

LIST OF MEMBERS AND AFFILIATIONS

Technical Advisory Panel on Uniform National Rules for NAEP Testing of Students with Disabilities

- Alexa Posny, Kansas Commissioner of Education (Chair)
  Former Director, Office of Special Education Programs
  U.S. Department of Education

- George Engelhard, Jr.
  Professor of Educational Studies (Educational Measurement and Policy)
  Emory University, Atlanta, GA

- Louis Danielson, Managing Director, American Institutes for Research
  Former Director, Research to Practice Division, Office of Special Education Programs, U.S. Department of Education

- Miriam Freedman, attorney and author
  Stoneman, Chandler & Miller, Boston, MA

- Claire Greer, Consultant for Autism, Severe, and Multiple Disabilities
  Exceptional Children Division
  North Carolina Department of Public Instruction

- Robert Linn, Professor of Education (Emeritus)
  Research and Evaluation Methods Program
  University of Colorado

- Debra Paulson
  Middle school math and special education teacher
  El Paso, TX.

- Martha Thurlow, Director
  National Center on Educational Outcomes
  University of Minnesota
# APPENDIX B

## ACCOMMODATIONS ALLOWED ON NAEP

<table>
<thead>
<tr>
<th>On state assessment (this student):</th>
<th>Accommodations allowed on NAEP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reading</td>
</tr>
<tr>
<td>Presentation Format</td>
<td></td>
</tr>
<tr>
<td>Has directions read aloud/repeated in English or receives assistance to understand directions</td>
<td>O</td>
</tr>
<tr>
<td>Has directions only signed</td>
<td>O</td>
</tr>
<tr>
<td>Has test items signed</td>
<td>O</td>
</tr>
<tr>
<td>Has occasional words or phrases read aloud</td>
<td>O</td>
</tr>
<tr>
<td>Has all or most of the test materials read aloud</td>
<td>O</td>
</tr>
<tr>
<td>Uses a Braille version of the test</td>
<td>O</td>
</tr>
<tr>
<td>Uses a large print version of the test</td>
<td>O</td>
</tr>
<tr>
<td>Uses magnifying equipment</td>
<td>O</td>
</tr>
<tr>
<td>Response Format</td>
<td></td>
</tr>
<tr>
<td>Responds in sign language</td>
<td>O</td>
</tr>
<tr>
<td>Uses a Braille typewriter to respond</td>
<td>O</td>
</tr>
<tr>
<td>Points to answers or responds orally to a scribe</td>
<td>O</td>
</tr>
<tr>
<td>Tape records answers</td>
<td>O</td>
</tr>
<tr>
<td>Uses a computer or typewriter to respond</td>
<td>O</td>
</tr>
<tr>
<td>Spell/grammar check not allowed</td>
<td></td>
</tr>
<tr>
<td>Uses a template to respond</td>
<td>O</td>
</tr>
<tr>
<td>Uses a large marking pen or special writing tool</td>
<td>O</td>
</tr>
<tr>
<td>Writes directly in the test booklet</td>
<td>O</td>
</tr>
<tr>
<td>Setting Format</td>
<td></td>
</tr>
<tr>
<td>Takes the test in a small group*</td>
<td>O</td>
</tr>
<tr>
<td>Takes the test one-on-one*</td>
<td>O</td>
</tr>
<tr>
<td>Takes the test in a study carrel</td>
<td>O</td>
</tr>
<tr>
<td>Receives preferred seating, special lighting, or furniture</td>
<td>O</td>
</tr>
<tr>
<td>Must have test administered by familiar person</td>
<td>O</td>
</tr>
<tr>
<td>Timing Accommodations (Note: NAEP takes only 90 minutes.)</td>
<td></td>
</tr>
<tr>
<td>Receives extended time*</td>
<td>O</td>
</tr>
<tr>
<td>Is given breaks during the test</td>
<td>O</td>
</tr>
<tr>
<td>Must be allowed to take subject test over several days</td>
<td>O</td>
</tr>
<tr>
<td>Other Accommodations</td>
<td></td>
</tr>
<tr>
<td>Uses a calculator, including taking or Braille calculator for computation tasks</td>
<td>O</td>
</tr>
<tr>
<td>Uses an abacus, arithmetic tables, graph paper</td>
<td>O</td>
</tr>
<tr>
<td>Uses dictionary, thesaurus, or spelling/grammar-checking software or devices</td>
<td>O</td>
</tr>
<tr>
<td>Receives the following accommodation(s) not listed above.</td>
<td>O</td>
</tr>
</tbody>
</table>
Fed Agencies Spar Over NAEP for Special Populations

At issue is how many ELLs, Spec. Ed. students to test

By Nirvi Shah

Despite a pending policy change aimed at including more students with disabilities and English-language learners in the "nation's report card," the federal agency that administers the national testing program appears to be softening the penalty for states that fail to improve inclusion rates.

The disagreement underscores the uneasy relationship between the National Center for Education Statistics, the federal agency that administers the national tests, and the National Assessment Governing Board, the independent body that sets policy for the exams. And it reflects an intensifying debate about how to ensure that the National Assessment of Educational Progress, a congressionally mandated set of tests designed to take the national pulse on student achievement, accurately allows for state-by-state comparisons of student achievement.

"These issues, as all issues with students with disabilities and English-language learners, are hot potatoes," said Cornelia Orr, the governing board’s executive director.

Two years ago, NAGB adopted a policy that takes effect in January, during the next administration of NAEP, to limit how many students with disabilities and English-learners states can be cut from the testing pool. The policy says, essentially, that only students with severe cognitive disabilities and English-language learners who have been in the country for less than one year should be excluded from taking the exams in reading, mathematics, and other subjects.

Nationwide, some 830,000 4th and 8th graders from nearly 18,000 schools will take the tests in reading and math next year.

"The impetus for the NAEP policy was to push states to smooth out those state exclusion rates, to have the same proportion of students being tested across states," Ms. Orr said.

As written, that policy would help make NAEP scores more comparable from state to state. As it now stands, states that exclude more students with disabilities and ELLs have a record of posting better scores than states that are more inclusive.

Case in Point

For example, in 2011, of 4th grade students with disabilities in the testing pool, Maryland included less than a third—31 percent—on the reading test. Other states included as many as 90 percent or more of

Grade 4 Reading

If a new policy about including more students on the National Assessment of Educational Progress had been in effect in 2011, fewer students with disabilities and English-language learners would have been excluded from taking the exam. Federal statisticians estimate that the resulting changes in some states’ exclusion rates would have led, in turn, to lower scores on the 4th grade reading exam that year.
those students, and the size of the testing pool—2,500 to 3,000 students—is the same in each state. Maryland posted among the highest 4th grade reading scores in the country that year, and it was one of the few states to improve its scores from previous years.

The discrepancies from state to state over which students are tested—and which are not—have been especially frustrating for states that have been more inclusive but have found their NAEP scores stagnating.

Florida's commissioner of education, Gerard Robinson, wrote to NAGB earlier this year, saying the board should consider a policy of only reporting or using state-level results if the minimum standards of inclusion are met.

NAGB's new policy says that the proportion of all students excluded from NAEP should not be more than 5 percent and that states should push to include 85 percent of all students with disabilities and ELLs identified to be part of the testing pool.

"This would ensure the validity of the reported results for the nation and for the participating states," wrote Mr. Robinson, whose state is among those with lower exclusion rates. "States not meeting the minimum standards should face funding sanctions."

From the beginning, the NCES, a branch of the U.S. Department of Education, disagreed with the policy, although the agency agreed with the greater goal of inclusion. ("NAEP Board Curbs Special Ed. and ELL Exclusions," March 17, 2010.) At the time it was adopted, Stuart Kerachsky, then acting commissioner of NCES, said that the statistics agency harbored concerns about "flagging" individual states' exclusion rates.

**Reason for Disagreeing**

"There is no statistical basis for such standards," he wrote in a letter just days before the policy was adopted. "For that reason alone, NCES is unable to support this recommendation: We would be implicitly impugning jurisdiction results... without cause."

By law, the NCES is required to implement NAGB policy but, as this episode demonstrates, it has some degree of discretion to do so as it sees fit.

As created, the NAGB policy envisions dinging states that continue to exclude students with disabilities and ELLs from the testing pool when scores were tabulated.

The penalty would operate this way: Under the technical rules that guide NAEP, the federal agency is directed to impute, or estimate, the scores of such excluded students. In other words, if students with disabilities are excluded, their scores would still count in the calculation, using the average scores of other students with disabilities who were tested.

"Since students with disabilities tend to score lower on average than other students, disabled students ... would receive the same scores as similar disabled students, thus lowering the average," said Peggy Carr, the NCES' associate commissioner in the assessment division.

So the NCES is not planning to enact that penalty, she said.

But the NCES' plans are "contrary to the NAGB policy," said Lawrence Feinberg, the governing board's assistant director for reporting and analysis. "There's no question about that."
Advocates Object

With the threat of lower scores removed, any pressure on states to be more inclusive of special education students and English-language learners evaporates, say advocates for those groups.

"We want the sample to be more exemplary of students" with disabilities, said Laura Kaloi, the public-policy director for the National Center for Learning Disabilities, in New York City. "Why are schools more focused on excluding students that they don't believe can pass than [on] looking at why so many can't pass a grade-level exam?"

One complicating wrinkle in that debate is that NAEP doesn't allow all of the same accommodations for students with disabilities or students learning English on its tests that states typically permit. Some states, for example, allow portions of their state exams, including the reading sections, to be read aloud as designated in a students' individualized education program, or IEP. But NAEP doesn't. However, NAGB wants most students with disabilities to take the exam even if there is an accommodation they are accustomed to but cannot use on the national assessment.

That's partly why so many students with disabilities in Maryland have historically not taken NAEP, said Mary Gable, the assistant state superintendent for academic policy. Schools have a legal responsibility to carry out students' IEPS, Ms. Gable said. She believes the state would be violating federal law if students whose plans say they are entitled to the read-aloud accommodation had to take NAEP without it.

There's a similar issue in Kentucky, which also has high exclusion rates.

No Stakes

Mr. Feinberg said NAGB's understanding is that students could take NAEP even without every accommodation their education plans require, especially because the tests have no stakes for any individual student, such as determining whether students should be promoted to the next grade, and no records are kept about which students were tested.

Beyond the read-aloud issue, nearly all other accommodations are allowed on NAEP, such as additional time for testing, one-on-one testing, small-group testing, bilingual Spanish-English test booklets for subjects other than reading and writing, additional breaks, and having directions read in sign language.

Including more students with disabilities on the math test may be less of an issue. NAEP only allows calculators on some portions of math, but some special education students are entitled to calculators any time they are working on that subject.

To encourage their participation, Ms. Carr said, those students will be assigned the portion of NAEP that allows calculators.

*Assistant Editor Stephen Sawchuk contributed to this report.*
### GRADE 4 READING

If a new policy about including more students on the National Assessment of Educational Progress had been in effect in 2011, fewer students with disabilities and English-language learners would have been excluded from taking the exam. Federal statisticians estimate that the resulting changes in some states' exclusion rates would have led, in turn, to lower scores on the 4th grade reading exam that year.

<table>
<thead>
<tr>
<th>State</th>
<th>2011 Exclusion Rate</th>
<th>Hypothetical Exclusion Rate</th>
<th>2011 Mean Score</th>
<th>Hypothetical Mean Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delaware</td>
<td>7.0%</td>
<td>1.2%</td>
<td>225.1</td>
<td>223.1</td>
</tr>
<tr>
<td>Kentucky</td>
<td>8.7</td>
<td>1.3</td>
<td>225.1</td>
<td>223.6</td>
</tr>
<tr>
<td>Maryland</td>
<td>10.3</td>
<td>0.6</td>
<td>230.8</td>
<td>228.3</td>
</tr>
<tr>
<td>New Jersey</td>
<td>9.1</td>
<td>0.5</td>
<td>231.2</td>
<td>228.8</td>
</tr>
<tr>
<td>Tennessee</td>
<td>7.1</td>
<td>1.5</td>
<td>214.6</td>
<td>212.6</td>
</tr>
<tr>
<td>Texas</td>
<td>9.9</td>
<td>1.6</td>
<td>218.3</td>
<td>216.3</td>
</tr>
</tbody>
</table>

* Students with disabilities and English-language learners combined

**SOURCE:** National Center for Education Statistics
## National Assessment Governing Board
### Nominations Committee

**December 1, 2012**
**7:15 – 8:15 am**

**AGENDA**

### Closed Session 7:15 – 8:15 am

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
</table>
| 7:15 – 7:20 am | Welcome, Introductions, and Agenda Overview  
**Tonya Miles, Chair** |
| 7:20 – 8:15 am | Discussion of 2013 Nominees  
**Committee Members** |
# National Assessment Governing Board

**Ad Hoc Committee on NAEP Background Information**

November 29, 2012  
2:00 p.m.-4:00 p.m.

## AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Speaker(s)</th>
<th>Attachment</th>
</tr>
</thead>
</table>
| 2:00 – 2:15 pm | Welcome and Introductions  
Overview of Committee Work  
*Terry Holliday, Committee Chair* |                                                                 | Attachment A |
| 2:15 – 2:50 pm | Measuring and Reporting on Socio-Economic Status  
*William Ward, NCES*  
*Charles Cowan, Analytic Focus, LLC* |                                                                 | Attachment B |
| 2:50 – 3:20 pm | Exploratory Analyses of NAEP Data  
*Alan Ginsburg, Consultant* |                                                                 | Attachment C  
To be sent separately |
| 3:20 – 3:50 pm | Implementation of Board Policy on NAEP Background Questions and Use of Contextual Data in Reporting  
*James Deaton, NCES* |                                                                 |            |
| 3:50 – 4:00 pm | Next Steps  
*Committee Members* |                                                                 | Attachment D |
Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting

INTRODUCTION

By statute, the purpose of the National Assessment of Educational Progress is to provide a “fair and accurate” measure of student achievement and achievement trends. Academic or cognitive questions are its primary focus; the American public is its primary audience. However, in addition to reporting on what American students know and can do, NAEP has collected data for more than 40 years that provide a context for reporting and interpreting achievement results. According to the statute, such factors, both in and out of school, must be “directly related to the appraisal of academic achievement.”

In each assessment NAEP administers background questionnaires for students, their teachers, and schools. The questionnaires deal with educational experiences and other factors, such as teacher training or out-of-school learning activities, that are related to academic achievement. Data on several hundred background or noncognitive variables are available on the Internet through the NAEP Data Explorer. However, for more than a decade, little use has been made of this information in NAEP reports. The data have received minimal attention and had little impact despite the considerable efforts expended in developing and approving questionnaires and collecting and tabulating responses.

In October 2011 the National Assessment Governing Board convened an expert panel to recommend how to make better use of existing NAEP background questions and to propose an analytic agenda for additional topics and questions that would be useful in developing education policy and of value to the public. The panel report, entitled, NAEP Background Questions: An Underused National Resource, was presented to the Board in March 2012 by Marshall Smith, former U.S. Under Secretary of Education, who chaired the six-member panel.

Many of the panel recommendations build on the Background Information Framework for the National Assessment of Educational Progress, adopted by the Governing Board after it received final authority from Congress over non-cognitive items on the assessment. The framework was adopted in 2003, but has not been fully implemented.
The following policies are based on recommendations by the expert panel. The Board has also taken into consideration a wide range of public comment and the analysis provided by the National Center for Education Statistics.

It is important to understand that the National Assessment is not designed to show cause-and-effect relationships. Its data should not be used to “prove” what schools should do. But, as the Background Information Framework declares, NAEP’s “descriptions of the educational circumstances of students…, considered in light of research from other sources, may provide important information for public discussion and policy action.” The Board believes the National Assessment should improve upon its efforts to collect contextual information and present it clearly to the public, which will add to NAEP’s value to the nation.

POLICY PRINCIPLES

1. NAEP reporting should be enriched by greater use of contextual data derived from background or non-cognitive questions asked of students, teachers, and schools. Such data will be used both in regular Report Cards and in special focused reports.

2. Reporting of background data will describe patterns and trends, including the educational experiences of different groups of students. Care should be taken not to suggest causation.

3. Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions in background questionnaires and their connection to student achievement. Such frameworks should be updated for each assessment cycle and provide the basis for new topics and questions.

4. An ad hoc committee of the Board will be established for one year to monitor implementation of this resolution, review the NAEP Background Information Framework, and recommend a permanent arrangement for Board consideration of background questions and the reporting of contextual data in NAEP.

IMPLEMENTATION GUIDELINES

For Questions and Questionnaires

1. Clusters of questions will be developed on important topics of continuing interest, such as student motivation and control over the environment, use of technology, and out-of-school learning, which could be used regularly or rotated across assessment cycles.
2. Modules will be prepared for special one-time studies to provide descriptive information on issues of current policy interest.

3. A thorough review will be conducted to eliminate duplicative or low-priority questions. Unproductive topics and questions will be dropped.

4. NAEP will include background questions from international assessments, such as PISA and TIMSS, to obtain direct comparisons of states and TUDA districts to educational practices in other countries.

5. Because of the value of preserving trends, consistent wording of questions should be maintained on topics of continuing interest. Changes in wording must be justified. However, as practices and circumstances change, new questions will be introduced in a timely manner to gather data on topics of current interest.

6. The development and use of improved measures of socio-economic status (SES) will be accelerated, including further exploration of an SES index for NAEP reporting.

**For Data Collection**

7. The maximum time for students to answer the background questionnaire will be increased from 10 to 15 minutes on new computer-based assessments. Consideration should be given to a similar increase in paper-and-pencil assessments.

8. Whenever feasible, assessment samples should be divided (spiral sampling) and background questions rotated in different years in order to cover more topics without increasing respondent burden. These practices will be initiated in the assessments of reading and mathematics, which are conducted frequently, and considered for other subject areas if the frequency of testing permits.

**For Reporting**

9. Special focused reports with data through the 2013 assessment will be issued on the following topics: private schools, charter schools, gender gaps, and black male students. Reports shall include significant contextual information as well as cognitive results. Advisory committees, composed of a range of knowledgeable persons, may be appointed to provide input on reporting issues.

10. Exploratory analyses will be carried out to determine if existing background questions may form the basis for additional focused reports. Such reports may be issued by the Governing Board as well as by the National Center for Education Statistics.
11. The NAEP Data Explorer should be further improved to make data more accessible to general, non-specialist users. Tables and very simple-to-construct charts will be prepared to present data on important topics of wide public interest. Additional means of disseminating information, using new technology such as simple apps that would allow parents, teachers, and others to access background and achievement data, will be explored.
National Assessment Governing Board

Ad Hoc Committee on NAEP Background Information

BACKGROUND AND PURPOSE

As part of the resolution on NAEP background questions, adopted by the Governing Board in August 2012, an Ad Hoc Committee on NAEP Background Information is to be established for one year. Its purposes are as follows:


2. Review the *NAEP Background Information Framework*, adopted August 1, 2003. Recommend revisions, additions, or replacement, as deemed necessary or desirable.

3. Recommend a permanent arrangement for Board consideration of background questions and the reporting of contextual data in NAEP. This work is now divided between the Reporting and Dissemination and Assessment Development committees of the Board.

SPECIFIC ISSUES

As listed in the resolution, these include:

- Making greater use of contextual data in NAEP Report Cards and focused reports.
- Using background data to describe patterns and trends, including the educational experiences of different student groups.
- Detailed frameworks to support the selection of non-cognitive topics and questions, including their connection to student achievement.
- Clusters of questions on topics of continuing interest, such as technology and out-of-school learning, to be used regularly or rotated across cycles.
- Modules on issues of current policy interest.
- Elimination of duplicative, low-priority, or unproductive topics and questions.
- Use of questions from international assessments, such as TIMSS and PISA.
- Improved measures of socio-economic status (SES), including exploration of SES index.
- Spiral sampling and rotation of background questions in different years.
- Increasing the maximum time for students to answer background questions.
- Additional focused reports with the appointment of advisory committees in some cases.
- Exploratory analyses of existing data that may form the basis for subsequent reports.
- Consistency of wording to preserve trends.
- Further improvements in the NAEP Data Explorer.
COMMITTEE COMPOSITION AND TIMELINE

The Ad Hoc Committee will include six or seven Board members with a variety of perspectives and membership in different standing committees of the Board. The Committee will convene during each quarterly meeting of the Governing Board, and is expected to make its final report in August 2013.
Background Information

Framework for the

National Assessment

of Educational Progress

EXCERPTS

National Assessment Governing Board
Adopted August 1, 2003
National Assessment Governing Board

Darvin M. Winick  
Chair

John H. Stevens  
Chair, Ad Hoc Committee on Background Questions

Charles E. Smith  
Executive Director

Lawrence Feinberg  
Project Officer

Background Framework Project

Mary Lyn Bourque, Writer/Editor  
Mid-Atlantic Psychometric Services

David Grissmer  
RAND

Consultants

Paul E. Barton  
Ina V.S. Mullis  
Michael Podgursky  
Mark Reckase  
Herbert J. Walberg
Preface

by the National Assessment Governing Board

The National Assessment of Educational Progress (NAEP) has been established by law to monitor the academic achievement of American students. In addition to its academic assessments, NAEP has collected information from hundreds of non-cognitive or background questions about students, their educational experiences in class and at home, their teachers, and their schools. Some of these questions provide data for NAEP’s reporting categories, but far more have been used to give context to NAEP results or to track factors associated with academic achievement. Some have been used by scholars in social science research.

Concerns have been raised about the selection of background variables, the quality of the information obtained, and the validity of inferences drawn from it. There is also concern about the burden that collecting background information places on respondents and on the NAEP program. After the National Assessment Governing Board was granted final authority over the background questions in early 2002, it adopted a policy to focus NAEP background data on the primary purpose of the National Assessment—to provide sound, timely information on the academic achievement of American students. The Board also initiated a process to prepare a general framework to guide the collection and reporting of background data.

It is important to understand the National Assessment is not designed to prove cause-and-effect relationships; it cannot prescribe what should be done. But its descriptions of the educational circumstances of students at various achievement levels—considered in light of research from other sources—may provide important information for public discussion and policy action.

This framework will define the purpose and scope of NAEP’s system of collecting background information, including background questionnaires and other sources of non-cognitive data. It will establish criteria for reporting background information as part of the National Assessment. The approach it suggests provides for asking various groups of questions to various samples of students at various times.

The framework reflects the following key principles:
• The selection of background topics and questions shall be designed to fulfill all legal requirements for the National Assessment and to carry out decisions regarding what NAEP will report and how to report it.

• Background information shall provide a context for reporting and interpreting achievement results and, as the statute provides, must be “directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information.”

• The collection of background data shall be designed to obtain information that is objective, valid, reliable, and of consistently high quality.

• The system of background data collection shall be efficient and designed to minimize the burden on respondents and on the NAEP program. As much data as possible should be obtained from school records and other reliable data sources.

• These principles shall apply both to the collection of general background information and to subject-specific background questions. The frameworks for the latter must be focused and prioritized, indicating a core set of variables for regular reporting and a more comprehensive set to be collected and reported less frequently.

• The priority order for background information is as follows: (1) reporting categories, as required by law; (2) contextual factors with a well-established relationship to achievement; and (3) subject-specific information.

There is one other consideration—the new role of the National Assessment in the No Child Left Behind Act of 2001. Under this law, all states receiving federal Title I aid are required to participate every two years in NAEP’s state-level samples of reading and mathematics in grades 4 and 8. The results will provide an independent yardstick to compare trends on NAEP with performance on each state’s own set of required exams.

Because No Child Left Behind places particular emphasis on closing the persistent performance gaps between various student groups, NAEP must be able to report on changes in achievement for all groups specified by law. Through its background questions, the National Assessment might also provide useful information about the students left behind and those who are ahead of them, including the sorts of schools that high-achieving and low-achieving students attend, the courses they take, the patterns of
how they are taught, and the qualifications of their teachers. Over time, such descriptive information will allow NAEP to track changes in contextual and instructional factors related to student achievement and in the distribution of important educational resources.

In sum, the purpose of this Background Information Framework is to focus the collection and reporting of background data by the National Assessment and to establish clear priorities and limits. We hope to make it possible that with far fewer non-cognitive questions than it has had in the recent past, NAEP will serve the purposes of law and provide the American public and decision makers with useful information. We are committed to improving the quality of data collected and the reporting of results.
Executive Summary

The National Assessment of Educational Progress (NAEP) is a federally authorized survey of student achievement at grades 4, 8, and 12 in various subject areas, such as mathematics, reading, writing, science, U.S. history, the arts, and foreign languages. The No Child Left Behind Act of 2001 (P.L. 107-110) requires the assessment to collect data on specified student groups, including race/ethnicity, gender, socio-economic status, disability, and limited English proficiency. It requires fair and accurate presentation of achievement data and permits the collection of background or descriptive information that is related to academic achievement and aids in fair reporting of results. The intent of the law is to provide representative-sample data on student achievement for the nation, the states, and subpopulations of students and to monitor progress over time.

The National Assessment Governing Board (NAGB) sets policy for NAEP and determines the content framework for each assessment. As a result of the No Child Left Behind Act, the Board is responsible for selecting and approving all of NAEP’s non-cognitive or background questions, as well as the cognitive items over which it has had final authority since 1988. This Background Information Framework will guide the development and selection of non-cognitive topics and questions, starting with the NAEP 2006 assessment. It will fulfill the purposes of law and implement Board policy.

When NAEP began in 1969-70, its background information was limited to gender, race/ethnicity, and literacy materials at home. During the 1980s the array of non-cognitive questions expanded greatly, both to provide more contextual information and in an effort—never fully realized—to use the assessment for educational research.

This background data framework will refocus the collection of non-cognitive variables on NAEP’s primary mission: providing a fair and accurate measure of student achievement and on achievement trends over time. Thus, the framework is a guide for gathering important information that will assist in reporting and understanding NAEP results. NAEP may contribute to research into improving education policy and practice, but its role in this respect is limited and the framework is not a comprehensive list of possible factors to explore.
Since by law NAEP may only collect information that is “directly related to the appraisal of academic achievement,” it must concentrate on non-cognitive variables that are known from other research to have such a relationship. The law also specifically prohibits NAEP from asking about personal or family beliefs and attitudes. These points are emphasized in the Governing Board Policy Statement on the Collection and Reporting of Background Data by the National Assessment (adopted on May 18, 2002). That policy is incorporated into this framework. It is attached in the appendix.

PRIORITIES

The following priorities for collecting and reporting non-cognitive information should be followed in planning background questionnaires, the frequency with which questions are asked, and the samples from which data are collected.

1. **Student reporting categories** that are required by law must be collected as a regular component of all NAEP assessments. These include race, ethnicity, gender, socio-economic status, disability, and limited English proficiency. A core of SES information should be collected in every assessment, such as type of community and poverty status. An expanded set of SES variables may be included periodically or administered to limited samples.

2. **Other factors that provide a context for results** should be sampled periodically, or on a rotating basis, over several NAEP cycles, although a limited set may be asked in every assessment. Contextual factors may include courses taken, student mobility, school safety and discipline, teacher-related factors such as demographics and experience, other factors related to students and schools, and educationally-relevant variables outside school. Although many non-cognitive variables may be of interest, they must be limited to meet the needs of NAEP reporting. In all cases, they must be clearly related to academic achievement or to the fair presentation of achievement results.

3. **Subject-specific background information** should be gathered at the same time that achievement in a subject is assessed. This may include relevant course content and requirements, teacher preparation, and other factors related to student achievement. Questions will not be designed to determine effective practices, but to show patterns and trends of factors of interest, based on previous research. Like the contextual information, most of these variables should be sampled periodically, or on a rotating basis, over several administrations of the subject exam, although a limited core set may be repeated every time the assessment is given.
SELECTION CRITERIA

Key criteria for selecting non-cognitive topics and questions are as follows:

- **Does the current or proposed non-cognitive variable relate to the primary purpose of NAEP and how?** The primary purpose of NAEP is to report on the academic achievement of students to the American public. It is not to report on the causes of that achievement. Other surveys with longitudinal data are far better suited to examining causality. NAEP’s choice of which non-cognitive variables to measure should be guided by how and to what extent the variables selected will support NAEP’s primary mission.

- **Do the current or proposed non-cognitive variables meet professional standards for reliability and validity?** The NAEP legislation requires that the assessment “use widely accepted professional testing standards (P.L. 107-110, Sec. 411 (b) (5).” This requirement applies equally to non-cognitive and academic variables.

- **How stable is the non-cognitive variable from period to period?** If a variable shows little change from year to year, it should be reviewed to determine whether it should be deleted or used on a periodic basis rather than in every assessment.

- **If new questions are added, have others been deleted in order to limit the burden and expense of NAEP’s background questionnaires?** There will always be pressure to collect more information. Mechanisms must be developed to make sure the burden of background questionnaires does not expand over time.

- **Does a question address specific behavior rather than conclusions?** Even for such questions, however, caution is advisable because self-reports are often unreliable.

- **Will the topic or question meet the test of broad public acceptability and not be viewed as intrusive or prying?** NAEP’s non-cognitive questions are not kept secure, and all of them are to be posted on the Internet. Possible objections should be considered in deciding whether or not a question will be asked.

- **Does the topic or question deal with a factor in which trends over time are important?**
DATA COLLECTION

Whenever possible, NAEP should use information from school records and other reliable data collections in order to improve the validity of the information collected and limit the background questionnaires in NAEP itself. In exploring the utility of different data sources, the following criteria should be considered: (1) reliability, (2) universality, (3) currency, (4) respondent burden, (5) logistics, (6) efficiency and cost-effectiveness, and (7) the impact on timeliness of NAEP reporting.

Of the student reporting categories in Priority 1, information on gender, race/ethnicity, disability status, and limited English proficiency shall be collected in a uniform manner in all NAEP samples. NAEP is also required to collect information on socio-economic status. This will continue to be done in all samples, although there may be some variation in the number of factors on which data are obtained with a uniform core and more extensive data gathering in some cases.

Because socio-economic status cannot be measured simply or directly, NAEP has used “proxy” variables, such as eligibility for free or reduced-price lunch (a measure of poverty), parent education, and number of reading materials in the home. The framework provides that NAEP explore development of a composite index for SES derived from the proxy variables currently collected. To the extent that the index can be sharpened by additional data from readily available sources, such as zip codes and census, this option should also be considered. Occasionally and in limited samples, more extensive SES questions may be asked. Although NAEP may never be able to produce a full composite of SES, based on family income, education, and occupation, efforts should be made to find an approximation that is more informative than the current set of proxy variables.

For the past two decades, NAEP has collected information on a lengthy list of student, teacher, school, and beyond-school factors that may provide a context for achievement results and are of interest to policymakers, researchers, and the public. Yet, NAEP’s design as a cross-sectional survey places serious limitations on the inferences that can properly be drawn from this information. We propose a careful review of the contextual factors in NAEP to focus on the most important variables related to public policy. All such information must be clearly related to student achievement, as shown by other research. Different questions should be cycled in and out of the assessment
periodically, and the use of data from non-NAEP sources should increase. Information should be collected at meaningful intervals in ways that may show significant patterns and change over time.

The collection of subject-specific background information should be focused, limited, and prioritized as part of the subject-matter frameworks adopted by the Board. For each subject there should be a small core set of background items administered to the full sample each time a subject is assessed. An additional, more comprehensive set of questions should be administered periodically or to smaller subsamples.

NCES will prepare for Board review and approval a plan indicating the frequency, sample size, and schedule of rotation for all background variables and questions on which information is to be collected by NAEP. This should include both questionnaires and alternate data sources to obtain core reporting data, subject-specific information, and data on achievement-related contextual variables from a variety of NAEP samples—national only, national and state, and a subset of the national sample. The plan should indicate the frequency and schedule of rotation for each of the questions proposed. It should also indicate any questions needed for quality control purposes. The recommendations should be prepared with input from researchers and state policy analysts, as appropriate, and updated on a regular basis.

In constructing questionnaires it is important to place strict limits on the respondent burden they impose. As much data as possible should be obtained from school records and other reliable data sources. The average individual response time to answer background questionnaires for each assessment, as calculated in accordance with Office of Management and Budget (OMB) procedures, shall be limited as follows: ten minutes for each student, 20 minutes for each teacher, and 30 minutes for each school.

**REPORTING**

NAEP reporting should include contextual variables and subject-specific background information to enrich and give perspective to results. Consistent with space and operational limitations, descriptive information should be part of NAEP Report Cards and summary and highlights reports. The reports should present information on patterns and trends of non-cognitive variables known to have a relationship to academic achievement and may contain disaggregated data on school conditions and practices for various groups of students. Data on courses taken before NAEP assessments (either from transcripts or questionnaires) is of great public interest and can be related to academic results.
In addition, supplemental reports may be prepared that focus on particular aspects of the background data collected. In all cases, NAEP reports published by the National Center for Education Statistics must not state conclusions as to cause and effect relationships and avoid simplistic presentations that imply best practice.

All background questions and data collected by NAEP should be posted on the Internet so the public may be able to consider them in discussing results. Complete data files should be made available to researchers for further analysis.

**RESEARCH**

As a cross-sectional survey without longitudinal data, the National Assessment is able to document school conditions and practices. It can report on achievement results. But it cannot properly be used to establish direct cause-and-effect relationships. Still, over the past three decades, NAEP has been part of two important research endeavors—exploring changes in the black-white test score gap since 1970 and seeking to establish the impact of state-level reforms during the 1990s. By monitoring achievement well, NAEP has provided sound data for researchers to use. NAEP results have been critical in identifying research hypotheses. Its large data sets have been combined with other information to tease out meaning and policy implications, though NAEP’s own reports have properly steered clear of these activities.

The Governing Board believes that by doing its main task of monitoring educational achievement well NAEP can make a valuable contribution to education research. The NCES program of secondary analysis grants for researchers to analyze NAEP data should continue. Educational researchers should be involved, under the auspices of NCES, in developing NAEP background questionnaires, validity studies, and other data collection efforts to carry out the provisions of this framework.

The primary purpose of NAEP is to provide fair and accurate information on student achievement. Its primary audience is the American public. The Governing Board believes that in serving its purpose and audience well, NAEP can contribute to educational research. It welcomes the interest and efforts of researchers.
A NEW MEASURE OF SOCIOECONOMIC STATUS (SES)
FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Submitted to the National Assessment Governing Board
by the National Center for Education Statistics

As part of the Background Information Framework for the National Assessment of Educational Progress, the Governing Board called for improvement in the measurement of socioeconomic status (SES) and its relationship to academic achievement as measured by NAEP. To date, NCES and its contractors have conducted several research initiatives and have enlisted the contributions of external experts to assist in accomplishing this goal.

On behalf of NCES, a panel of experts was convened and met three times in 2010 to 2012 to discuss and provide recommendations for a new SES, with a focus on theoretical aspects of SES measurement. Panel members were invited on the basis of expertise in the fields of socioeconomic measurement, education, statistics, poverty, economics, and sociology. The panel developed a consensus definition of SES, and proposed both a core SES measure, reflecting parental education, parental occupation, and family income, and an expanded measure reflecting in addition, school and neighborhood factors. The panel reviewed the history of SES measurement, suggested various approaches to measuring its components, and discussed issues and methods for combining SES components into a single composite for reporting.

The panel concluded with a series of six recommendations covering core and expanded definitions of SES, the importance of conducting research on explanatory variables associated with SES and on the use of subjective SES in educational applications, the value of a composite measure of SES, and the importance of conducting further research on linking to U.S. Census Bureau data to improve the measurement of SES. The panel summarized their deliberations and recommendations in a white paper.

The background of the initiative and the white paper will be presented at the meeting.
A NEW MEASURE OF SOCIOECONOMIC STATUS (SES)
FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Expert Panel Members

- Charles D. Cowan, Analytic Focus, LLC
- Robert M. Hauser, University of Wisconsin-Madison; National Research Council
- Robert Kominski, U.S. Census Bureau
- Henry M. Levin, Teachers College, Columbia University
- Samuel R. Lucas, University of California-Berkeley
- Stephen L. Morgan, Cornell University
- Margaret Beale Spencer, University of Chicago
- Christopher Chapman, National Center for Education Statistics (Ex officio)
IMPROVING THE MEASUREMENT OF SOCIOECONOMIC STATUS FOR THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS: A THEORETICAL FOUNDATION
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert Panel</td>
<td>3</td>
</tr>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>4</td>
</tr>
<tr>
<td>1. Setting the Stage</td>
<td>8</td>
</tr>
<tr>
<td>2. Background</td>
<td>10</td>
</tr>
<tr>
<td>Measuring SES</td>
<td>11</td>
</tr>
<tr>
<td>3. Objectives</td>
<td>13</td>
</tr>
<tr>
<td>Provide a Definition of SES</td>
<td>14</td>
</tr>
<tr>
<td>Identify Components of SES</td>
<td>14</td>
</tr>
<tr>
<td>Review Data Collection and Measurement Approaches</td>
<td>14</td>
</tr>
<tr>
<td>Create an SES Composite (or Justify Use of Multiple Single Variables)</td>
<td>14</td>
</tr>
<tr>
<td>Consider Implications of a New Measure of SES</td>
<td>15</td>
</tr>
<tr>
<td>4. Defining SES</td>
<td>15</td>
</tr>
<tr>
<td>Panel Recommendation: A Definition of SES</td>
<td>19</td>
</tr>
<tr>
<td>5. Components and Correlates of SES</td>
<td>19</td>
</tr>
<tr>
<td>The “Big 3”</td>
<td>19</td>
</tr>
<tr>
<td>Neighborhood SES</td>
<td>20</td>
</tr>
<tr>
<td>School SES</td>
<td>21</td>
</tr>
<tr>
<td>Psychological Process Variables</td>
<td>21</td>
</tr>
<tr>
<td>Subjective SES</td>
<td>22</td>
</tr>
<tr>
<td>Panel Recommendations: Identifying Components and Correlates of SES</td>
<td>23</td>
</tr>
<tr>
<td>6. Approaches to Measuring SES components</td>
<td>23</td>
</tr>
<tr>
<td>Family income</td>
<td>23</td>
</tr>
<tr>
<td>Other indirect measures of family income</td>
<td>24</td>
</tr>
<tr>
<td>Household composition</td>
<td>25</td>
</tr>
<tr>
<td>Parental educational attainment</td>
<td>26</td>
</tr>
<tr>
<td>Parental Occupational Status and Employment Status</td>
<td>26</td>
</tr>
<tr>
<td>Neighborhood SES</td>
<td>27</td>
</tr>
<tr>
<td>School SES</td>
<td>27</td>
</tr>
<tr>
<td>Panel Recommendation: Review Data Collection and Measurement Approaches</td>
<td>28</td>
</tr>
<tr>
<td>7. SES Composite</td>
<td>30</td>
</tr>
<tr>
<td>General model</td>
<td>31</td>
</tr>
<tr>
<td>Arbitrary weighting</td>
<td>32</td>
</tr>
<tr>
<td>Empirical weighting</td>
<td>33</td>
</tr>
<tr>
<td>Measurement invariance goals</td>
<td>35</td>
</tr>
<tr>
<td>Missing data issues</td>
<td>36</td>
</tr>
<tr>
<td>Panel Recommendation: Create an SES Composite</td>
<td>36</td>
</tr>
<tr>
<td>8. Implications</td>
<td>37</td>
</tr>
<tr>
<td>Reporting and Implications for Trend</td>
<td>37</td>
</tr>
<tr>
<td>Data conditioning</td>
<td>38</td>
</tr>
<tr>
<td>Use by other units, departments, agencies</td>
<td>39</td>
</tr>
<tr>
<td>Anticipated effects and unanticipated side effects</td>
<td>40</td>
</tr>
<tr>
<td>Panel Recommendation: Consider Implications of a New Measure of SES</td>
<td>40</td>
</tr>
<tr>
<td>9. Discussion</td>
<td>41</td>
</tr>
<tr>
<td>Key Recommendations</td>
<td>43</td>
</tr>
<tr>
<td>References</td>
<td>44</td>
</tr>
</tbody>
</table>
EXPERT PANEL

Charles D. Cowan
Partner and Chief Executive Officer
Analytic Focus, LLC

Robert M. Hauser
Vilas Research Professor Emeritus
Department of Sociology, University of Wisconsin-Madison

Executive Director
Division of Behavioral and Social Sciences and Education
National Research Council

Robert A. Kominski
Assistant Chief for Social Characteristics
Social, Economic, and Housing Statistics Division, U.S. Census Bureau

Henry M. Levin
Professor of Economics and Education
Teachers College, Columbia University

Samuel R. Lucas
Professor
Department of Sociology, University of California-Berkeley

Stephen L. Morgan
Jan Rock Zubrow ’77 Professor in the Social Sciences
Department of Sociology, Cornell University

Margaret Beale Spencer
Marshall Field IV Professor of Urban Education
Social Sciences Division, Comparative Human Development
University of Chicago

Chris Chapman
Acting Associate Commissioner
Early Childhood, International, and Crosscutting Studies Division
National Center for Education Statistics (Ex officio)
EXECUTIVE SUMMARY

At the request of the National Assessment Governing Board (NAGB, 2003), the National Center for Education Statistics (NCES) convened a panel of experts to provide recommendations concerning socioeconomic status (SES) as a construct, with the understanding that their recommendations might ultimately lead to a new measure of SES that could be used for the National Assessment of Educational Progress (NAEP). The current, most prominent NAEP measure of student SES, National School Lunch Program (NSLP) eligibility, has become less valid over time. The panel’s main focus was on the theoretical foundations of SES.

Objectives

Several objectives guided the panel’s recommendations:

- provide a definition of SES,
- identify components of SES,
- review data collection and measurement approaches,
- create an SES composite, and
- consider implications of a new measure of SES.

Defining SES

The panel developed the following consensus definition of SES:

*SES can be defined broadly as one’s access to financial, social, cultural, and human capital resources. Traditionally a student’s SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate adjustment for household or family composition. An expanded SES measure could include measures of additional household, neighborhood, and school resources.*

Components and Correlates of SES

The panel concluded that the components of a core student SES measure were the “big 3” variables (family income, parental educational attainment, and parental occupational status),
but also suggested that home neighborhood and school SES could be used to construct an expanded measure of SES. Identifying such variables and including them in an expanded SES composite could help improve the explanatory power of SES in accounting for NAEP scores. In addition, some psychological process variables (e.g., coping mechanisms, emotional control, or perceptions of the environment) and some subjective measures (i.e., how one views one’s SES), might be understood as useful contextual and potentially explanatory variables that could help interpret student NAEP scores. Although psychological process and subjective factors were not included as components of a core or expanded SES as developed by the panel, it is important that research be conducted to evaluate the effects of these factors on achievement.

**Approaches to Measuring SES Components**

In addition to current measures of family income, additional variables, such as housing tenure (rent or own), number of moves in the past year, presence of a household member needing healthcare assistance, and others, could be studied for potential use as indirect measures of family income. Parental educational attainment is currently measured through the NAEP questionnaire, but only for 8th- and 12th-graders, and parental occupational status – one of the big three variables – is not collected in the Student Questionnaire, nor is it available through school records. Cognitive laboratory studies should be conducted on various question types for collecting student reports on parental occupation.

There are currently no direct measures of neighborhood components of a possible expanded SES measure, although NAEP student questionnaire items and information from school records could be aggregated to serve as neighborhood measures. American Community Survey data could be used to provide much of the information not available through NAEP questionnaires and school records. The upcoming Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) (NCES, 2012a), which tests 4th-graders in 2014, represents an ideal opportunity to inform SES measurement.

**SES Composite**

There are reporting and interpretation advantages and disadvantages for treating SES as a single measured variable, as several single measured variables, or as a composite of several measured variables. The advantages of a composite variable over the use of single variables outweigh the disadvantages. There are a variety of schemes by which SES components could
be combined into a composite measure. A challenge in developing an SES composite is determining whether weights should vary depending on factors such as location or grade level. A review of the existing literature and data quality should be conducted before proposing a recommendation on a component weighting scheme.

**Implications**

A new SES measure will affect NAEP reporting, including whether and how to characterize SES levels, whether a bridge study must be conducted to link new and old measures of SES, and how a new SES measure will affect NAEP’s conditioning model. The research, framework, and findings associated with the development of a new SES measure could benefit other programs that measure SES, both within and outside NCES. For example, states are continually seeking better measures of SES. In addition, the development of a new SES measure is likely to incur both anticipated and unanticipated side effects, including the requirement to coordinate with other federal programs within and outside NCES, and consequences such as attention given to equity and educational resource distribution.

**Key Recommendations**

The panel made four key recommendations to improve measurement and reporting of SES:

1. Family income and other indicators of home possessions and resources, parental educational attainment, and parental occupational status should be considered components of a core SES measure, and should be the subject of immediate focus for NAEP reporting.

2. Neighborhood and school SES could be used to construct an expanded SES measure, and measures of these variables could contribute to an expanded SES.

3. Composite measures have many advantages, such as being a single summary useful for reporting, greater reliability, and representing the full range of SES factors. In addition, treating SES as a composite measure does not preclude reporting on relationships between individual SES components and achievement. Therefore, attempts should be made to develop an SES composite measure.

4. The validity of NSLP eligibility has been decreasing due to jurisdiction-wide eligibility and other factors, and that trend is likely to continue. Furthermore, there is concern over the quality
of student reports, particularly regarding parental educational attainment (for 4th-graders) and occupational status (for all grades). Due to these data quality issues, along with burden considerations, attempts should be made to explore the possibility of linking to Census data on SES components.
1. SETTING THE STAGE

The National Assessment of Educational Progress (NAEP) is a congressionally authorized project of the National Center for Education Statistics (NCES) within the Institute of Education Sciences of the U.S. Department of Education. The Commissioner of NCES is responsible for carrying out the NAEP project, while the National Assessment Governing Board (Governing Board) oversees and sets policy for NAEP. NAEP measures student progress over time in a variety of subject areas, including reading, writing, mathematics, science, and U.S. history. NAEP does not report individual student scores; rather, the assessment is designed to produce public-domain data about student achievement at the group level. Because NAEP results are meant to inform educators, policymakers, and the general public about the performance of students at the 4th, 8th, and 12th grade levels, reports include overall results as well as scores for student subgroups that are of interest to the target audiences, such as gender, race/ethnicity, and socioeconomic status (SES).

In response to a call by the Governing Board (NAGB, 2003) to improve the measurement and reporting of SES and its relationship to academic achievement in the context of NAEP, NCES convened a panel of experts in the fields of economics, education, statistics, human development, and sociology with substantive expertise in the effects of poverty and disadvantage on student achievement as well as methodological expertise in the measurement of socioeconomic standing. The panel was asked to provide recommendations concerning SES as a construct with the understanding that those recommendations might ultimately lead to a new SES measure that could be used in programs such as NAEP. The guidance was to focus on the theoretical aspects of SES measurement, not on operational aspects. Specifically, the panel was tasked with considering issues surrounding SES, including the creation of a composite measure of SES, how a new SES variable could be used in a reporting context, and how its derivation could be explained to both technical and general audiences. The panel met three times between 2010 and 2012.

This report reflects the discussions and recommendations of the panel and provides the context and background for those discussions. The report was prepared with key NCES stakeholders in mind, including the general public and education policymakers at both the state and national levels. The panel discussed the reporting of SES in NAEP and other large-scale assessments, such as the Program for International Student Assessment (PISA) and the Trends in
International Mathematics and Science Study (TIMSS), to learn from those assessments and also to seek to inform them.
2. BACKGROUND

There is a long history of SES being reported to correlate with educational achievement (Cuff, 1934; Holley, 1916; Lynd & Lynd, 1929). The Equality of Educational Opportunity Commission Report (Coleman et al., 1966) played a major role in bringing this finding to prominence in policy circles. Since then, measures of SES have been routinely included in educational research studies as background variables. Researchers and policy makers are interested in SES as a contextual variable to study educational equity and fairness issues, as a covariate with achievement to examine the effects of other variables such as class size or school governance policies, and as a matching variable to ensure the equivalence of treatment and control groups in educational intervention studies. NAEP treats SES as one of five background reporting variables (see law excerpt, below) and scores are reported separately for different SES subgroups. NAEP is mandated to report scores by SES by the No Child Left Behind Act of 2001 (P.L. 107-110, 2002), which acknowledges the importance of SES in educational achievement:

The Commissioner, in carrying out the measurement and reporting described in paragraph (1), shall—‘‘(G) include information on special groups, including, whenever feasible, information collected, cross tabulated, compared, and reported by race, ethnicity, socioeconomic status, gender, disability and limited English proficiency (Sec. 411. National Assessment of Educational Progress, Paragraph (b) Purpose; State Assessments; Subparagraph (2) Measurement and Reporting). (115 STAT. 1898)

However, the mandate does not provide specific guidelines on how SES is to be measured, nor even on how it is defined. Current NAEP practice is to measure SES through a set of proxy variables, most notably eligibility for the National School Lunch Program (NSLP; 2008), but also through school Title 1 status, parental educational attainment, and reading materials in the home. For reporting purposes, all of these are treated as individual variables, rather than as a composite SES variable.

It is instructive to review how socioeconomic status is treated in NAEP score reporting. Typically no mention is made of SES per se, but NAEP scores are reported by different variables that might be interpreted as SES measures. In the recent 2009 NAEP Science report (NCES, 2011), for example, for 4th- and 8th-grade students, NAEP scores were reported by eligibility for NSLP in three categories (not eligible, eligible for a reduced-price lunch, and
eligible for a free lunch). NSLP eligibility was reported to be “an indicator of low income” (p. 60). The Technical Notes section of the report states that scores were not reported by NSLP eligibility for 12th-grade students “because students’ eligibility for free or reduced-price school lunch may be underreported at grade 12.” (p. 60) (See discussion in the Measuring SES section, below.) These data are obtainable from the NAEP Data Explorer, however. For 8th-grade and 12th-grade students, NAEP scores were reported by parental educational attainment, which is widely regarded as a central component of SES.

A broad and widely accepted definition of SES in the scientific literature emphasizes its role in reflecting access to resources. Furthermore, students’ SES is traditionally defined as a combination of family income, parental occupational status, and parental educational attainment. Although the proxy variables currently used in NAEP reflect these factors to some extent, questions have been raised about the quality of the data, the narrowness of the measure, and the lack of a composite SES measure.

Consequently, there have been calls to explore alternative SES measures. Among the suggestions have been to create a composite measure rather than relying on single proxy measures (Barton, 2003), and to use data linked from other sources, such as the U.S. Census Bureau’s American Community Survey, which provides data on income, parental occupation, and parental educational attainment (Hauser & Andrew, 2007). The problems identified with current NAEP practice in measuring SES, along with conceptual and empirical developments in understanding SES, suggest that the time is right to consider alternatives in developing a new SES measure for NAEP.

**Measuring SES**

The history of SES measurement and the identification of possible explanatory correlates show that SES is defined as a broad construct, ideally measured with several diverse indicators. However, there are some advantages to using NSLP eligibility as an operational SES measure for NAEP. First, NSLP eligibility is available through school records for almost every student in the U.S., making data collection inexpensive and minimizing problems with missing data. In addition, NSLP status is a three-level categorical variable, which is convenient for reporting purposes and easily understood by a variety of audiences. Finally, NSLP eligibility status is also tied to federal definitions of poverty, which means that maintenance or updating is handled automatically through updating of federal poverty guidelines.
On the other hand, there are problems with using NSLP eligibility as the main measure of SES in NAEP reporting. These problems can be summarized as follows:

1. **NSLP eligibility measures only one SES component, family income** (adjusted for household composition). NSLP eligibility does not reflect parental educational attainment or occupational status.

2. **Due to the process of eligibility certification, NSLP eligibility may not be the most reliable measure of family income** (Harwell & LeBeau, 2010). Approximately 20 percent of students either are not eligible but are deemed eligible or are eligible but are not recognized as such (Food and Nutrition Service, 1990; Harwell & LeBeau, 2010; Hauser, 1994). The problem of eligible students failing to apply (whether due to social stigma or some other cause) increases with grade level, and is particularly prevalent for 12th-graders (Office of Research, Nutrition, and Analysis, 1994). Failure to apply when eligible is also thought to correlate with immigration status and to be more prevalent among students who speak English as a second language.

3. **Because there are only three levels of NSLP eligibility, there are large within-category SES differences, particularly in the non-eligible category.** Furthermore, the categories contain uneven shares of the distribution; there is approximately an 8:1 ratio of students in the free vs. reduced-price lunch categories.

4. **School-level and jurisdiction-level eligibility threatens the validity of NSLP eligibility as a measure of an individual student’s family income.** All students in a school with greater than 80 percent eligibility are categorized as NSLP eligible, regardless of their family income. Likewise, all students in some jurisdictions, such as Puerto Rico, and many of the urban districts are declared eligible regardless of family income levels.¹

The remainder of this paper is organized into seven additional chapters. **Objectives** reviews project goals, which are to articulate a definition of SES; identify SES components; address

---

¹ Regarding this last point, the most significant problem with the NSLP eligibility measure for the future is the introduction of Community Eligibility (Provision 4) through The Healthy, Hunger-Free Kids Act of 2010 (PL 111-296). Community Eligibility means that schools will no longer be required to keep individual student eligibility information once they have determined a baseline percentage of eligible students, which may result in missing or inaccurate individual student eligibility information. This change in eligibility certification is expected to be phased in, but would nevertheless affect the validity of the NSLP eligibility measure.
data collection issues, which should lead to a new SES composite; and consider implications of a new SES measure. *Defining SES* emphasizes a broad definition of SES as a student’s access to resources and reviews its expected relationship to achievement. *Components and Correlates* presents the idea of SES as a composite of “the big 3” variables—family income, parental educational attainment, and parental occupational status—and additional variables, most notably neighborhood and school SES. Also included here is a discussion of variables that could be considered as either components or correlates of SES, including subjective (perceived) SES, cultural capital, and other factors. The chapter also reviews variables that correlate with SES and variables (moderators) that interact with SES in its relationship with achievement. *Approaches to Measuring SES Components* reviews how the proposed components of SES can be measured. *SES Composite* reviews various ways in which SES components can be weighted and combined, and discusses issues with missing data. *Implications* focuses on the consequences of a new measure of SES for reporting, including anticipated and unanticipated side effects, and discusses possible uses of the new SES measure by other units, departments, and agencies. Finally, *Discussion* provides a general summary of the paper and concludes with key recommendations.

### 3. OBJECTIVES

The panel was to provide recommendations for a new SES measure for NAEP that would continue to meet the requirements of reporting SES and also improve the measurement and reporting of SES through the collection of higher quality data.

The primary purpose of proposing a new measure of SES was to meet the requirements of the No Child Left Behind Act of 2001 (P.L. 107-110, 2002) in the best feasible way. The law mandates the reporting of scores by SES in acknowledgement of the importance of SES in educational achievement. The law was not specific about how SES was to be measured. The primary objective was to have a panel provide expert guidance and interpretation on how the law’s mandates could be realized.

Specifically, the panel was charged with the following study objectives:
Provide a Definition of SES

Although it is possible to develop and use a measure of socioeconomic status without a clear definition of what it is—for example, basing it on measures that have been used or are currently used in different projects—there are many advantages to communicating a clear definition of SES.

Identify Components of SES

Historically, SES has been defined as some combination of family income, parental educational attainment, and parental occupational status. Other variables have also been considered as part of SES, including various school factors, community or neighborhood factors, and subjective measures of socioeconomic status, such as where individuals see themselves on a status ladder. An objective for this study was to identify which of the various components should be included as part of SES for NAEP reporting.

Review Data Collection and Measurement Approaches

Some SES measures, such as eligibility for the National School Lunch Program, have been obtained from school records. Others, such as parental educational attainment, have been obtained from the student questionnaire. A methodology that has been discussed, and experimented with, involves obtaining geographically aggregated Census data (aggregated over ZIP code tabulation areas, Census blocks, or Census tracts) to impute student family data, such as family income, household status (e.g., single vs. dual head of household), and parental occupations. Census data can be obtained either from the United States Census 2000 long form (to analyze previously-collected data for research purposes only), or from the ongoing American Community Survey 5-year estimates (which could be analyzed for both research purposes and operational use). An objective of this study was to review data collection and measurement pertinent to these various approaches.

Create an SES Composite (or Justify Use of Multiple Single Variables)

An initial panel objective was to consider alternatives and make recommendations on how an SES composite could be formed. However, the scope was widened to include the possibility of using multiple single variables to measure SES rather than a composite. Thus an objective for
the study was to consider the pros and cons of an SES composite vs. multiple single-variable measures of SES. The charge was also to consider various issues in forming a composite, such as how to weight the components of a composite, and whether to vary or keep weights constant across grades, whether to adjust weights (such as income) for locality, how to deal with the issue of missing data, and so forth.

Consider Implications of a New Measure of SES

A new measure of SES would have implications for the reporting of NAEP scores. For example, a new measure of SES might show greater achievement differences between low and high SES groups, compared to free lunch vs. non-subsidized lunch groups. A sudden change in how SES was defined might therefore disrupt trends in the relationship between SES and NAEP achievement scores, which would create significant challenges to interpreting SES estimates over time. Eligibility for a free or reduced-price lunch is a variable with three categories, which is convenient for reporting. A new measure of SES could be a continuous variable, in which case a decision would have to be made about whether to transform it into a categorical variable, or treat it in some other fashion. An objective for this study was to consider these and other implications of a new measure of SES.

4. DEFINING SES

SES is measured by different variables in different studies (e.g., Sirin, 2005), which makes it difficult to appreciate exactly what it is, or what researchers or policy makers mean by SES. However, studies on the relationship between SES and educational achievement cover more than nine decades of research (e.g., Bryant, Glazer, Hansen, & Kursch, 1974; Coleman et al., 1966; Cowan & Sellman, 2008; Harwell & LeBeau, 2010; Holley, 1916; Lynd & Lynd, 1929; White, 1982). It is useful to consider this history in developing a definition of SES.

SES emerged as a concept because of observations that students of parents with low educational attainment, working in low-status jobs, or with low income performed more poorly in school and on tests that reflected school achievement. One of the earliest SES conceptualizations was Taussig’s (1920) classification, which was based solely on father’s occupational status, classified into seven categories. In a later study by Cuff (1934), the Sims (1927) Score Card was employed as a measure of SES. The Sims Score Card contained a survey with 23 questions about home possessions (books, telephones), rooms in the home,
extracurricular and cultural activities, parents’ educational attainment, and father’s occupation. The Chapin (1933) scale was a rating scale based on the idea that socioeconomic status reflected cultural and material possessions (furniture, accessories), income, and participation in community activities, and which were reflected in and therefore could be measured by home possessions in and the condition of one’s living room.

The development of instruments such as the Sims Score Card and the Chapin scale led to increased measurement sophistication of SES. An example was Sewell’s (1940) classic study of the measurement of SES in farm families, one of the earliest sociological applications of factor analysis. Ganzeboom, De Graaf, and Treiman (1992) developed a model-based approach in which they proposed that occupational status mediated the relationship between education and income. They then computed occupational status accordingly. The Ganzeboom et al (1992) measure is currently used in The Program for International Student Assessment (PISA) to measure occupational status. Hauser and Warren (1997) similarly took into account educational levels in measuring occupational status.

In the present day, large-scale international assessments routinely include measures of SES. PISA, for example, includes items administered to fifteen-year-old students that form an SES composite called the PISA index of economic, social, and cultural status (ESCS) (OECD, 2010a; see pp. 131). The ESCS is a weighted composite (based on a principal component analysis) of three variables:

- occupational status of the parent with the higher occupational status (based on the Ganzeboom et al. [1992] model, described above),
- educational attainment of the parent with the higher educational attainment, and
- an index of home possessions.

The index of home possessions is itself a composite of three variables (derived from 16 survey questions related to home possessions) and a categorical measure of total number of books in the home:

- wealth (room of their own, Internet link, dishwasher, DVD player, and 3 country-specific measures),
- cultural possessions (classic literature, books of poetry, classic art), and
home educational resources (desk and quiet place to study, a computer available for school work, educational software, books to help with school work, technical reference books, and a dictionary)
number of books in the home (four categories: 0-10; 11-100; 101-500; over 500).

Numerous studies over the years have attempted to provide an explanation for why SES correlates with academic outcomes. The Wisconsin Model developed by William H. Sewell and colleagues (Sewell, Haller, & Portes, 1969), based on the Blau and Duncan model (1967), was one of the first attempts to account for educational and occupational attainment by proposing a recursive model including personal aspirations, the influence of peers, educational achievement, parents' SES, and cognitive ability. Along these lines, SES is related to the kind of school and the kind of classroom a student attends (Reynolds & Walberg, 1992), with schools differing characteristically in the kind of instruction offered, materials provided, teacher experience, and access to teachers (Wenglinsky, 1998), as well as the kind of relationship that exists between school staff members and parents (Watkins, 1997).

It may not be family income or poverty per se that drives the relationship between SES and achievement, and life success (Mayer 1997). Spaeth (1976) suggested that SES might indicate the complexity of a child’s cognitive environment and that exposure to cognitively challenging home environments prepares students better for the challenges of school. Levin & Belfield (2002) suggested several “pathways” or home environment variables through which SES might affect student achievement. These include the learning environment, language and literacy, parent-child interactions, and daily routine. Low SES children are less likely to have a “school-like” home and follow a daily routine; they have weaker language interaction with parents, weaker literacy engagement, and more conflicting interactions. Walpole (2003) noted that low SES students also tend to have less access to cultural capital (specialized or insider knowledge not taught in schools) and social capital (contacts in networks that can lead to personal or professional gains; Coleman, 1988), which have been argued to be key components of a student’s educational success. Recent research in genetics suggests that SES may limit opportunities for children to pursue and benefit from educational experiences congruent with genetically-influenced intellectual interests (Tucker-Drob & Harden, 2012). There also is research linking family socioeconomic resources, including a consideration of family size and structure, to student test scores (Duncan & Magnuson, 2005).
Together, these studies suggest that SES may broadly be seen as a general variable that indexes resources available to the student, including economic, social, and cultural resources. Furthermore, the “big 3” variables discussed earlier can be thought to capture different aspects of resources available to students.

Recently, the American Psychological Association (2007a) created an Office on Socioeconomic Status and issued a report from a specially commissioned American Psychological Association Task Force on Socioeconomic Status. The commission provided a framework for defining and developing SES measures. They characterized three models for understanding SES and social class-related inequalities, across three domains: education, health, and human welfare. One model, reflecting most of the SES literature as reviewed here, was what they referred to as the traditional materialist model. Another model emphasized social gradients and individuals’ positions relative to others’, which motivates the use of subjective SES measures. A third model focused on social capital, but seemed not to have resulted in specific SES measurement approaches.

Several studies have investigated what kinds of variables have been used in studies of educational achievement to measure SES. White (1982) conducted a meta-analysis on studies before 1980, and Sirin (2005) conducted a meta-analysis of studies published between 1990 and 2000. Both studies indicated a medium to strong relationship between SES and achievement with some measures showing stronger relationships than others. Sirin (2005) found that measures could be placed into the SES categories of parental educational attainment (30 studies), parental occupational status (15 studies), family income (14 studies), free or reduced-price lunch (10 studies), neighborhood (6 studies), and home resources (4 studies). Parental educational attainment was also the most commonly used measure in the studies White (1982) reviewed, and parental occupational status and family income were frequently used as SES measures.

Based on both the history of SES and the measures used to assess SES in studies of educational achievement, the primary measurement of SES over the years has been the “big 3” variables: (a) family income, (b) occupational status of heads of household, and (c) educational attainment of heads of household, consistent with what Hauser (1994) pointed out. However, school and neighborhood variables have also been included in SES measurement for some time (Hauser, 1969). For example, Fertig (2003) examined student peer group achievement heterogeneity on student achievement using PISA data. Van Ewijk and Sleegers (2010)
conducted a meta-regression analysis of the effects of peer socioeconomic status on student achievement, and showed effects at both the individual and class levels.

Panel Recommendation: A Definition of SES

A consensus definition of SES is as follows:

*SES can be defined broadly as one’s access to financial, social, cultural, and human capital resources. Traditionally a student’s SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate adjustment for household or family composition. An expanded SES measure could include measures of additional household, neighborhood, and school resources.*

5. COMPONENTS AND CORRELATES OF SES

In thinking about how SES should be defined for NAEP, it is useful to draw a distinction between components and correlates of SES. An SES component is a variable that should be included as part of SES—that is, as part of the measurement of SES. An SES correlate is simply a variable that correlates with SES, but should not be considered part of SES. It is a high priority in future NAEP data collections to include SES components, while collecting data on SES correlates is a lower priority.

This chapter focuses on determining the components and correlates of SES without regard to the practical aspects of measurement. The next chapter focuses on possibilities for measuring the components of SES.

The “Big 3”

Given the history of SES and how it has been defined over the years, and given the common ways it has been measured in research (see chapter 4), the “big 3”—family income, parental educational attainment, and parental occupational status—should be considered components of SES. Home possessions could be used to measure family income, but there are several drawbacks to using possessions as such a measure: possessions are not typically measured in surveys, do not necessarily represent an accurate picture of family income, and vary over the life cycle in uneven ways with respect to income. Still, possession measures are widely used as
SES measures in student educational surveys because they are less intrusive than income measures.

There are additional factors that could be considered components of SES. Specifically, insofar as SES is defined as access to financial, social, and human capital resources, particularly as these factors relate to schooling, they could also be considered components of SES.

**Neighborhood SES**

The argument for including neighborhood SES information in an expanded measure of student SES is that not all financial, social, and human capital resources available to the individual student come from the family. Some resources come from the neighborhood or community in which the student resides. The resources shape the home environment, broadly conceived, and have been shown to be associated with school achievement.

Traditional indicators of neighborhood SES include the percentages of families below the poverty line, unemployed adults in a neighborhood, and the adults in the neighborhood with a low education level (e.g., percentage without a high school credential). Additional indicators could include the percentage of single parent homes and the percentage of homes where English is not spoken well. In addition, there are social and physical resources associated with neighborhoods, both negative (e.g., the presence of abandoned buildings and roads and walkways in poor condition), and positive (e.g., the availability of parks, recreational areas, and public libraries), that could also be considered part of a neighborhood SES construct. There also are family and household characteristics of a neighborhood, such as aggregated family possessions (e.g., number of rooms in residents’ homes, books in their homes, and backyard facilities) which may indicate social and cultural status of a neighborhood.

Empirically, it is not necessarily the case that neighborhood SES data adds information not already available from individual level data. For example, some previous analyses of 8th-graders found that neighborhood data from Census added very little to the relationship between student and parent SES reports (Rivas & Hauser, 2008). However, there is a conceptual distinction between individual family and neighborhood measures of SES, and neighborhood SES should be considered an additional SES component.

There are advantages in including neighborhood SES as part of an individual-level measure of student SES. For one, neighborhood SES can be critical to understanding how student
psychological processes (see description in the *Psychological Process Variables* section, below) interact with the context “in real time,” and these processes may be influenced by, for example, the creation or expansion of libraries and parks or the diminishment of features such as abandoned buildings and unsafe walkways. Additionally, Census variables that might not be linked at the individual level could be used at the neighborhood level.

Defining what is meant by “neighborhood” (e.g., ZIP code, tract, block group) is difficult, however, and should be considered an operational decision to be decided later. There is also an important distinction to be made between school neighborhood and living neighborhood, as the neighborhood where students live may not have the same characteristics as the neighborhood of the school the student attends, even if they are located in the same ZIP code. For these reasons and others, the prospects for creating components of neighborhood SES that are specific enough to increase the prediction of individual-level NAEP tests scores are uncertain. The odds are sufficiently high that additional work is warranted.

**School SES**

Many students attend school in the neighborhood in which they live, but some students attend schools outside of their neighborhood due to school choice initiatives and other factors. School choice is a major movement that may lead to more disconnect between neighborhood SES and the SES composition of the schools that students attend. Therefore, both school and neighborhood SES information could be included as distinct components in an expanded measure of SES. School SES can be defined as the aggregate of the individual students’ SES. Currently, school SES is commonly measured by Title 1 status and percentage of students eligible for NSLP.

There are other characteristics of schools (e.g., school safety, physical surroundings) that are relevant for student achievement. However, they should not be considered direct components of an expanded SES measure.

**Psychological Process Variables**

Research has shown that students at different SES levels have varied levels of exposure to experience with events such as frequent moving or having contact with law enforcement in different ways. Research has also suggested that that low SES is associated with significant
risk exposure and low protection factors, and these are likely to influence achievement. Student perceptions of parental involvement and parental monitoring may affect NAEP outcomes. In addition, certain neighborhoods may lead students to adopt coping mechanisms that may not function well in a school environment, or inhibit the development of noncognitive skills such as emotional control. While these are important variables for understanding how students make sense of their environments, psychological process variables, such as coping mechanisms, perceptions, and emotional control, are variables best understood as consequences or correlates of SES rather than as necessary components of SES.

**Subjective SES**

Research on subjective SES suggests that how one thinks of one’s status subjectively can be as important as objective SES measures in relating to outcomes. For example, subjective SES has been shown to predict physical and mental health outcomes after controlling for objective SES (Demakakos, Nazroo, Breeze, & Marmot, 2008). That is, believing you are high status might compensate for lower objective status.

Measurement of subjective SES has relied extensively on the SES ladder technique (e.g., Demakakos, et al., 2008), in which respondents are shown a picture of a ten-rung ladder designed to reflect SES and asked to indicate where they think they (or their family) would stand on the ladder. Other methods for measuring subjective SES include a simple “get along” measure, asking whether the student or student’s family has enough money to get along, which has the advantage of being a relative measure that is adjusted over time. Gallup has administered a “get along” question for several decades in various adult surveys as a means to obtain a subjective estimate of poverty level (e.g., Citro & Michael, 1995).

A number of measurement challenges could hinder development of a valid measure of the subjective SES of students, particularly for 4th-graders. For example, the meaning of subjective SES may vary based on geographic location. A subjective SES measure also might not capture distinctions between high earners with modest educational backgrounds and highly educated middle-level earners. A subjective SES measure could be susceptible to reference group effects (Crede, Bashshur, & Niehorster, 2010), that is, differences in how students see themselves due to the reference group to whom they are comparing themselves. For example, students from homogeneous neighborhoods might interpret objectively small neighbor-to-neighbor differences as large because their reference group is the neighborhood in which they live.
Effects of school heterogeneity on self evaluations has been studied in international surveys [Lafontaine & Monseur, 2007], but it seems that comparable studies have not been conducted with neighborhood heterogeneity. However, if a valid measure of subjective SES could be developed, it might prove useful as a way to capture whether the child perceives that they have the resources to succeed. This would not, however, be consistent with a measure of SES that indexes access to actual resources of various types.

**Panel Recommendations: Identifying Components and Correlates of SES**

1. The primary components of SES are the “big 3” variables—family income, parental educational attainment, and parental occupational status.

2. Additional components of an expanded SES measure could include neighborhood and school SES.

3. Psychological variables and some subjective measures of SES may be useful contextual and potentially explanatory variables that could help interpret NAEP scores.

**6. APPROACHES TO MEASURING SES COMPONENTS**

The purpose of this section is to review ways of measuring the SES components identified in the previous section. The focus is on measuring the “big 3” and neighborhood and school SES. This section, reviews existing measures of each of the SES components, including school records, the NAEP student background questionnaire, the NAEP 2012 pilot student background questionnaire, and the American Community Survey (ACS). The ACS measures are included because they provide alternative socioeconomic measures, and they may be useful in characterizing geographic areas.

**Family income**

As reviewed previously, NSLP eligibility, obtained through school records, is a measure of income (adjusted for family composition), and is featured prominently in NAEP reporting. The NAEP student background questionnaire also includes items yielding data that could be understood as reflecting family income:
• Books in the home
• Encyclopedia in the home
• Magazines in the home
• Computer in the home

The 2012 NAEP pilot student background questionnaire includes additional items that may yield data pertaining to family income:

• Home possessions (internet access, clothes dryer, dishwasher, more than one bathroom, your own bedroom)

The ACS includes items pertaining to income:

• Income (9 questions, total) (for each member of the household)
• Home possessions (8 items)
• Rooms in the home (2 items)

Other indirect measures of family income

Several other variables could be considered indirect measures of family income, but are not currently measured in NAEP background questionnaires. These include:

• Housing tenure (rent or own)
• Number of moves in the past year
• Presence of household member needing healthcare assistance
• Immigration status (recency of immigration)
• School resources
• Student’s perceived level of support (home, school, neighborhood)

Housing tenure (owning as opposed to renting one’s place of residence) is an indicator of income and wealth and of residential stability. In addition, there is considerable evidence regarding its relationship to age-grade retardation and high school dropout (Frederick & Hauser, 2008; Hauser, Frederick, & Andrew, 2007; Hauser, Pager, & Simmons, 2004; Hauser, Simmons, & Pager, 2004). Number of moves in the past year serves as an indirect measure of housing tenure, and also as a measure of instability and high risk status. Presence of household
member needing healthcare assistance can drain family financial resources. Immigration status is an indirect indicator of English language proficiency, social capital, and wealth. School resources is not typically thought of as measuring family income, but could be considered indirect measures of family resources, and school resources reflect resources available to the student. Student’s perceived level of support at home, at school, and in the neighborhood also reflects the availability of resources to the student. Many of these measures could be collected through the student (and teacher and school) questionnaires, and some might be obtainable through school and Census (ACS) records.

The 2012 NAEP pilot student background questionnaire includes an item that may yield data indicating students’ English language proficiency, social capital, and wealth:

- How long have you lived in the United States?

**Household composition**

Household composition—number of parents and siblings—should be included when measuring family income. Partly this is due to the fact that family income has to be distributed across the members of the household, and so financial resources available to the individual student will be a function of both family income and the number of individuals that income is spread across. One or two parents in the household will have an opposite effect, as two parents may provide more social and emotional support than one. NSLP eligibility itself implicitly includes household composition, as its Income Eligibility Guidelines (based on the federal income poverty guidelines) are stated by household size. There are no additional questions on household composition in the NAEP student questionnaire.

The 2012 NAEP pilot student background questionnaire includes the following household structure questions:

- Size of household (total, number of adults)
- Household structure (single- vs. dual-parent, and other relatives)

There may be some ACS variables that could be added to this list, such as number of workers in the household and number of earners in the family.
Parental educational attainment

The NAEP student questionnaire includes two parental educational attainment questions:

- Mother’s educational attainment (8th and 12th grade only)
- Father’s educational attainment (8th and 12th grade only)

The ACS includes educational attainment questions for each member of the household:

- Whether currently attending school (level and type)
- Educational attainment
- Major (for bachelor’s degree holders)

Parental Occupational Status and Employment Status

The NAEP student questionnaire does not include any questions about parental occupation and employment status, nor is such information available from school records. Therefore this SES component has been absent from NAEP reporting.

The 2012 NAEP pilot student background questionnaire includes the following question about parental and household resident employment status:

- How many adults living in your home have a job?

The ACS includes the following employment status and occupation questions for each member of the household:

- Employment status (working for pay or not, part-time vs. full-time, etc.; 22 questions total)
- Occupation (6 questions total)
The National Education Longitudinal Study (NELS:88) (Ingels, 1990) asked 8th-graders for their mother’s and father’s occupation in an open-ended question. But it also included a closed (multiple-choice) question: “what kind of work do you expect to be doing when you are 30 years old?” The response choices included categories such as craftsperson or operator, farmer or farm manager, professional business or managerial, and so on.

Cognitive laboratory studies must be conducted on various question types for collecting student reports on parental occupation. If questions could be developed to provide reliable information on parental occupation, then it would be useful to use these data in creating a better measure of SES, even if such information does not reach the same reliability and validity level as other questionnaire responses.

There are upcoming opportunities to collect data on new SES component measures. For example, the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) (NCES, 2012a) wave will test 4th-graders in 2014. That study will collect data from both parents and students. Certain questions concerning parental educational attainment, occupation, home possessions, or any other SES-related questions could be inserted into the parent and the student ECLS-K:2011 questionnaires for the 2014 study. A comparison could be made between the responses to evaluate the validity of 4th-grade student data.

**Neighborhood SES**

There are currently no direct measures of neighborhood SES from either the NAEP student questionnaire or school records. However, the 2012 NAEP pilot student background questionnaire includes the self-reported ZIP code item, from which neighborhood information could be obtained.

There are a variety of ways to measure neighborhood SES. In addition to measures such as percentage below poverty, unemployed, and with low educational attainment, other variables include the availability of parks and libraries in the neighborhood, the absence of abandoned buildings, proportion of single-parent households, and the proportion of households in which English is not spoken well.

Some neighborhood SES information could be obtained through the extended school questionnaire. Some items from the student questionnaire and from school records could also be aggregated so as to serve as neighborhood SES measures.
Additionally, the ACS includes a number of items that could be treated as measures of neighborhood SES, including unemployment, education, and income levels, household overcrowding, poverty, home ownership, and perhaps some indicators of vulnerability. ACS data would be suitable for neighborhood measures of SES, though there are some challenges in using ACS data, such as heterogeneity in neighborhoods at the tract level and above. Information obtained from the ACS 5-year estimates (with the least sampling error and provided at the smallest geographical unit) would not reflect rapid changes in a neighborhood, but neighborhoods tend to be very stable, and this is not likely to be a problem. Another challenge is that the size and boundaries of a neighborhood for the purposes of creating a neighborhood SES variable are undefined. Perhaps Census blocks (the smallest geographic area for which data are collected and tabulated), or block groups (optimal size of 1,500 people), or possibly even Census tracts (optimal size of 4,000 people), or ZIP code tabulation areas could serve as neighborhood boundaries for this purpose. Using Census blocks or tracts would require obtaining more precise location information for student households than student ZIP codes and would require special arrangements with the Census Bureau.

School SES

As noted above, school SES is most commonly measured by Title 1 status and percentage of students who are eligible for NSLP. However, additional school SES variables could be formed as aggregations of student-level variables obtained through school records or student questionnaires, such as percentage of English language learners, average level of parental educational attainment, average home possessions, and so forth. In addition, characteristics of the schools and school climate could be obtained through teacher and school questionnaires, and these could be treated as part of a school SES variable. Because most students attend neighborhood schools, it is likely that school and neighborhood SES measures would correlate highly, but it also may be that there is unique information in the school and neighborhood SES measures.

Panel Recommendation: Review Data Collection and Measurement Approaches

1. Additional variables could be studied as indirect measures associated with family income, including housing tenure (rent or own), number of moves in the past year, presence of a household member needing healthcare assistance, immigration status (and recency of
immigration), school resources, and student’s perceived level of support (home, school, neighborhood).

2. Family/household composition and structure—size of household and whether single or dual-parent—are also important variables to consider both because single parenthood is generally considered a disadvantage and because household resources are diluted in large households.

3. Parental education is currently measured through the NAEP questionnaire, but only for 8th- and 12th-graders. The ACS includes parental educational attainment questions, which could be used to obtain this measure for 4th-graders. However, a strategy would have to be developed to determine how to link ACS data to NAEP.

4. Cognitive laboratory studies should be conducted on various question types for collecting student reports on parental occupation. If a proper format could be identified for collecting reliable information on parental occupation, then it might be useful to include such questions in future questionnaires even if the reliability and validity level were not as high as is expected for other questionnaire responses.

5. There are currently no direct measures of neighborhood SES from either the NAEP student questionnaire or school records. However, if student ZIP code could be obtained, it may be possible for NAEP data to be linked to ACS data in order to compute neighborhood SES measures for the students’ residential neighborhoods (such as unemployment, education, and income levels, household overcrowding, poverty, home ownership, and perhaps some indicators of vulnerability). Research is needed to determine whether ZIP-code defined “neighborhoods” will yield useful additional components for an expanded SES measure. Data from student questionnaires and information from school records also could be aggregated to serve as neighborhood SES measures, although only for neighborhood schools.

6. School SES can be measured using Title 1 status and percentage of students eligible for NSLP. Additional school SES variables could be computed as aggregations of student-level variables, obtained through school records or through student questionnaires, such as percentage of English language learners, average level of parental educational attainment, and average home possessions. School characteristic and climate variables could be obtained through teacher and school questionnaires and these could be part of a school SES variable.
7. An ideal opportunity to inform SES measurement is available through participation in the upcoming Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011) (NCES, 2012a), which tests 4th-graders in 2014. The study will collect data from both parents and students, enabling a comparison of parent and 4th-grade student reports to test the validity of the student reports.

7. SES COMPOSITE

In the literature and in official reports SES is sometimes reported as a single variable, such as parental educational attainment level or NSLP eligibility, and sometimes as a composite variable with several component variables summed together. Initially, the panel was to consider alternatives and make recommendations on how an SES composite could be formed. However, during panel deliberations the scope was widened to include the possibility of the use of a single variable (or multiple single variables) rather than a composite to measure SES. Thus an objective for the panel was to consider the pros and cons of an SES composite vs. single-variable measure of SES. The charge was also to consider various issues in how to form a composite, such as how to weight the components of a composite, and whether to vary or keep weights constant across grades, whether to adjust weights (such as income) for locality, whether to change weights every year, or periodically, and so forth.

An advantage of treating SES as a single variable is that the meaning of a single variable is typically clear and easy to communicate. For example, audiences know what it means to have a parent who completed college, or to be eligible for a free lunch through the NSLP. A related advantage is that the meaning of different SES levels when defined as a single variable can be transparent—it is clear what the differences are between groups of students who are eligible for a free lunch, eligible for a reduced price lunch, or not eligible. It is also clear what the differences are for students whose parents completed high school versus completed college.

A disadvantage to treating SES as a single variable is that SES is typically understood as having multiple components, including family income, parental occupational status, and parental educational attainment. Treating SES as only one of these components is at odds with the conventional definition of SES. There also is more measurement error in a single variable compared to a composite variable. Some of these problems could be overcome by treating SES as multiple separate variables. However, doing so complicates reporting and interpretation. The separate variables constitute separate operational definitions of SES, which could lead to
potentially conflicting data about the relationship between achievement and SES, defined in different ways. For example, achievement for levels of parental educational attainment might show greater variability than achievement for different levels of income.

A composite variable combines information from all the components in a single variable, maintaining simplicity in reporting and avoiding conflicting stories about relationships to achievement. This could mask differences between components in their relationship to achievement, of course, and that could be a potential disadvantage to a composite variable. Nevertheless, the advantages of a composite variable generally outweigh the disadvantages. The remainder of this chapter focuses on a composite measure of SES.

**General model**

There are several ways to think about what a composite SES variable might look like and how it could be formed. In the psychometrics literature there are two kinds of latent variable models, formative and reflective (Bollen, 2002; Edwards & Bagozzi, 2000). A *reflective* measurement model is one in which the latent variable is assumed to be the cause of the measures or indicators (i.e., which are commonly called reflective or effect indicators, Blalock, 1964). For example, cognitive ability and personality are commonly assumed to cause responses to particular tests or test items. Changes in the latent variable cause changes in the indicator variables. Cronbach’s (1959) alpha, factor analysis, and classical test theory are all reflective measurement models—covariation among indicators is assumed to be caused by an underlying latent variable. A *formative* measurement model is one in which the latent variable is assumed to be caused by the indicators (i.e., which are commonly called formative or cause indicators). That is, changes in the indicators cause changes in the latent variable. SES is commonly understood as a latent variable in a formative measurement sense because SES does not cause income, educational attainment, or occupational status; rather, income, educational attainment, and occupational status cause (or determine) SES. A fundamental difference between reflective and formative measurement is that latent variables in reflective measurement are defined by the degree to which indicators covary, and in fact the pattern of covariances can be used to compute latent factor scores. The situation is different with SES and other latent variables in formative measurement. Here, indicator variables have no necessary relationship with one another and can be uncorrelated or negatively correlated with each other. An example of formative measurement could be variables such as the stress scale (Holmes & Rahe, 1967), which is simply a count of the number of stress-inducing events experienced by
an individual within a relatively short period of time (e.g., a year). Indicators are life events, such as death of a spouse, imprisonment, personal injury, or pregnancy, none of which have any necessary relationship to each other (i.e., they can correlate positively, negatively, or be uncorrelated). However, the stress scale predicts future events, such as subsequent illness. In the same way, SES is useful for its ability to predict present and future academic achievement and other life outcomes.

To produce a composite index or score in formative measurement, some scheme must be used to weight the components in some fashion. Below is a review of several ways to do this.

**Arbitrary weighting**

An infinite number of arbitrary weighting schemes are possible for forming an SES composite. For example, the number of years of parental educational attainment could be added to annual family income and a rating of job status to form an SES composite. However, the weight of these measures would be related to the variance of the components, and somewhat arbitrary rescalings of components (e.g., changing from income in dollars to income in cents or to a three-level value, such as 1 = high; 2 = medium; 3 = low) could have dramatic effects on the composite. Putting components on the same scale (e.g., through the use of standard scores, or z scores) would be a way to avoid this problem. An analyst or policy maker might believe that parental educational attainment is the most important component of SES in an educational application such as NAEP, and so parental educational attainment could be given more weight (e.g., twice the weight) than the other components in forming an SES composite.

An advantage of arbitrary weighting is that it is easy to communicate the rules by which components are combined to form a composite. For example, the “misery index” is the sum of the employment rate plus the inflation rate; the United Nations Development Program’s “human development index” is a more complicated geometric mean of normalized indices, but is nevertheless arbitrary. However, the arbitrariness of these indexes is a visible feature, in that their makeup is clear and transparent. A disadvantage of arbitrary weighting is that it is arbitrary. There is no reason to prefer one set of weights to another, and different weights might give different answers to substantive questions (such as, what is the relationship between SES and achievement?)
Empirical weighting

If SES were treated as a latent variable with reflective indicators then component weights could be developed using factor analysis or principal component analysis of the indicators. A rationale for treating SES as a latent variable with reflective indicators is that its components correlate. Treating SES as a latent variable with reflective indicators implies that changing SES would result in a change in income, parental occupational status, and parental education, which seems implausible. Still, the reflective indicators assumption and approach to identifying and weighting an SES composite is used in PISA (see e.g., OECD, 2010a). In PISA, SES (PISA’s Economic, Social, and Cultural Status index, or ESCS index) is computed from highest parental educational attainment (in number of years of education), highest parental occupation (converted to a status index), and number of home possessions (summing over 20 items), including books in the home. Weighting these three components is determined by a principal component analysis (conducted separately for each participating country) based on the covariances among the three components, and it has typically yielded approximately equal weights for the three components (although with job status given the most weight, education the second most, and home possessions the smallest). Thus PISA treats SES, at least partly, as a latent variable with reflective indicators.

However, as discussed previously, SES is more commonly thought of as a latent variable with formative indicators, because it is assumed that SES is caused by its indicators rather than the other way around. If SES is treated as a latent variable with formative indicators, then weights cannot be assigned by a covariance-based approach (e.g., principal component analysis) that only considers the components of SES. Such a system weights components according to their centrality (similarity or correlation) with respect to each other, but components do not have to be correlated with each other in a formative measurement model. Instead, under a formative variable assumption, an approach to forming a composite with non-arbitrary weights would be to compute weights through multiple regression analysis using an outcome variable. An outcome variable, such as NAEP Mathematics scores, could be regressed on the SES component variables, and the estimated weights could be used to form an SES composite optimized for predicting NAEP Mathematics scores for that particular grade and year. There are potential drawbacks to producing composite scores this way. One perspective is that SES emerged as a construct because of its predictive relationship with educational outcomes. It is therefore fair and reasonable to weight SES components according to the regression weights of those components when predicting educational outcomes, such as NAEP scores. A complexity
related to this perspective is that regression weights will, in general, change depending on which NAEP scores are being predicted (e.g., 4th-grade reading, 8th-grade mathematics, 12th-grade civics, etc.), and in what year they are being predicted (e.g., 2011, 2013). This issue is revisited in the next section of this chapter.

An alternative viewpoint is that SES should have an identity separate from its ability to predict particular achievement outcomes. This viewpoint reflects the perspective that little is learned by studying the relationship between SES and achievement if SES becomes little more than a set of variables optimally weighted to predict achievement. Instead, the relationship between SES and achievement should be a finding rather than an optimization exercise.

There are two ways out of this impasse between a tailored (i.e., regression-weighted) and independent (weights determined without regard to the composite’s prediction of achievement) SES composite. If a composite predicts achievement equally well under a range of composite weights (e.g., unit weights, weights determined by regression with 4th-grade mathematics, or with 12th-grade reading), then the distinction between the two composite weighting approaches is of little practical importance. There is some evidence for this perspective (Noel-Miller & Hauser, 2011; Wilks, 1938).

Another way to define an SES composite empirically without tuning component weights to maximize prediction with NAEP scores is to consider additional outcome measures. This was an approach originally suggested by Hauser and Goldberger (1971) as the multiple indicator multiple cause (MIMIC) model (for a recent discussion regarding how this strategy helps identify formative latent variable models, see MacCallum & Browne, 1993). For example, SES is a widely used construct in the health literature, and a health outcome (e.g., absenteeism due to illness) could be used as an additional outcome variable that could be regressed on the SES components. Such a model could be estimated using a structural equation modeling (SEM) approach. Doing so would lead to an SES composite that was not being tuned specifically to the prediction of NAEP achievement. (Multiple NAEP population groups, for example, with various achievement scores at different grade levels, could also be used for this purpose, but the generalization would be to NAEP achievement, not to life outcomes in general.) At the same time, component weights would not be arbitrary, but would be based on the predictiveness of the SES composite across diverse outcomes.
Measurement invariance goals

In developing an SES composite, regardless of whether arbitrary or empirical weights are used and whether SES is a latent variable with formative or reflective indicators, there is an issue concerning the degree to which the composite should be defined in a consistent (i.e., invariant) way (i.e., given the same component weights) across situations; that is, across grades, across NAEP subject areas, across time, across locations, and so forth. One (extreme) option would be to have a specific SES composite for every measurement occasion. That is, there could be a 4th-grade mathematics SES composite for New York in 2013, and a separate SES composite for New Jersey, and separate SES composites for each grade, for each subject, and for each testing year. There could be other categories by which SES composites could be separately formed, such as urban, suburban, and rural, or by cost-of-living areas (a given family income, say $40,000/year, might indicate different socioeconomic status depending on whether the family resided in Manhattan or the rural south). A National Research Council report on a new poverty measure (Citro & Michael, 1995) recommended adjustment for geographic differences in the cost of housing and insurance.

However, there is a benefit of having an SES composite that maintains the same component weights across all measurement occasions (i.e., across grades, subjects, locations, and years), namely, consistent SES measurement can simplify reporting and interpretation. In current reporting, SES indicators, such as parental educational attainment, NSLP eligibility, and home possessions, are measured the same across all contexts. That is, these indicators are invariant in the raw (manifest) metric across grades, subjects, locations, and years.

The issue of how and the degree to which the SES composite can be kept invariant is affected by the form of the SES composite. For example, with arbitrary weights for the components (sum of unit weighted parental educational attainment, parental occupational status, family income), educational attainment, occupational status, and income could be standardized separately within 4th, 8th, and 12th grade, or they could be standardized across grades (e.g., parental educational attainment could be placed on a common scale across all three grades, or on separate scales, one for each grade). With regression-based weighting for SES components, weights could be obtained from a regression analysis for one grade and one subject, and applied to other grades and subjects, or separate regression analyses for each grade and subject could be conducted, or weights could be averaged across subjects or grades (c.f., Noel-Miller & Hauser, 2011).
Weights could be identical or similar across subjects and years, which might make interpretation and reporting simpler. However, this might not be possible across grades due to the differences in information that can be collected from 4th-graders versus 8th- or 12th-graders. Based on prior research (e.g., Dawes, 1979; Noel-Warren & Hauser, 2011; Wilks, 1938), the weighting scheme might not have much impact on the identity of SES (i.e., applying two sets of component weights to construct an SES composite would likely result in two versions of SES that were highly correlated). Therefore a simpler approach, such as using unit weights or average weights (where average weights involve averaging the component weights obtained in one context with comparable weights obtained in another context), might be advisable. Examining the existing literature (Cohen, 1990) and closely reviewing the quality of data (once it was determined what data would be collected and from what sources) is the most appropriate course of action in determining weighting.

Missing data issues

Dealing with the issue of missing data may be more critical in the case of composite variables compared to single variables such as parental educational attainment (or NSLP), simply because there are more opportunities for data to be missing (e.g., through skips by the respondent). If casewise deletion were invoked any time any of the component items for an SES composite were missing, that could result in both a relatively high number of missing values, and the introduction of bias if data were not missing completely at random (using the standard terminology from Little & Rubin, 1987).

However, there are probably no special problems associated with imputing missing data in the case of computing the SES composite. For example, a standard practice (e.g., used in PISA) is to impute missing values for students with missing data for one of the SES components using data from the other two components. In general, either a maximum likelihood approach for handling missing data in the context of modeling the data, or a multiple imputation approach similar to that used for handling missing achievement data, could be used and would be worth exploring for this purpose (Enders, 2010).

Panel Recommendation: Create an SES Composite

1. The advantages of treating SES as a composite of several variables rather than as a single variable or multiple single variables outweigh the disadvantages.
2. The formative-reflective measurement model distinction was important in considering how to combine SES components into a composite measure. The literature and data quality should be examined before proposing a recommendation on a component weighting scheme.

3. Further study is necessary to address missing data issues in SES measurement.

8. IMPLICATIONS

Adopting a new measure of SES would have various implications on the reporting of NAEP scores. To begin with, a new measure would have to be clearly explained and communicated, because a new measure of SES might show greater achievement differences between low and high SES groups, compared to free lunch versus non-subsidized lunch groups. A sudden change in how SES was defined might therefore disrupt trends in the relationship between SES and NAEP achievement scores, which would create significant challenges to interpreting SES estimates over time.

Reporting and Implications for Trend

As reviewed in chapter 2, achievement scores are disaggregated in NAEP reports by individual SES proxy variables, most notably eligibility for NSLP (not eligible, eligible for a reduced-price lunch, and eligible for a free lunch). Eligibility for a free or reduced-price lunch is a variable with three categories, which is convenient for reporting. A new measure of SES could, and likely would, be a continuous variable. In that case, a decision would have to be made about whether to transform the continuous variable into a categorical variable, or treat it in some other fashion. If it were transformed into a categorical variable, a decision would have to be made about how many categories it could be reported by (e.g., three, more?) and how these categories would be labeled (e.g., low, medium, and high SES).

A new measure of SES would not have to be treated as categorical, however. In PISA (OECD 2010a, Figure II.1.3, p. 32), for example, SES data are reported on a continuous scale, with scatter plots of achievement scores and the PISA index of economic, social, and cultural status (ESCS), and a regression line of achievement on ESCS. With ESCS presented as a continuous variable, PISA reporting makes considerable use of presentations (e.g., tables and scatter plots of ESCS against a variety of variables), and the use of ESCS as a control variable in examining factors such as single-parent families and the like. PISA also computes “socio-economic
gradients” that characterize the within-country relationships between ESCS and achievement, facilitating country-to-country comparisons on that measure.

A continuous SES variable could be used in NAEP reporting, but it would not have to be limited to presentations in scatter plots, or as a gradient index. For example, expected SES achievement at, say, the 20th, 50th, and 80th percentiles of SES could be presented, or at the mean SES and at a level one standard deviation above and below the mean of SES. These displays would take a form similar to that taken by NSLP eligibility.

For understanding trends in variables undergoing changes, as SES would be if a new measure were adopted, it is useful to conduct bridge studies, such as those conducted as a result of new race/ethnicity classifications introduced in Census 2000 (Parker, Schenker, Ingram, Weed, Heck, & Madans, 2004). For SES, a carefully constructed study enabling bridging to NSLP eligibility could be useful for understanding trends. For example, for a reporting cycle or two, both SES and NSLP eligibility could be reported as the audience became familiar with the new scale. This would allow readers to compare SES effect sizes (on achievement) with SES measured by NSLP eligibility versus SES measured by a new composite.

**Data conditioning**

NAEP uses a balanced incomplete block design for administering only subsets of the item pool to particular students (i.e., each student only takes 2 of 11 blocks of items). Background information, including SES, along with data from the items actually administered, is used to estimate scores on the items that are not administered to a particular student, a process referred to as conditioning (Mislevy, 1991). Changing the measure of SES would likely lead to changes in the conditioning model, and changes in the posterior distributions of student responses from which plausible values that secondary analysts use are drawn. A question is how severe a difference in the conditioning model would likely result from a change in the makeup of SES (e.g., from NSLP to a new SES index). There is literature suggesting that the demographic variables are the most important background variables affecting the conditioning model (Thomas, 2002). And SES is likely to be among the more important demographic variables. That same literature, however, suggests that background variables are not as important to the conditioning model as the cognitive variables themselves (i.e., the responses to the cognitive items that are administered). As with the proposed bridge study, current variables, such as
NSLP, could be retained, and differences in the conditioning model due to the inclusion of a new SES measure could be studied.

**Use by other units, departments, agencies**

The focus of the present effort is developing a new SES measure for NAEP. A new SES measure could have direct effects in reporting NAEP scores, such as providing a more valid estimate of the relationship between SES and achievement. In addition, SES is used for the conditioning model in NAEP to assist in the estimation of proficiency scores, and a better measure of SES could be more predictive of proficiency scores and thereby more useful for data conditioning. The quality of NAEP data reported could therefore improve as a result of a better SES measure. In addition to these specific benefits for NAEP reporting, there would be additional benefits based on secondary analysis of NAEP. SES, or proxy measures such as NSLP, is widely used in secondary analysis of NAEP data (e.g., Harwell & LeBeau, 2010; Sirin, 2005).

NCES programs beyond NAEP might benefit from the work conducted in defining and developing a new SES measure. NCES Fast Facts (2012b) provides a list of NCES surveys, many of which use SES measures of some kind. These include adult literacy surveys (National Assessments of Adult Literacy [NAAL], the Program for International Assessment of Adult Competencies [PIAAC]), international comparative surveys (Trends in International Mathematics and Science Study [TIMSS], Progress in International Reading Literacy [PIRLS], the Program for International Student Assessment [PISA]), longitudinal surveys (the Early Childhood Longitudinal Study [ECLS], Baccalaureate and Beyond [B&B], Beginning Postsecondary Students Longitudinal Study [BPS]), and so forth. For some of these studies there could be a fairly direct transfer of findings on improving SES measurement. For other studies some of the research, methods, and lessons learned in developing an improved SES measure could be incorporated into future study designs.

Investigating new methods for measuring SES could produce benefits that extend beyond NCES and the U.S. Department of Education. For example, in the health sector, there is an extensive literature that relates SES to women’s health, public health, and psychological health (APA, 2007b); to specific conditions, such as cancer (Singh, Miller, Hankey, Edwards, 2003) and cardiovascular disease (Winkleby, Jatulis, Frank, Fortmann, 1992); and to other health and wellness issues. Agencies such as the National Institutes of Health (NIH) and the Center for
Disease Control (CDC) may benefit from research conducted for NAEP by NCES in developing improved measures of SES.

**Anticipated effects and unanticipated side effects**

Developing a new SES measure is likely to involve both anticipated effects and unanticipated side effects. It is reasonable to assume that developing the new measure will involve an interagency agreement and collaboration between NCES and the Census Bureau. Such interagency collaborations are beneficial, but often introduce scheduling complications, new costs, and other challenges that require flexibility, patience, and a willingness to consider a variety of approaches to solving potential problems. Another relatively minor change will be a requirement to collect ZIP code information from respondents, perhaps on the NAEP questionnaire. Privacy issues are also likely to be important to resolve.

It is always difficult to anticipate the unanticipated side effects of measurement changes, but as with any assessment, new measurement is often accompanied by the element of consequential validity (Messick, 1995). Consequential validity refers to the changes in practice or culture that accompany changes in assessment. For example, introducing writing assessments can lead to an increased emphasis on writing instruction in the schools; introducing a high-stakes noncognitive skills assessment can lead to more emphasis on developing noncognitive skills. It is not entirely predictable what changes might accompany the introduction of a new SES measure, but if such a measure proves to be more valid than current measures, it is possible that more attention could be given to the importance of the SES-achievement relationship and to a more equitable distribution of educational resources.

**Panel Recommendation: Consider Implications of a New Measure of SES**

1. There are reporting and psychometric implications that should be considered before implementation of a new SES measure. They include whether and how to characterize SES levels, whether to conduct a bridge study linking new and old measures of SES, and studying the implications of a new SES measure on the conditioning model used by NAEP to generate plausible values.
The goal of this panel was to provide recommendations for a new measure of SES that could be used in NAEP. The role of the white paper was to serve as technical documentation of the panel deliberations and to bring this work to the attention of stakeholders and the research community to engage discussion about SES and its measurement.

NAEP is required by law to report scores by SES. Current SES measures, such as NSLP eligibility and parental educational attainment, are single proxy variables, which are limited in several ways. Historically, SES has been defined as a composite measure reflecting resources available to the individual, as expressed in family income, parental educational attainment, parental occupational status, and sometimes neighborhood resources. A common view, as reflected in other large-scale educational assessments such as PISA, is that composite measures that include all of the SES components may be more informative than single measures.

A second limitation of current SES measurement concerns the quality of the data. Student reports of some SES components (such as parental educational attainment) may be unreliable and biased, and reports on variables like these by 4th-graders are likely to be particularly unreliable. This is not to say that they are unusable. Attempts to collect data from 4th-graders on parental educational attainment and perhaps even parental occupational status should be revisited. However, additional data sources such as NCES and state assessment databases and private data sources should also be considered to help bolster the quality of an SES measure.

Perhaps the most critical data quality issue in current SES measurement concerns NSLP eligibility. Measures of NSLP eligibility have several problems, including large errors in eligibility certification and jurisdiction-wide eligibility which fails to differentiate poverty levels within schools or jurisdictions where everyone is declared NSLP eligible (Harwell & LeBeau, 2010; Hauser, 1994). Most importantly, that trend is likely to continue and even get worse.

Given the current limitations of how NAEP measures SES, a major contribution of the panel was to devise a consensus definition of SES, based on a review of various perspectives on SES:

*SES can be defined broadly as one’s access to financial, social, cultural, and human capital resources. Traditionally a student’s SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate*
adjustment for household or family composition. An expanded SES measure could include measures of additional household, neighborhood, and school resources.

Note that this definition outlines and provides a justification for both a core SES measure, which should be the subject of immediate focus for operational reporting, and a more expanded measure, which could be treated as a research project intended to illuminate some of the more contextual and explanatory aspects of SES.

There are other potential components of SES, such as subjective SES and psychological factors. These are best understood as contextual and explanatory variables that could help in the interpretation of SES-achievement relationships, but these contextual factors should not be considered part of a core SES meeting the charge of a congressionally mandated reporting variable. A research program studying these variables, however, could be critical for understanding the importance of measuring SES in the context of an educational achievement survey.

The panel reviewed existing and proposed new measures of SES components from sources including school records, the student questionnaire, additional potential NAEP questionnaire items that were pilot-tested in 2009, 2011, and 2012, and questions from the American Community Survey (ACS). Measures reviewed included ones pertaining to family income and home possessions, parental educational attainment, parental occupational status, and neighborhood wealth and resource indicators. Additional measures that might be related to family income and resources, such as housing tenure, number of residence moves, household members’ healthcare needs, immigration status, and household composition measures were also considered. Some of these can be obtained from ACS data. Although NAEP 4th-grade questionnaires do not ask students to indicate parental educational attainment in the questionnaire due to low data quality, such information can be obtained from ACS data. Occupational information is not asked about in the NAEP questionnaires, again due to concerns with low data quality, but such information can be obtained from ACS data. An extensive amount of neighborhood SES data could be obtained from the ACS, including neighborhood poverty levels, unemployment, educational attainment, presence of parks and libraries, abandoned buildings, single-parent households, and non-English speaking households. However, there are challenges in obtaining these kinds of data from ACS and for linking ACS data to NAEP data, such as determining how best to aggregate data in linking datasets.
There are a wide variety of ways to combine all the information on components of SES. A composite can be assembled by summing variables reflecting family income, parental educational attainment, parental occupational status, and neighborhood SES indicators. The primary distinction is in whether the summing would occur by arbitrarily weighting the components (e.g., unit weighted), or by allowing the components to be weighted to best predict some outcome, such as student achievement. There are advantages and disadvantages to both approaches. Another important consideration would be how to maintain the meaning of SES across grades, across locations (e.g., varying cost-of-living regions), and across time. However, with respect to the issue of component weighting, there is some evidence that this may be merely academic and that practically how variables are weighted might not make much difference in what SES is (Noel-Miller & Hauser, 2011). That is, an SES composite with a set of weights determined from one context is likely to be highly correlated with an SES composite based on a set of weights determined from a different context, given that the components themselves tend to be highly correlated, and neither will be much different from a unit-weighted composite, as has been long known (Wilks, 1938).

Developing a new SES measure for NAEP has implications for reporting and elsewhere. If a new measure were developed, it might be useful to report achievement results disaggregated by SES, measured both by the current measures (e.g., NSLP eligibility, parental educational attainment) and by the new composite measure. While it may be valuable to treat SES as a continuous variable, it could also be treated as a categorical variable (e.g., low, medium, and high SES). The research and findings resulting from developing the new SES measure for NAEP would benefit other federal programs both within and outside NCES.

**Key Recommendations**

Summarized below are the panel’s key recommendations for improving the measurement and reporting of SES.

*Recommendation 1.* Family income and other indicators of home possessions and resources, parental educational attainment, and parental occupational status (the “big 3”), should all be considered components of a core SES measure; that is, part of the measurement of a core SES variable. The core SES measure should be the subject of immediate focus for operational reporting. This recommendation reflects the academic literature on SES.
**Recommendation 2.** An expanded SES measure could include additional variable components besides family income, parental educational attainment, and parental occupational status. These additional components could include resources available in the student’s neighborhood or community and resources available at school. Consideration should be given to the development of an expanded SES measure in addition to the core SES measure.

**Recommendation 3.** The advantages of treating SES as a composite—e.g., a single summary for reporting, greater reliability, and representation of the full range of SES factors—outweigh the disadvantages, especially because the use of the composite would not preclude using and reporting on single measures. Therefore, attempts should be made to develop an SES composite measure.

**Recommendation 4.** The validity of the most widely used measure of SES—NSLP eligibility—has been decreasing due to jurisdiction-wide eligibility and other factors, and that trend is likely to continue. There will be growing pressure to replace NSLP eligibility with a new, more valid measure. Burden issues prohibit a longer questionnaire, and there is concern about the reliability of student reports on SES components, particularly educational attainment (for 4th-graders) and occupation (for all grades). Because of data quality issues, along with burden considerations, attempts should be made to explore the possibility of linking to Census data on SES components. Studies should be conducted with the U.S. Census Bureau to determine the feasibility of linking Census data to NAEP and to evaluate the quality of the data that would result from various linking strategies.

**REFERENCES**


Chapin, F. S. (1933). *The measurement of social status, 3*, University of Minnesota Press.


Sims V. M. (1927).*The measurement of socioeconomic status.* Bloomington, IL: Public School Printing Co.


Background Information Framework for the National Assessment of Educational Progress

National Assessment Governing Board
Adopted August 1, 2003
Chapter Three

Topics and Types of Background Data

This chapter will cover the non-cognitive topics that are required for reporting under the No Child Left Behind Act of 2001 (P.L. 107-110), as well as those that should be considered for inclusion in NAEP on a cyclical basis. It discusses socio-economic status (SES), contextual factors of interest to public policy, and subject-specific variables.

Demographic Reporting Categories

The demographic variables currently collected by NAEP are gender, age, race/ethnicity, and two elements of socio-economic status (SES)—participation in Title I, and eligibility for free or reduced-price lunch, which is based on family income. In addition, information is obtained on disability status and on students who are classified as limited English proficient. All of this information is collected on an administration roster, completed from school records in advance of testing. In addition, data on race/ethnicity is also collected on the NAEP student questionnaire, and students are asked to report on two other SES variables: the number of reading materials at home and the highest level of each parent’s education.

A more extensive questionnaire is completed by school staff on each student selected for NAEP who is classified as either disabled or limited English proficient (LEP). For students with disabilities (SD), the questionnaire collects data on the specific disability and its severity, the student’s Individualized Education Plan (IEP), type of curriculum, whether the student participates in standardized testing (with or without accommodations), and the accommodations allowed on state and district standardized tests in presentation, response, setting, and timing. For LEP students, the questionnaire covers native language, number of years of academic instruction in English, percent of instruction in English and/or native language, and the testing accommodations provided under district or state policy. In the future, NAEP might also identify students who recently exited from LEP programs and track their achievement.
NAEP is required to collect information on all of these categories (except age), but has some discretion in determining definitions and aggregating responses. These data will continue to be collected in a uniform manner in every NAEP assessment, although, for socio-economic status, as explained in the section below, there may be some variation, with a uniform core and more extensive data-gathering in some cases.

Socio-Economic Status (SES)

Under current law, NAEP is required to collect information on socio-economic status. SES also is clearly a factor that has been shown to be related to academic achievement in many research studies, beginning with the Equality of Educational Opportunity Commission Report (Coleman et al., 1966). The research community’s consensus over the past four decades has been to deal with the influence of SES on other achievement-related variables by holding SES constant while examining the other effects, for example, adjusting for SES while looking at effects of class size or teacher training. NAEP does not adjust for SES, but it does report on the relationship between student achievement and SES proxy variables like parents’ education or Title I participation.

NAEP has not been able to measure SES directly, using its present set of questions and data sources, i.e., the student, teacher, and school questionnaires. The assessment has used “proxy variables” for SES, including students’ eligibility for the National School Lunch program, participation in Title I, parents’ education, and the number of reading materials in the home (newspapers, magazines, books, etc.)—information on the latter two factors being reported by students in the assessment samples. In addition, NAEP uses census data to classify schools by type of location, based on Census Bureau definitions, such as central city, suburban/large town, and rural/small town.

Strictly speaking, these are individual proxy variables and are not combined into a composite variable. However, both the questions on parent education and home environment are coded in a pseudo-composite manner. For example, the parent education related to the student is the higher of either the mother’s or father’s education level. On the four home environment questions, student responses are coded differently for a “yes” answer to two questions or fewer,
“yes” to three questions, and “yes” to four questions, as well as omitted responses (Allen, Carlson, and Zelenak, 1999).

At the lower grade levels, students’ reports of their parents’ education are questionable at best, while the National School Lunch program sorts students only into three categories (Yes, No, and Unknown) and Title I into two categories (Yes or No). For many years, NAEP used a reporting category of disadvantaged urban schools, which was constructed from information provided by school principals. This was discontinued in the mid-1990s because the category lacked a consistent definition from year to year and between different state samples. There also were serious doubts about the reliability of the information on which it was based. In short, there has been considerable concern over many years about the quality of the SES measures in NAEP, both for reporting to the public and for analysis by researchers.

Barton (2002) suggests two alternative approaches for improvement: (1) a composite index for SES, or (2) a parent questionnaire. A composite index is viable using the same information that is currently collected in NAEP, or perhaps augmented with a few targeted questions or census data, possibly the zip code of student home addresses. The necessary analytical work should be initiated through small research studies using extant NAEP data sets in order to check systematically the validity of a composite index as a better measure of SES in NAEP samples. The results could vary by grade level, in which case, adjustments might be needed in the way the data are collected, augmented, and/or confirmed. NAEP may never be able to produce a full composite of income, education, and occupation, but efforts ought to be made to find an approximation that is more reliable than the current set of individual proxy variables.

The argument in favor of this approach is that it advances the goals of the current law without impacting data collection in unforeseen ways. Barton suggests that such an index would enable NAEP to report results in terms of SES quartiles (much the same way that the National Educational Longitudinal Survey, NELS, does). Further, it would allow the assessment to report cross-tabulations on distributions of students in the NAEP achievement level categories by SES. A good measure of SES would improve the monitoring of achievement gaps among various racial/ethnic groups, although
sample sizes may not be large enough within all ethnic groups or types of schools. Finally, a composite SES index may be beneficial to states and districts in the Trial District Assessment, enabling NAEP to compare the performance of groups of students with the same socio-economic status, which is a factor of high public and policy interest.

The argument against such an approach is that SES would continue to be measured indirectly, i.e., by using proxy variables, albeit through a composite index. There would also be disagreements about precisely which variables to include in the index and how to weight different factors. For example, Armor (D. J. Armor, personal communication, December 18, 2002) has suggested that two variables recently deleted from the NAEP student questionnaire be reinstated, namely, the number of siblings in the home and family status (student lives with both parents, mother or father, neither). These variables were dropped because of concerns about intrusiveness, but they may be of considerable importance in constructing an SES index. The Board will have to weigh the considerations involved, and may decide there is value in using them periodically or in limited samples.

A parent questionnaire has been proposed as a more reliable means of collecting SES data than relying on student reports, school records, or census data. Other National Center for Education Statistics surveys, for example, NELS and the Early Childhood Longitudinal Study, have employed parent questionnaires that ask direct questions regarding occupation and income.

However, the National Assessment of Educational Progress involves far more students than any of these research surveys. Accordingly, a parent questionnaire on NAEP would entail far more respondent burden and might arouse more controversy, making it more difficult to accomplish the primary mission of the assessment to measure student achievement. A parent questionnaire has been considered by NAGB in the past, but rejected as too burdensome and intrusive. Because these considerations are still persuasive, particularly as the scope of NAEP has expanded, no work should be undertaken on developing a parent questionnaire.

*In sum, because of its importance and the requirements of law, information on socio-economic status must be collected in all*
NAEP samples, although there may be some variation in the number of factors on which data are obtained. Research should be conducted into creating a composite index of SES.

A core of SES information should be collected in every assessment, such as type of community (e.g., central city, rural, etc.), poverty status (e.g., eligibility for free or reduced-price lunch and Title I participation), reading materials in the home, and level of parent education—though steps must be taken to ensure that such data are reliable. An expanded set of SES variables may be included periodically and administered to limited samples, including such factors as number of siblings and parents at home, possession of computers, and parent occupation.

NAEP should explore the use of an SES index derived from proxy variables currently in either the administration roster or student questionnaire. To the extent that an index can be sharpened by additional information from readily available sources, such as zip codes and/or census data, this option should be considered as well.

Public Policy Contextual Factors

For the past two decades NAEP has collected information on student, teacher, school, and beyond-school factors that are of interest to policymakers and the public. For students, some of these factors have included course-taking patterns, television watching, homework, and use of computers. For teachers, the contextual factors have included educational background, credentials, years of experience, and participation in professional organizations, to name a few.

The lists of factors have been long. They have become burdensome both to respondents and to the efficient scoring, analysis, and reporting of the NAEP survey. The way they have been reported—through simple one-way tabulations—has encouraged unwarranted conclusions about cause-and-effect relationships.

We propose a careful review of the contextual factors on which information is collected by NAEP to focus on the most important variables related to public policy. All such information must be clearly related to student achievement, as shown by other research. Data should be obtained periodically, on a rotating basis, over several NAEP cycles, although a limited set of factors may be included
in every assessment. Information should be collected at meaningful intervals in ways that may show significant patterns and change over time.

Two documents are helpful in surveying the research base and presenting alternatives for NAGB to consider. The first is *Monitoring School Quality: An Indicators Report* (Mayer, Mullens, and Moore, 2001), prepared by Mathematica Policy Research, Inc., for NCES. This report presents a research synthesis, indicating factors for which there is a research base showing a strong relationship to academic achievement. The synthesis, involving a review panel as well as statistical analyses, identifies the following as factors related to student achievement results: the academic skills of teachers, teacher assignments (such as out-of-field teaching), course content, student discipline and school safety, class size, and focus on academic achievement. Other sources of information are available on all of these factors, but only through NAEP can they be related to the achievement of broad groups of students over time.

The second document, *Making Connections* (Greenberg, Stancavage, Farr, and Bohrnstedt, 2001), was prepared for NCES by the American Institutes for Research and presents an elaborate typology of non-cognitive variables that could be measured by NAEP. It is organized into seven broad categories of non-cognitive information related to students, instructional content and practice, teachers, schools, school community factors, beyond school factors, and federal, state, and district policy. The listing goes beyond what NAEP can and should handle, but its discussion is thoughtful and the document is useful for planning.

**Subject-Specific Background Data**

For each subject assessed by NAEP, additional subject-specific background information has been collected from students, teachers, and schools. These data fall into the broad category of instructional content and practice. Under that umbrella come such topics as the curriculum taught, course offerings, class management and style, ability grouping, and modes of instruction. Subject-specific data collection has expanded enormously over the past two decades, and in recent years has included five to ten minutes of questions for students, about 30 minutes of questions for teachers, and 30 to 45 minutes of questions for school administrators.
Now is the time for these questions to be focused, limited, and prioritized. Future subject-matter frameworks adopted by the Governing Board should spell out clearly what these priorities will be.

A design for doing this was presented to the Board in the 1996 report of the Design/Feasibility Team of prominent researchers (Forsyth et al., 1996). The group recommended that a core set of non-cognitive questions should be administered to students each time a subject is assessed by NAEP. In addition, a more comprehensive questionnaire would be given whenever a new framework is introduced and repeated every eight to ten years. For example, an extensive set of background questions in reading and mathematics (grades 4 and 8) was administered in 2003, the baseline year for the No Child Left Behind legislation. Another complete set should be administered in mathematics in 2005 and in reading in 2009, the years in which revised frameworks are first used, and then should be repeated every eight years. In the intervening years, only the more limited core modules will be administered. Similar patterns should be established for the school and teacher questionnaires.

The NAEP assessments in other subjects, such as writing, science, history, geography, and civics, should have a core set of non-cognitive questions administered to the full sample, with longer, more extensive questionnaires being administered to smaller sub-samples. With states now required to participate in NAEP every two years, the total number of students tested has expanded substantially. This makes even more compelling the case for limiting the NAEP background questionnaires and rotating the background questions.

**NCES should prepare for Board review and approval a plan indicating the frequency, sample size, and schedule of rotation for all background variables and questions on which information is to be collected by NAEP. This should include both questionnaires and alternate data sources to obtain core reporting data, subject-specific information, and data on achievement-related contextual variables from a variety of NAEP samples—national only, national and state, and a subset of the national sample. The plan should indicate the frequency and schedule of rotation for each of the questions proposed. It should also indicate any questions needed for quality control purposes. The recommendations should be prepared with input from researchers and state policy analysts, as appropriate, and updated on a regular basis.**
Table 1 presents a model schedule for comprehensive and core sets of subject-related variables through 2013. It is based on the schedule of assessments approved by the Board in May 2003.

Table 1. Model Data Collection Schedule for Comprehensive and Core Sets of Non-Cognitive Variables by Subject Area

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Data Collection Year for Comprehensive Set of Variables</th>
<th>Data Collection Year for Core Variables Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreign Language (12)</td>
<td>2004, 2012</td>
<td></td>
</tr>
<tr>
<td>World History (12)</td>
<td>2010</td>
<td>TBD</td>
</tr>
<tr>
<td>Economics (12)</td>
<td>2006</td>
<td>TBD</td>
</tr>
<tr>
<td>Arts (8)</td>
<td>1997, 2008</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>2000, 2009</td>
<td>2005</td>
</tr>
<tr>
<td>US History</td>
<td>2001, 2006</td>
<td></td>
</tr>
<tr>
<td>Geography</td>
<td>2001, 2010</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Based on schedule approved by NAGB on May 17, 2003.
NAEP BACKGROUND QUESTIONS:
AN UNDERUSED NATIONAL RESOURCE

A Report to the National Assessment Governing Board by the Expert Panel on Strengthening the NAEP Background Questions

February 22, 2012

Chair: Marshall S. Smith

Members: Naomi Chudowsky, Alan Ginsburg, Robert Hauser, Jennifer Jennings, and Sharon Lewis
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive Summary</td>
<td>3</td>
</tr>
<tr>
<td>Introduction</td>
<td>9</td>
</tr>
<tr>
<td><strong>Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important Background Questions around High-Priority Areas</strong></td>
<td>13</td>
</tr>
<tr>
<td>Questionnaire Overview</td>
<td>13</td>
</tr>
<tr>
<td>Recommendation 1a. Continually review and refine a core and second-tier set of background topics and questions that are common across NAEP surveys</td>
<td>15</td>
</tr>
<tr>
<td>Recommendation 1b. Extend NAEP background questions to inform topics of current policy interest in the subjects assessed</td>
<td>16</td>
</tr>
<tr>
<td>Recommendation 1c. Select clusters of questions that collectively best measure different aspects of research-based theoretical frameworks for major educational topics</td>
<td>18</td>
</tr>
<tr>
<td>Recommendation 1d. Use consistency over time as a criterion to consider for question selection and wording</td>
<td>20</td>
</tr>
<tr>
<td>Recommendation 1e. Delete duplicative or low-priority questions to make time for the Panel's higher priority items</td>
<td>21</td>
</tr>
<tr>
<td><strong>Recommendation Area 2. Strengthen the Validity, Reliability and Coordination of the Measures and Clusters of Measures for the Background Questions</strong></td>
<td>22</td>
</tr>
<tr>
<td>Recommendation 2a. Improve the validity and reliability of the current measures NAEP uses for its mandated student reporting categories</td>
<td>23</td>
</tr>
<tr>
<td>Recommendation 2b. Enhance the validity of student responses at different grade levels</td>
<td>25</td>
</tr>
<tr>
<td>Recommendation 2c. Accurately measure the multi-dimensional nature of learning-to-learn skills including student learning behaviors, motivation and expectations</td>
<td>26</td>
</tr>
<tr>
<td>Recommendation 2d. Improve question reliability by replacing imprecise phrases such as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or more a week”</td>
<td>27</td>
</tr>
<tr>
<td>Recommendation 2e. Coordinate NAEP background questions with those asked on international or domestic surveys</td>
<td>28</td>
</tr>
<tr>
<td>Recommendation 2f. Build on current NCES cognitive interview techniques by using cognitive laboratories, such as small standing panels, to field test questions to establish their validity and reliability</td>
<td>29</td>
</tr>
<tr>
<td><strong>Recommendation Area 3. Reform NAEP Sampling to Enhance the Scope of the Background Questions While Maintaining Sampling Accuracy</strong></td>
<td>30</td>
</tr>
<tr>
<td>Recommendation 3a. Support NCES's exploration of a spiral sample methodology to expand the scope of background questions, while assessing the possible loss in the representativeness of disaggregated data</td>
<td>31</td>
</tr>
<tr>
<td>Recommendation 3b. Consider other item-sampling reforms to obtain the needed questionnaire time including lengthening the student survey; establishing a 4-year interval between administration of some background questions; and pooling item responses across survey administrations</td>
<td>32</td>
</tr>
<tr>
<td><strong>Recommendation Area 4. Reinstitute the Analysis and Regular Reporting of the NAEP Background Questions</strong></td>
<td>32</td>
</tr>
</tbody>
</table>
Recommendation 4a. Prepare special reports highlighting the background question findings.

Recommendation 4b. Prepare an online compendium of key background indicators for States and participating urban districts.

Recommendation 4c. NAEP’s reports should not indicate causal interpretations using the background questions. However, the NAEP data offer some unique advantages for generating relationships and hypotheses about factors that may be associated with performance and these findings should guide more rigorous in-depth follow-on analyses.

Recommendation 4d: NAEP should encourage others to conduct exploratory studies of the background questions. However, the NAEP data offer some unique advantages for generating relationships and hypotheses about factors that may be associated with performance and these findings should guide more rigorous in-depth follow-on analyses.

Recommendation 4e. Further improve the powerful online NAEP tools for data analysis.

5. Implementing the Panel Recommendations

Recommendation 5a. Exploit existing background data through special reports focused on issues and topics informed by background questions.

Recommendation 5b. Initiate a set of activities to build clusters of core and second-tier questions around high-priority topics for the 2015 NAEP administration.

Recommendation 5c. Further improve the usability of the Data Explorer and other NAEP online tools, which are already of high quality.

Recommendation 5d. Champion implementation by creating a single NAGB subcommittee responsible for the background questions; provide adequate resource support, while ensuring efficient resource use; and publicize background question products and findings.

REFERENCES

EXPERT PANEL MEMBERSHIP AND AFFILIATIONS
Executive Summary

For more than four decades the National Assessment of Educational Progress (NAEP) has tracked the achievement of U.S. students in major academic subjects. This national resource is the only assessment that states and now many urban districts can look to as an objective yardstick of their performance over time, relative to national benchmarks, and compared with other jurisdictions. Less known, but complementing the NAEP assessments, is a rich collection of student, teacher and school responses to background questions that can help in understanding the context for NAEP achievement results and give insights into how to improve them.

Currently, the NAEP background questions are a potentially important but largely underused national resource. The background questionnaires have been cut back over the past decade. They now cover only a small fraction of important student, teacher, and school issues and have been little used in recent NAEP reports, in contrast to the first state-level NAEP Report Cards in the early 1990s.

NAEP should restore and improve upon its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, these data could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education including:

- Describing the resources available to support learning (opportunity-to-learn) for students with differing home backgrounds and over time.
- Tracking progress in implementing key instructional, curricular, and technological changes and educational policy initiatives, such as the Common Core standards.
- Monitoring student motivation and out-of-school learning as research-based factors affecting student achievement.
- Benchmarking high-performing states and urban districts and those with high achievement growth to identify factors that differentiate high-performers from lower-performers on NAEP. This domestic effort would parallel the extensive reporting of background variables in PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) that have become starting points for U.S. international benchmarking analyses to describe the characteristics of high-performing and low-performing education systems.

The panel proposes building a strategy to make the NAEP background questions an important national resource for educators, policymakers, and the public. The panel sees the need to expand the scope and quality of the existing questions, move into important new areas directed by research and policy, make better use of the questions though regular publications, and improve the capacity for analysis by users around the world.
We offer recommendations in four areas (see Exhibit A):

1. Ask Important Questions.
2. Improve the Accuracy of Measures.
3. Strengthen Sampling Efficiency.
4. Reinstitute Meaningful Analysis and Reporting.

### Exhibit A. Expert Panel Recommendations to Strengthen NAEP Background Questions in Four Areas

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Core questions</td>
<td>Valid</td>
<td>Spiral sampling</td>
<td>Special background question reports</td>
</tr>
<tr>
<td>Rotated questions</td>
<td>Reliable</td>
<td>Extended questionnaire time</td>
<td>Online compendium of responses</td>
</tr>
<tr>
<td>Policy questions</td>
<td>Coordinated (with domestic and international surveys)</td>
<td>Alternate surveys</td>
<td>Report descriptive not causal findings</td>
</tr>
<tr>
<td>Theoretical frameworks</td>
<td>Cognitive labs</td>
<td>Pooling item responses across surveys</td>
<td>Externally conducted research</td>
</tr>
<tr>
<td>Consistent questions</td>
<td></td>
<td></td>
<td>Improve online tools</td>
</tr>
<tr>
<td>overtime</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete duplicative or low-priority questions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Establish a single NAGB committee overseeing background questions
- Review budget including need for staff to implement recommendations

### Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important Background Questions around High-Priority Areas.

To the extent that you don’t ask and analyze important questions, you can’t expect to get back important answers. The panel recommends identifying topics falling into three question groups.

- **A common core** set of background questions to include three question clusters: (1) the congressionally required student background characteristics; (2) instructional practices and school learning opportunities and resources; and (3) student motivation and control over the environment.

- **A second tier** of priority background question clusters would be rotated across assessment cycles. Important topics that might be explored include school-parent cooperation, school climate and discipline, school administration including support for learning, and out-of-school learning time.

- **A third tier** would be a set of policy issues that would be examined for six years and then rotated out with new ones added. For example, the initial set might start
with questions on implementation of the Common Core standards. Two years later, a set of questions or module on teacher evaluations could be added, and two years after that a module on project-based or online learning.

Once question topics are identified, the panel urges the selection of clusters of questions that collectively best portray different important aspects of research-based theoretical frameworks for the major educational topics. Such frameworks should be published, as they are for TIMSS and PISA, to explain the theoretical rationale and research evidence that underlie the selection of the background questions and their connection to student learning and achievement.

The Panel recommends two additional considerations to maximize the information worth of the questions chosen. The first is to pay greater attention to the consistency of question selection and wording to produce reliable time-series that measure change over time. A review of 400 questions asked about teachers found that about 300 are no longer used, with many replaced by just slightly different wording. A second recommendation is to balance the number of questions asked about a topic with the information value gained. Eight questions are asked about technology use in mathematics but there are no questions about student expectations despite the strong research connection with achievement.

**Recommendation Area 2. Strengthen the Validity, Reliability and Coordination of the Measures and Clusters of Measures for the Background Questions.**

The panel urges attention to strengthening the validity, reliability and coordination of NAEP background questions. An important first step in this overall effort would be to improve the validity, reliability and coordination of the current measures NAEP uses for its mandated student reporting categories. The panel strongly supports the current review of the SES variables as it is critical to respond to the known limitations of the school-lunch proxy. These problems will worsen with expansion of the Department of Agriculture state pilots, which allow whole-school eligibility for schools serving concentrations of low-income students. The panel also believes that an expanded cognitive interview capability, such as a small standing panel of respondents to test out questions, would improve question validity and reliability. We recognize that this may increase costs but it would help make NAEP a better source of information.

The panel recommends improving question wording by replacing imprecise terms such as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or more a week.” Furthermore, major information benefits would accrue from coordinating the NAEP background questions with those asked on other international and domestic surveys. To illustrate, the PISA international survey covers number of hours of math instruction in-school and out-of-school; NAEP only asks about days taught math in-school and only about participation in math instruction outside of school and nothing about frequency.
Recommendation Area 3. Reform NAEP Sampling to Enhance the Scope of the Background Questions While Maintaining Sampling Accuracy.

The panel recommends that NAEP should consider expanding the depth of its background questions through a variety of strategies including spiral sampling (already under study), expanded questionnaire time and rotating background questions across samples. The panel notes that the depth of student information in particular is limited by the ten-minute questionnaire time limit compared with 30 minutes used for TIMSS and PISA. A combination of these strategies would allow NAEP to obtain far richer information while maintaining sampling accuracy and still keeping respondent burden to acceptable levels.

Recommendation Area 4. Reinstitute the Analysis and Regular Reporting of the NAEP Background Questions.

This set of recommendations would bolster the analysis and reporting of the background questions by means of separate publications, online tables, and improvements to the Data Explorer. The recommendations also include a reiteration of current policy to not use causal interpretations of point-in-time data.

The panel strongly recommends NAEP consider two initial special reports, one organized around learning opportunities in school and a second around learning opportunities and conditions out of school. Exhibit B displays an illustrative overview table for in-school learning opportunities for math that suggests the rich potential information payoffs from background question analyses. A third benchmarking report should also be considered that explores the correlates of high-performing states and districts or those with high achievement growth. These synthesis reports would also provide a way to assess the information value of current and past questionnaire items.

Implementation of Recommendations

The panel urges the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES) to move quickly to begin implementing its recommendations to make the background questions a more useful resource, while also recognizing that implementation will take time.

Initial implementation should be undertaken through a three-part plan:

- Immediately produce special reports on the background data that analyze the considerable quantity of data already collected, but is largely unreported and unanalyzed.
Exhibit B. Illustrative Table of Background Question Indicators with a Grade 8 Math Focus: School Districts Participating in the 2011 Trial Urban Development Assessment

| Jurisdictions         | Grade 8 All Students | Eligible for National School Lunch | Grade 8 Students Absent 5 or more days last month | Grade 8 Students in Algebra | Grade 8 Students 5 or more Hours of Math Per Week | Grade 8 Students 1 Hour or More Math Homework | Grade 8 Does Math At An After School or Tutoring Program | Grade 8 Has Math Major/Minor/Special Emphasis | Grade 8 Full-time Math Specialist At School | Grade 8 Assigned to Math By Ability | Grade 8 26% Students in Math Class | Grade 8 Students in Algebra | Grade 8 Students Absent 5 or more days last month | Grade 8 Students 5 or more Hours of Math Per Week | Grade 8 Students 1 Hour or More Math Homework | Grade 8 Does Math At An After School or Tutoring Program | Grade 8 Has Math Major/Minor/Special Emphasis | Grade 8 Full-time Math Specialist At School | Grade 8 Assigned to Math By Ability | Grade 8 26% Students in Math Class | Grade 8 Computers Available to Teachers and Students |
|-----------------------|----------------------|-----------------------------------|-----------------------------------------------|----------------------------|-----------------------------------------------|-----------------------------------------------|------------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| National              | 284                  | 44                                | 7                                             | 42                          | 37                                           | 17                                           | 21                                           | 17                                           | 38                                           | 37                                           | 37                                           | 76                                           | 45                                           | 84                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Albuquerque           | 275                  | 60                                | 8                                             | 37                          | 65                                           | 13                                           | 20                                           | 27                                           | 33                                           | 32                                           | 66                                           | 59                                           | 77                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Atlanta               | 266                  | 82                                | 5                                             | 27                          | 75                                           | 38                                           | 57                                           | 57                                           | 95                                           | 61                                           | 59                                           | 37                                           | 90                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Austin                | 287                  | 59                                | 8                                             | 23                          | 61                                           | 27                                           | 30                                           | 42                                           | 57                                           | 58                                           | 53                                           | 52                                           | 89                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Baltimore City        | 261                  | 85                                | 9                                             | 46                          | 93                                           | 41                                           | 28                                           | 28                                           | 78                                           | 53                                           | 86                                           | 37                                           | 71                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Boston                | 282                  | 76                                | 9                                             | 66                          | 76                                           | 39                                           | 30                                           | 13                                           | 69                                           | 12                                           | 61                                           | 47                                           | 56                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Charlotte             | 285                  | 52                                | 8                                             | 35                          | 87                                           | 18                                           | 29                                           | 44                                           | 47                                           | 33                                           | 86                                           | 76                                           | 70                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Chicago               | 270                  | 84                                | 4                                             | 32                          | 67                                           | 47                                           | 37                                           | 23                                           | 84                                           | 20                                           | 45                                           | 65                                           | 88                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Cleveland             | 256                  | 108                               | 11                                            | 29                          | 69                                           | 33                                           | 25                                           | 6                                            | 58                                           | 14                                           | 51                                           | 44                                           | 90                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Dallas                | 274                  | 85                                | 7                                             | 32                          | 46                                           | 27                                           | 39                                           | 61                                           | 66                                           | 13                                           | 45                                           | 24                                           | 57                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Detroit               | 246                  | 79                                | 17                                            | 24                          | 81                                           | 46                                           | 37                                           | 11                                           | 83                                           | 39                                           | 48                                           | 85                                           | 61                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| District of Columbia (DCPS) | 255                       | 70                                | 12                                            | 53                          | 65                                           | 29                                           | 39                                           | 57                                           | 68                                           | 40                                           | 53                                           | 20                                           | 86                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Fresno                | 295                  | 88                                | 10                                            | 51                          | 32                                           | 11                                           | 26                                           | 6                                            | 37                                           | 23                                           | 91                                           | 75                                           | 59                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Hillsborough County (FL) | 282                        | 54                                | 9                                             | 87                          | 20                                           | 13                                           | 22                                           | 40                                           | 35                                           | 29                                           | 95                                           | 3                                             | 86                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Houston               | 279                  | 76                                | 6                                             | 29                          | 63                                           | 26                                           | 37                                           | 56                                           | 63                                           | 25                                           | 84                                           | 58                                           | 68                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Jefferson County (KY) | 274                  | 60                                | 7                                             | 40                          | 68                                           | 14                                           | 20                                           | 21                                           | 34                                           | 36                                           | 77                                           | 80                                           | 80                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Los Angeles           | 261                  | 82                                | 6                                             | 67                          | 44                                           | 40                                           | 27                                           | 39                                           | 67                                           | 37                                           | 75                                           | 52                                           | 74                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Miami-Dade            | 272                  | 72                                | 5                                             | 36                          | 43                                           | 47                                           | 25                                           | 38                                           | 72                                           | 25                                           | 90                                           | 13                                           | 88                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Milwaukee             | 254                  | 81                                | 13                                            | 30                          | 78                                           | 43                                           | 31                                           | 37                                           | 74                                           | 82                                           | 28                                           | 86                                           | 78                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| New York City         | 272                  | 87                                | 10                                            | 28                          | 83                                           | 26                                           | 39                                           | 36                                           | 65                                           | 36                                           | 60                                           | 83                                           | 79                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| Philadelphia          | 265                  | 88                                | 10                                            | 34                          | 89                                           | 27                                           | 27                                           | 24                                           | 54                                           | 32                                           | 30                                           | 76                                           | 89                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |
| San Diego             | 276                  | 60                                | 8                                             | 69                          | 48                                           | 13                                           | 27                                           | 11                                           | 40                                           | 17                                           | 78                                           | 72                                           | 80                                           |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |                                              |

Source: NAEP Data Explorer

- Move quickly to initiate a long-term effort to improve the relevance, quality, coherence, and usefulness of a core and rotated set of background variables while implementing recommended improvements to improve measurement accuracy and sampling efficiency.

- Further improve the usability of the Data Explorer and other NCES online tools, which are already valuable analytic supports.

The panel suggests that NAGB establish a separate standing committee to review all background questions and plans to improve their use. Currently, the Board’s responsibilities for background questions are divided between two of its standing committees. These subgroups do not coordinate their work and the background questionnaires are of secondary interest to both of them. A unified standing committee should regularly monitor and report on implementation of the panel’s recommendations by NCES and Governing Board staff.

In addition, the panel believes that the background questions and how they used in NAEP reporting warrant a periodic, rigorous, and independent evaluation similar to that conducted in the past on NAEP cognitive assessment items.

The panel recognizes that implementing its recommendations will involve resource considerations in terms of time, money, and personnel. One approach to this problem may be to reduce costs in certain areas. For example, efforts should be made to eliminate
lower-priority activities, such as the duplicative collection of racial data and the disproportionate number of questions asked in areas such as technology. Another approach should be to make a clear and powerful case for the usefulness of having a coherent set of relevant and valid background variables to help explain NAEP results and to take this case to the Department of Education, the Office of Management and Budget (OMB), and Congress.

In conclusion, the NAEP background questions are a unique national information resource. The Governing Board and NCES have a responsibility to develop this resource to better understand academic achievement and the contexts in which it occurs and, hopefully, to help spur educational improvement.
Introduction

The National Assessment of Educational Progress (NAEP) is a unique American education resource. For more than four decades the assessment has tracked the achievement of U.S. students in major academic subjects. This national resource is the only assessment that states and now many urban districts can look to as an objective yardstick of their performance over time, relative to national benchmarks, and compared with other jurisdictions.¹

Representative samples of students regularly take NAEP assessments in reading, mathematics, science, and writing and the national, state, and urban district levels. Other subjects, including U.S. history, civics, and the arts, are tested at the national level only. Less known, but complementing the NAEP assessments, is a potentially rich collection of student, teacher and school responses to background questions that can help in understanding the context for NAEP achievement results and give insights into how to improve them.

Currently, the NAEP background questions are a potentially important but largely underused national resource. The background questions have been cut back over the past decade. They now cover only a small fraction of important student, teacher and school issues, and have been little used in recent NAEP reports, in contrast to the first state-level NAEP Report Cards in the early 1990s.

NAEP should restore and improve upon its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, these data could provide rich insights into important questions about the nature and quality of American primary and secondary education. What are the racial, ethnic and economic characteristics of schools at different achievement levels? What are the sources of curriculum content? What resources are available for students? What are the common instructional approaches teachers employ, and how do they adjust approaches to differing student needs? What preparation and training do teachers receive? How is teacher performance evaluated?

In turn, the answers to these survey questions can support important NAEP analyses. The analyses should focus on the unique advantages of NAEP for collecting data and trends over time on education-related background factors paired with achievement results that are representative of states and many urban districts. The following three examples

¹ Although this report focuses on the lack of reporting the background variables for the main NAEP, a similar weakness occurs in not reporting the background variables for the long-term trend NAEP. The report on the 2008 long-term trend assessments did include data on higher level course taking in math in 2008 in relation to that year’s NAEP scores, but surprisingly did not report results for earlier years, although available.
illustrate potentially significant descriptive findings from the NAEP background questions for mathematics with respect to:

- Describing the resources available to support learning (opportunity-to-learn) for students with differing home backgrounds and over time.
  - In Arizona, a Hispanic grade-8 student is only 57 percent as likely to have a teacher of mathematics who has a major in mathematics as a white grade-8 student. In California, their chances are nearly equal.

- Tracking progress in implementing instructional, curricular, and technological changes and key education policy initiatives.
  - The proportion of students in schools with no eighth-graders enrolled in algebra is 15 percent nationally. Among urban districts, Miami-Dade and Houston have only 5 percent of their students in schools without a grade-8 algebra course, but Detroit and Milwaukee have over 80 percent of eighth-graders in such schools.

- Monitoring student motivation and out-of-school learning as factors affecting student achievement.
  - More than 45 percent of the grade 4 students in several Southern states (Louisiana, South Carolina and Texas) participated in after-school math instruction. But in several highly rural states (Maine, Oregon and Vermont) the participation rate in after-school math instruction was only about 25 percent.

Moreover, the extensive reporting of the background variables in PISA and TIMSS have become starting points for U.S. international benchmarking analyses to describe the characteristics of high-performing education systems (Darling-Hammond, 2010). These data have been used to examine characteristics of high-performing systems, such as Singapore and Korea, and to study the nature of instruction in subjects such as math and science, where the U.S. performs poorly. In a similar fashion the NAEP data could be used to guide benchmarking of high-performing states and urban districts or jurisdictions experiencing substantial performance growth. This benchmarking activity would be a means to generate hypotheses for further verification though in-depth study. Specific examples of the use of NAEP background questions for domestic benchmarking might include examining:

- A high overall-performing state such as Massachusetts or a state like Texas that has a relatively small white-Hispanic performance gap compared with other states.
- A high-performing district such as New York City that has low-income students achieving above the national average for all low-income students in both reading and math at grades 4 and 8.
- The nearly one standard deviation growth in grade 4 math since 1990 and the instructional, curriculum and teacher changes that occurred over this period.
The panel recognizes the justifiable concern over misuse of the NAEP background variables in making causal interpretations. NAEP is not able to reduce countervailing explanations for causation like a well-designed experiment. Also, successive NAEP assessments will sample different students in the same grade, so the data are not a measure of change over time for the same students as in a true longitudinal design. However, the panel believes that a valid concern over causal interpretations has led to a serious and unjustified overreaction. NAEP’s national and state representative data uniquely address many important descriptive questions. These data can track progress on variables shown by research to be important for achievement. The NAEP background questions can inform national policies by providing descriptive data about the quality of implementation. Also, because NAEP is already in the schools to administer its assessments, data can be collected at relatively low cost compared with other survey vehicles.

Yet for the past decade NAEP has stopped publishing all but the most minimal background information.

- NAEP no longer systematically reports on the responses to the background questions when publishing its assessment results, except for the congressionally required student reporting categories (e.g., race/ethnicity, low-income).
- In-depth special reports using the background questions are rare (e.g., the 2010 report on American Indian Educational Experiences was an exception).
- Data are made available almost entirely through an online database called the NAEP Data Explorer. This is a useful tool, but it is not a substitute for carefully prepared summary data tables and analyses. Most educators, policy makers and members of the public do not have the time or inclination to master use of the Data Explorer, but many would pay attention to focused reports and make use of summary tabular information.

Reporting the background questions would be a great service to the nation in identifying and tracking important national and state trends in education. Here, the panel finds that the NAEP background questionnaires severely limit their potential usefulness by not explicitly asking questions about the progress and challenges of implementing key national policies in different states and urban districts. Yet the *NAEP Background Information Framework* (2003), which sets out principles to guide background question selection and reporting, explicitly recognizes that the background questions should “focus on the most important variables related to public policy.”

NAEP’s de-emphasis of the background questions is in marked contrast to the significance that all the major international surveys – PISA (Program for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), and PIRLS (Progress in International Reading Literacy Study) – give to background

---

2 In 2011 NAEP began to use the background variables again in its main assessment reports, but with only a single background table related to instruction for each subject and grade. The 2010 Civics, Geography and U.S. History reports also contained a background table related to instruction for the different grades.
variables in participating countries.

The panel believes NAEP should return to its earlier practice of making much greater use of background data, but do so in a more sound and research-supported way. With proper attention, the questions could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education and the context for understanding achievement and its improvement. The panel believes there is a need to expand the scope and quality of the existing questions, move into important new areas directed by research and policy, make better use of the questions though regular NAEP publications, and improve the capacity for analysis by data users.

To do so the panel has developed recommendations for improvements in four areas:

1. Ask Important Questions.
2. Improve the Accuracy of the Measures.
3. Strengthen Sampling Efficiency.
4. Reinstitute Meaningful Analysis and Reporting.

Within each area, Exhibit 1 identifies the specific individual recommendations.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Core questions</td>
<td>Valid</td>
<td>Spiral sampling</td>
<td>Special reports</td>
</tr>
<tr>
<td>Rotated questions</td>
<td>Reliable</td>
<td>Extended questionnaire time</td>
<td>Online compendium of responses</td>
</tr>
<tr>
<td>Policy questions</td>
<td>Coordinated (with domestic and international surveys)</td>
<td>Alternate surveys</td>
<td>Report descriptive not causal findings</td>
</tr>
<tr>
<td>Theoretical frameworks</td>
<td>Cognitive labs</td>
<td>Pooling item responses across surveys</td>
<td>Externally conducted research</td>
</tr>
<tr>
<td>Consistent questions overtime</td>
<td></td>
<td></td>
<td>Improve online tools</td>
</tr>
<tr>
<td>Delete duplicative or low-priority questions</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The panel recognizes that these recommendations would require commitments of resources and that the Governing Board and the Commissioner of Education Statistics are in the best position to decide on any tradeoffs between existing and proposed features of NAEP that may be required within NAEP’s budget.
Recommendation Area 1. Identify Core, Rotated and Theoretically Coherent Groups of Important Background Questions around High-Priority Information Areas

To the extent that you don’t ask and analyze important questions you can’t expect to get back important answers. This section recommends strategies for focusing clusters of questions on important information topics within the confines of NAEP questionnaire timelines and administration procedures. Consistent with the NAEP framework, important questions are ones that would primarily focus on the factors that research has shown are related to student achievement. Background questions would also address the implementation of major national policies where NAEP surveys can provide a view from the field state-by-state. In this way, NAEP can report on the distributions and trends of many of the factors and policies important for student achievement.

Questionnaire Overview

With each administration of the subject area assessment, NAEP includes separate student, teacher and school background questionnaires. Although a few questions about subgroups are specified in the NAEP legislation, the Governing Board has the discretion to determine most questions. Exhibit 2 displays the overall number of questions and general question content for each of the three respondent questionnaires on the most recently-reported reading and mathematics surveys.

| Exhibit 2. Overview of the Most Current NAEP Mathematics and Reading Background Questionnaires for Students, Teachers and Schools |
|---|---|---|---|
| **Students 10 Min** | **Teachers 30 Min** | **Schools 30 Min** |
| Questions: - Student & family background and out-of-school learning - Subject specific: self-perception and school courses content | Questions (subject specific): - Teachers Background: education and training; - Classroom Organization and Instructional practices | Questions: - School Characteristics (including a special charter school survey) - Subject specific: course, student placement, staff composition, training, technology |
| Math: 2011 | 31 | 30 | 40 | 48 | 31 | 39 | 49 | 48 |
| Reading: 2011 | 32 | 26 | 34 | 30 |  |

A 10-minute student questionnaire consisting of approximately 30 questions asks about family background, school and home experiences, and out-of-school learning activities.

- Since NAEP does not administer a questionnaire to survey parents, the student questionnaire is the primary source of information on students’ home characteristics and out-of-school learning activities. (School records do provide an alternative source for race, ethnicity and school lunch eligibility data).
- With respect to socio-economic status, grade 4 students are only asked about household items (computers in the home, numbers of books). Students in grades 8 and 12 are also queried about their mother’s and father’s highest level of education.
- A few questions are asked about students’ out-of-school learning-related activities -- talk about things studied in school, read for fun on your own time, or studying and reading at an after-school program.
- A few items are included about student self-perception and enjoyment of a specific subject, for example whether reading and math are favorite subjects.
- Students are asked a number of questions about their classes in the subject assessed – for example, the frequency of reading aloud and discussing what they read in class, and in math many questions about using technology (calculators, graphing programs and spreadsheets).

A 30-minute teacher questionnaire of 30-40 questions is filled-out by the teacher in grade 4 or 8 in the subject assessed, usually the classroom teacher at grade 4 and the English or mathematics teacher at grade 8. This questionnaire covers:

- Teacher background information on race/ethnicity, education, certification and experience and professional development.
- Classroom organization items about class size, hours of instruction and ability grouping.
- Instructional items about topic emphasis, instructional approach, homework, evaluating student progress and access to resources and technology. The math questionnaire includes extensive questions about calculators of all types, computers, the Internet and CD-ROMs.

A 30-minute school questionnaire of about 40 questions covers:

- Overall school characteristics including grades, status as a charter, student composition and turnover, teacher absenteeism, volunteerism, and Title I federal program participation.
- Subject-specific items about specialist staff, structuring of content with standards and assessments, resource availability with emphasis on technology,
- Special charter school questionnaire about legal status and focus of charter.

Looking across the surveys, several issues of questionnaire coverage emerge:

- The student questionnaire includes items obtainable elsewhere and may be duplicative. For example, student-reported information on classroom instructional approaches overlaps with information on the teacher questionnaire.
Although the three surveys collectively cover a broad range of important background topics, the surveys omit a few topics with a strong base supporting their relationship to achievement. Two examples are the degree to which schools reach out to parents, and school discipline and the climate for learning.

The questionnaires largely ignore major national policy issues prominent over the last decade involving the response to federal mandates for state-based student testing and high-stakes accountability.

The panel believes there is a need to address these and other issues of questionnaire content through a systematic process for identifying topics and questions that best relate to understanding NAEP student achievement results without being excessively burdensome or invasive.

**Recommendation 1a. Continually review and refine a core and second-tier set of background topics and questions that are common across NAEP surveys.**

- NAEP should build on its current process for specifying a common core set of background questions to include three question clusters: (1) the congressionally required student background characteristics; (2) instructional practices and school learning opportunities and resources; and (3) student motivation and control over the environment.
- NAEP should develop a second tier of priority background question clusters that could be rotated across assessment cycles. Important topics that might be explored include school-parent cooperation, school climate and discipline, school administration and support for learning; and out-of-school learning time.
- NAEP should prioritize core and second tier items in terms of information value and respondent time, select high-priority items, and eliminate current low-priority items.
- NAEP should regularly publish its background questionnaires and provide justifications for all questions asked in terms of research and policy. Core and second-tier background questions should be identified.

**Discussion**

This recommendation would expand NAEP’s current set of core background questions focused primarily on the congressionally required student subgroups. The panel recommends including as an additional part of the core, a second cluster for instructional and other school learning opportunities. This cluster would allow examination of student learning environments by describing the curriculum, instructional approaches, and teacher qualifications. Many of these types of questions are now included in the teacher questionnaire and would be folded into this category.

A third core cluster of core questions is recommended to cover the area of student motivation and control over the environment. Measures such as whether students believe that success depends more on ability than effort or students’ locus of control have been
documented over several decades as strongly related to academic performance (Coleman, 1966; Chen & Stevenson, 1995). Also, students’ educational expectations predict their educational achievement and occupational expectations predict occupational attainment (ETS, 2010). When good teachers and a positive school environment influence student motivation and expectations this in turn will lead to improved achievement.

A second tier set of question clusters is proposed to focus on items for which there is strong research backing of their relation to achievement, but for which rotated items across alternate assessments (e.g., every four years) would be acceptable. As noted above, these second tier clusters could describe school-parent cooperation, school climate and discipline, school administration and support for learning; and out-of-school learning time. Specific clusters should vary across time as achievement levels and educational practices and policies change.

Together these clusters of items would view gains in school achievement as driven by a simple theory that sees gains in learning as a function of the curriculum, learning time, quality of instruction and student motivation These core and second-tier clusters meet the principle in the Board’s Background Information Framework that “The information obtained be of value in understanding academic performance and taking steps to improve it” (2003 Background Information Framework).

The Panel recognizes that in defining these clusters NAEP will have to establish tradeoffs in terms of meeting the constraints of questionnaire length and cost. These decisions should be based on the priority of a question or question cluster in terms of information value balanced against respondent burden and costs. To make room for new high-priority items NAEP should consider eliminating or reducing low-value or duplicative questions as noted below. Time constraints may also be addressed by rotating questions on alternate survey administrations (i.e., four-year intervals) NAEP also constrains the student questionnaire length to ten minutes when TIMSS even at grade 4 is 30 minutes.

**Recommendation 1b. Extend NAEP background questions to inform topics of current policy interest.**

- Implementation of this recommendation could focus on three rotating sets of policy questions each extended over a six-year period. For example, the initial set might start with questions on implementation of the Common Core standards. Two years later, a set of questions or module on teacher evaluations would be added, and two years after that a module on project-based or online learning. After six-years, questions on a new policy issue would be introduced to replace the first. Using this approach each of the question sets would have three observations over the six-year time.

- The panel concurs with the 2003 Background Report caution to include only policy-relevant questions that are answered on the basis of fact rather than opinion. That is, the responses to policy-relevant questions should be objective and not reflect personal beliefs. Questions should ask about policy responses, such as training received to understand new standards or the extent to which new
standards have changed instructional content or approaches. Questions should not elicit judgments about personal policy preferences.

- The policy information collected should not duplicate what can be obtained from other sources, such as description of the law or state implementation plans. Instead, NAEP is uniquely positioned to obtain ground-level information by surveying teachers and principals about policy implementation and challenges. This would not be designed nor suited to address legal compliance with federal policy, which is the role of program monitoring. Instead, it would provide information to improve the quality of policy and practice.

- Indeed, many national policies such as the Common Core are not federal at all. In this example, NAEP would track the implementation of standards in the Common-Core states, identifying changes in instructional content and emphasis compared with non-Common cores states. NAEP teacher surveys could further address the extent of staff training and understanding of the new standards and instructional challenges.

Discussion

The panel’s review of the current background questionnaires concluded that they insufficiently incorporate questions about school and teacher responses to policies that could strengthen policy implementation and promote student achievement. Examples of policy-relevant issues that NAEP could but currently does not report on include characteristics of instruction in schools that made adequate yearly progress, the degree to which teacher evaluations incorporate student outcomes, or the nature and extent of coordination between school and after-school instruction.

This recommendation would reinforce NAGB (2003) guidance that identifies “informing educational policy” as a reason for collecting non-cognitive information. It would also support NCES commitments to convening “a policy/contextual issues panel when needed to identify policy/contextual issues that NAEP might address in the future, and to outline the relevant constructs and identify data needed to address these issues.”

The panel recognizes that policy issues should be regularly refreshed as new policies emerge that build on or replace prior strategies. Our proposal aims for roughly a six-year issue cycle to give policies sufficient time to be implemented and effect improvements. The three policies suggested in the recommendations reflect the likely timeframe of implementation. The initial focus is on Common Core implementation, which is already underway in many states. Next a question set would be added on how schools evaluate their teachers. This would include questions on how evaluations of teachers take into consideration the outcomes of a teacher’s students, as this relatively new policy takes

---

hold. The third suggestion of project-based and online learning reflects expectations that the role of technology in providing instruction will substantially increase.

**Recommendation 1c. Select clusters of questions that collectively best measure different aspects of research-based theoretical frameworks for major educational topics.**

- Such frameworks should be published, as they are for TIMSS and PISA, to explain the theoretical rationale and research evidence that underlie the selection of the background questions and their connection to student learning and achievement. NAEP unlike TIMSS or PISA currently fails to publish clearly defined, research-based theoretical frameworks that guide question selection. Accordingly, NAEP should make explicit and publically available the underlying theoretical frameworks for question selection. The Panel recognizes that the research basis for the theoretical justifications may be less than perfect and are sometimes subject to post-hoc rationalizations. Nonetheless, the objective syntheses of research across a variety of settings to form theoretical frameworks for clusters of variables significantly enhances the odds of collecting survey information that will accurately and usefully inform practice and policy.
- Background questions should fit together to portray different important aspects of a topic (e.g., the different dimensions of SES).

**Discussion**

The 2003 Background Information Framework for NAEP states the principle that “Background information shall provide a context for reporting and interpreting achievement results and, as the statute provides, must be “directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information.” NAEP to its credit employs panels involving contractors and multiple external groups in its question development.

However, currently, NAEP does not formally publish an accompanying document with each assessment that lays out the theoretically-based frameworks that underlie the selection of the background questions and their connection with learning and achievement.

NCES has a good start toward building the necessary research foundation for developing such frameworks in the papers prepared by the Education Testing Service (ETS). ETS (2010) has developed three in-depth literature reviews, one each to support the topics currently or potentially addressed in the student, teacher and school questionnaires. The student and school questionnaire reviews also compare the current NAEP content items with the content measured in other large-scale national and international assessments.

The panel’s proposal would build-on the current literature reviews by:
• Using the research to develop theoretical frameworks that identify for major topics the component variables around which to build clusters of questions. The current ETS literature reviews although useful, are largely a description of discrete findings. Exhibit 3 is an example of how PISA presents a research-based, theoretical framework to organize background questions around the components of student engagement in reading and reading strategies. In this example, PISA operationalizes engagement in reading in terms of five components: reading for school, enjoyment of reading, time spent reading for enjoyment, diversity of reading materials, and diversity of online reading activities. Multiple questions then ask students about their reading behaviors with respect to each component.

• Organizing literature reviews around topics, which is preferable to the current organization around three separate questionnaires. Some topics may cut across the student, teacher and school questionnaires. For example, the current ETS literature review considers family involvement only in terms of the student questionnaire and the items describing home learning activities and resources. A broader research-based theoretical framework around the issue of parental involvement would extend the construct to include how teachers and schools reach out and support families, not just what families do by themselves. Indeed, Title I longitudinal evaluations have shown that student achievement improves when schools reach out and support parental involvement. (USED, 2001).

Once developed, these research-based frameworks would form the basis for developing valid and reliable questions to measure the different aspects of a topic domestically and to coordinate measurement with major international surveys. (Section 2 below).

Exhibit 3. PISA Analytic Framework for Student Engagement in Reading and Learning Strategies to Inform Decisions about Improving Reading

Source. OECD, PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices
Recommendation 1d. Use consistency over time as a criterion to consider for question selection and wording.

- NAEP’s inconsistent inclusion of background questions weakens its potential to track trends and improvements within a subject area and topic.
- Recognizing that NAEP needs to periodically refresh its question set, nonetheless NAEP question selection seems haphazard – important questions may not be asked for two or more assessments and then they may reappear with changed wording that disrupt the time series reporting.
- Rather than total eliminating some potentially important survey questions on a topic, NAEP should consider rotating questions so that a question may be asked only once every 4-6 years.
- When rewording is necessary, NAEP should do bridge studies to link the new question responses with prior ones to form an unbroken time series of responses.

Discussion

The opportunity to assess progress on a background indicator over time is lost when NAEP no longer asks a prior question or disrupts the time series by asking essentially the same question in a somewhat different way. Because NAEP is the only major regular state-by-state assessment, question disruption results in a loss of important information to understand changes in a state educational context.

The panel examined the extent to which time series are available on the background question items for a sample of five broad questionnaire categories (Exhibit 4). The examination computed the percentage of questions asked under each category on the 2011 questionnaire for which there was also information for the same question for 2005 or earlier (at least a six-year trend).

- Between 70%-80% of the 2011 items about student characteristics or school demographics could be traced back to 2005 or earlier years.
- The three remaining categories that dealt with more judgmental measurement had much weaker time series availability. Only one-third of the 2011 questions asking about course offerings yielded at least a 6-year trend. No 2011 questions about curriculum or school resources were found on the 2005 or earlier questionnaires.

Some question categories become confusing to the user because of the considerable number of questions no longer asked. A case in point under the group of teacher factor questions is the “Preparation, Credentials and Experiences” category that contains over 400 questions of which more than 300 are no longer used, with many replaced by just slightly different wording. Moreover, what appears to be the exact same question maybe listed a number of times and in different places. Each instance of this all too common occurrence requires the user to search through and find all similar items and try and identify the one, if any, that is available and relevant.
Recognizing that at times changes in question wording may be necessary, the Panel recommends conducting *bridge studies* that would compare responses in the same year for prior and newly revised questions on a topic. NAEP’s 2004 assessments in math and reading conducted a bridge study to compare results from students randomly assigned to the original and revised versions of the assessment (NCES, 2004). Bridge studies were also conducted for the new frameworks in reading and 12th grade math that were introduced in 2009. A similar process could be developed to bridge question changes in important areas of the background questionnaires.

Strategies for holding down the added expense of bridge studies should be carefully explored. Recognize that in conducting a bridge study on background questions, smaller representative samples of the kind used for polling may be adequate and preferable in minimizing error to having no bridge study at all. Also, it may be feasible to add background questions to other bridge studies such as those employed for the assessment.

<table>
<thead>
<tr>
<th>Question Category</th>
<th>Total Questions 2011</th>
<th>Total Number Asked in 2005 or Earlier</th>
<th>% of 2011 questions Asked in 2005 or Earlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Characteristics</td>
<td>10</td>
<td>8</td>
<td>80%</td>
</tr>
<tr>
<td>Curriculum</td>
<td>34</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Course Offerings</td>
<td>78</td>
<td>28</td>
<td>36%</td>
</tr>
<tr>
<td>School Demographics</td>
<td>18</td>
<td>13</td>
<td>72%</td>
</tr>
<tr>
<td>School Resources</td>
<td>43</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

*Source: NAEP Data Explorer*

**Recommendation 1e. Delete duplicative or low-priority questions to make time for the Panel's higher priority items.**

- Several question groups on the student questionnaire are duplicative of information asked on the school or teacher survey. With the 10-minute limited time constraints on the student survey, these duplicative items should be reviewed for elimination and replaced by higher-priority items in the areas recommended by the panel.
- There seem to be an excessive number of background variables collected around a particular topic in some subjects.

**Discussion**

With the student questionnaire currently only 10 minutes long, each question must bring information value or be eliminated and replaced by a high-value item. The Panel has identified two item clusters as duplicative and candidates for elimination.

- Student’s race/ethnicity asked on the student questionnaire is also obtainable from
### Exhibit 5. NAEP’s 2011 Grade 8 Student Questionnaire Asks 8 Questions About Technology Use

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How often do you use these different types of <strong>calculators</strong> in your math class?</td>
<td>a) Basic four-function (addition, subtraction, multiplication, division) b) Scientific (not graphing) c) Graphing</td>
</tr>
<tr>
<td>2. When you take a math test or quiz, how often do you use a calculator?</td>
<td>a) Never b) Sometimes c) Always</td>
</tr>
<tr>
<td>3. For each of the following activities, how often do you use a <strong>calculator</strong>?</td>
<td>a) To check your work on math homework assignments; b) To calculate the answers to math homework problems; and c) To work in class on math lessons led by your teacher.</td>
</tr>
<tr>
<td>4. What kind of <strong>calculator</strong> do you usually use when you are not in math class?</td>
<td>a) None; b) Basic four-function (addition, subtraction, multiplication, division); c) Scientific (not graphing); d) Graphing</td>
</tr>
<tr>
<td>5. How often do you use a <strong>computer</strong> for math at school?</td>
<td></td>
</tr>
<tr>
<td>6. Do you use a <strong>computer</strong> for math homework at home?</td>
<td></td>
</tr>
<tr>
<td>7. On a typical day, how much time do you spend doing work for math class on a <strong>computer</strong>?</td>
<td>Include work you do in class and for homework.</td>
</tr>
<tr>
<td>8. When you are doing math for school or homework, how often do you use these <strong>different types of computer programs</strong>?</td>
<td>a) A spreadsheet program for math class assignments; b) A program to practice or drill on math facts (addition, subtraction, multiplication, division). c) A program that presents new math lessons with problems to solve d) The Internet to learn things for math class e) A calculator program on the computer to solve or check problems for math class f) A graphing program on the computer to make charts or graphs for math class g) A statistical program to calculate patterns such as correlations or cross tabulations h) A word processing program to write papers for math class. i) A program to work with geometric shapes for math class</td>
</tr>
</tbody>
</table>

In addition to direct item duplication, inefficiencies in question selection come about through an imbalance of questions in an area that is disproportionate to its information importance. Exhibit 5 lists the sixteen questions about technology on the 2011 student questionnaire for the eighth grade assessment in mathematics. This is over one-quarter of the items and, while easily measurable, the level of detail may be hard to justify in terms of information value.

### Recommendation Area 2. Strengthen the Validity, Reliability and Coordination of the Measures and Clusters of Measures for Background Questions.

The panel urges attention to strengthening the validity, reliability and coordination of NAEP background questions.

A validity study of the NAEP background questions would assess whether they capture the concept NAEP intends the questions to measure. Concepts such as student socioeconomic status, student expectations, teacher qualifications, instructional content are challenging to define and quantify.
Two common approaches to assessing validity are:

1. Construct validity assesses whether the question or set of questions accurately captures the underlying construct being measured, which is often multidimensional. Socio-economic status is a multidimensional concept about family and community position in society that is incompletely captured by a discrete measure of poverty status—eligibility for a free or reduced-price school lunch.

2. Concurrent and predictive validity assesses whether the questions measuring a concept relate well at the same time or in the future with another established measure of that concept. The different aspects of family involvement that relate to current or future achievement meet the concurrent or predictive validity test.

A reliable measure yields consistent results over repeated measures. Asking teachers a question about frequency of a behavior in terms such as how much emphasis do you place on a subject is imprecise and subject to the subjective opinion and local norms. A more reliable question would ask do you teach this subject once a week, twice a week or very day.

Coordination among a set of questions maximizes information content. A duplicative question yields no added information content. Matching a NAEP set of questions with comparable questions on international assessments is highly efficient as it potentially adds considerable information content at little or no extra respondent burden.

The following recommendations suggest improvements to the validity, reliability and coordination of the NAEP background questions.

**Recommendation 2a. Improve the validity and reliability of the current measures NAEP uses for its mandated student reporting categories.**

- Support the current NAGB and NCES reviews of the best way to measure student socioeconomic status (SES). The known limitations of the current school lunch proxy and the likelihood that even this proxy will no longer be available make this review critically important.
- Assess the implications of changes in multi-racial student populations for the racial/ethnic student classification.
- Examine the accuracy of state-by-state or urban school system performance differences because of variation in the percentages of special education students receiving accommodations.

**Discussion**

_The panel supports the current NAGB and NCES reviews to identify the best way to measure SES variables within the confines of the NAEP questionnaire structure._
This review is critically important given the well-documented limitations of the current school lunch proxy and that the first three State systems are piloting free school lunches for all students in very high-poverty school systems.

Limitations of the current school lunch measure include:

- The current measure divides the population only into two groups of free and reduced price school-lunch eligibles and ineligibles and is therefore insensitive to income differences above and below the income eligibility thresholds. SES is more accurately reflected by continuous measures. For example, this is consistent with studies showing student achievement results are sensitive to income levels over a broad income range.4

- School lunch eligibility is known to be underreported in secondary schools. Secondary students may not want the stigma of making known their families low-income and secondary students may not eat lunch at school. In fact, the grade 12 NAEP did not include school lunch for its 2009 report because of the problems of underreporting.

- The lengthy research literature on measuring SES consistently recommends multidimensional SES indices (Hauser & Warren, 1997) involving family resources, education and occupation. However, NAEP only reports the single student school lunch eligibility measure. NAEP’s SES Project Progress Report (Noel-Miller and Hauser, August 2011) shows that a simple weighted average of indicators of home possessions and parental educational attainment does quite as well as independently estimated regression estimates in predicting math and reading achievement across grade-levels and race-ethnic subgroups.

- The 2010 Healthy, Hunger-Free Kids Act includes a “community eligibility” option, which would permit schools in high-poverty areas to provide free breakfast and lunch to all students without sending home individual paper applications for parents to submit income data. Three states have been selected for 2011-12 pilot eligibility (Illinois, Kentucky and Tennessee) and more states are scheduled to participate in successive years. Moreover, one urban school system Cleveland already counts 100 percent of its students as eligible for school lunch.

Consistent with the research literature, PISA incorporates questions for age 15 respondents to support an international multidimensional, socio-economic index. PISA’s SES index elements consist of: occupational status of the father or mother, whichever is higher; the level of education of the father or mother, whichever is higher, converted into

---

4“In data from the Early Childhood Longitudinal Study (ECLS) measuring kindergarten students achievement on the ECLS reading achievement assessment, low-income students scored at about the 30th percentile, middle-income students scored at about the 45th percentile, and upper-income students scores at about the 70th percentile.” (Lacour & Tissington, 2011)
years of schooling; and the index of home possessions, obtained by asking students whether they had a desk at which they studied at home, a room of their own, a quiet place to study, educational software, a link to the Internet, their own calculator, classic literature, books of poetry, works of art (e.g. paintings), books to help them with their school work, a dictionary, a dishwasher, a DVD player or VCR, three other country-specific items and the number of cellular phones, televisions, computers, cars and books at home.

The panel recommends that NAEP also move toward a multidimensional index for SES using current background questions. The panel further supports a long-run direction along the lines NCES is exploring of a two-pronged approach: (1) Creating an enhanced student background questionnaire with items that probe resources in the home, parents’ education level, and parents’ employment status; and (2) Using geocoding software to link students’ home addresses to aggregate SES data available from the United States Bureau of the Census. The geocoding would reflect neighborhood and community factors that influence student performance.

In this context, the panel strongly supports the current NCES pilot to “generate SES information from the Census American Community Survey (ACS) data using school catchment zones, and which would make the collection of students’ home address unnecessary for any assigned (non-choice) school.”

The Panel recommends assessing the potential implications of changes in multi-racial student populations for the valid measurement of the racial/ethnic student classification.

Starting in 2011 NAEP collected multi-racial data from school records and included it in the main subject-matter reports. In 2008, the U.S. Census (2011) reported the multiracial population at 7.0 million or 2.3% of the population. This number is for the full U.S. population and the percentage for the school age children would be expected to be higher to reflect the growing number of inter-racial families in the U.S. NAEP now collects these race/ethnicity data two ways – from school records and student reports. The student reports allow students to check more than one box within racial and ethnic categories. NAEP should compare the self-identified reports with the official school records.

Recommendation 2b. Enhance the validity of student responses at different grade levels.

- Assess whether the same construct (e.g., SES) is best measured by different and increasingly more valid items across grades 4, 8 and 12.

Discussion

A younger (grade 4) NAEP respondent is likely to have more difficulty accurately going

---

5 Quote from NCES Jan. 26, 2012 memo from Peggy Carr to Larry Feinberg.
through a typical question-answer process, which involves 4 steps: (1) understanding and interpreting the question being asked; (2) retrieving the relevant information from memory; (3) integrating this information into a summarized judgment; and (4) reporting this judgment by translating it to the format of the presented response scale (Borgers & Hox, 2000).

The Panel recognizes that NAEP questionnaire design already gives considerable attention to differences in the ability of students at different age groups to go through these four steps to respond accurately to background questions. Thus, NAEP dropped a question about parent’s education for grade 4 students because of research suggesting that responses from grade 4 students were less reliable than from older students. However, balanced against possible student response error is the loss of potentially useful information from eliminating questions. The Panel recommends NAEP explore the inclusion in the grade 4 questionnaires of questions that ask about mother’s and father’s highest education. The exploration should compare the error rates in estimating SES with and without the grade 4 parent education item.

The Panel also recommends that NAEP consider how the same construct (e.g., SES) can be measured by increasingly more valid and multi-dimensional clusters of items for students in upper grades.

**Recommendation 2c. Accurately measure the multi-dimensional nature of learning-to-learn skills including student learning behaviors, motivation and expectations.**

- Learning-to-learn skills refer to a cluster of personal qualities, habits and attitudes and include learning strategies, motivations and expectations. These soft-skills have shown a strong predictive relationship with math and reading achievement and workforce performance over decades (Coleman report, ETS paper on ECLS, NAEP, TIMSS and PISA). The Panel also notes that motivation and expectation questions are a regular component in major NCES national longitudinal surveys and international surveys at the primary and secondary level. However, developing questions that accurately measure non-cognitive skills through subjective responses to survey questions is challenging and should build on the considerable existing body of measurement in this area.

**Discussion**

To accurately measure some of the hard-to-measure concepts the Panel has recommended (1c above) that NAEP develop clusters of questions that collectively provide a good measure of different aspects of theoretically-based frameworks. Currently, the NAEP background questionnaire, especially the student questionnaire, is highly restricted by time constraints and does not contain the rich set of items needed to validly measure many learning attributes associated with student achievement.
Exhibit 6 provides an example of how PISA’s in-depth questioning draws out students’ approaches to understanding a particular type of text. In essence, the questionnaire creates more authentic learning situations from which to document students’ behaviors.

**Exhibit 6. PISA’s In-Depth Student Questions Of How They Would Approach Remembering Information in a Text Approximates An Authentic Assessment Item**

Reading task: You have to understand and remember the information in a text.

How do you rate the usefulness of the following strategies for understanding and memorizing the text?

<table>
<thead>
<tr>
<th>Possible strategy</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not useful at all</td>
</tr>
<tr>
<td>a) I concentrate on the parts of the text that are easy to understand.</td>
<td>○</td>
</tr>
<tr>
<td>b) I quickly read through the text twice.</td>
<td>○</td>
</tr>
<tr>
<td>c) After reading the text, I discuss its content with other people.</td>
<td>○</td>
</tr>
<tr>
<td>d) I underline important parts of the text.</td>
<td>○</td>
</tr>
<tr>
<td>e) I summarize the text in my own words.</td>
<td>○</td>
</tr>
<tr>
<td>f) I read the text aloud to another person.</td>
<td>○</td>
</tr>
</tbody>
</table>

Source: OECD PISA 2009 Student Questionnaire

The Panel recommends that NAEP explore including these rich behavior questions for grades 8 and 12 even if it would require expanding the student questionnaire time for completion.

**Recommendation 2d. Improve question reliability by replacing imprecise phrases such as “infrequent” or “a lot” with more precise terms such as “once a month” or “twice or more a week”**.

**Discussion**

NAEP should ask questions involving frequency of behaviors or intensity of services in a form that elicits the most precise meaning to these terms. In this regard, some NAEP questions are not specific and the reliability of responses to these questions may be low.

The following illustrates two questions on the NAEP 2009 teacher questionnaire asking teachers about frequencies of time spent on science. Question a) asks about time spent on physical science in terms using categories such as “Little”, “Some” or “A lot” that could
mean quite different amounts of time depending on teacher norms. By contrast, question b) uses the preferred wording in which response times are expressed in clear distinct time intervals.

Question a): In this class, about how much time do you spend on physical science?
   Answers: None = 4%, Little = 9%, Some = 27%, A lot = 60%

Question b): About how much time in total do you spend with this class on science instruction in a typical week
   Answers: Less than 1 hour = 1%, 1-2.9 hours = 4%, 3-4.9 hours = 60%, 5-6.9 hours = 25% , 7 hours or more = 9%

NAEP should specify responses to questions about frequency and intensity in a specific quantifiable format wherever feasible.

**Recommendation 2e. Coordinate NAEP background questions with those asked on international or domestic surveys.**

- NAEP should explore framing its questions with as identical wording as feasible to similar questions found on international assessments.
- NAEP should examine the feasibility of NAEP coordinating with the NCES household survey to administer the household survey to families of students who participate in the NAEP subject assessments. This coordination between the two surveys would link the results of adults in the household survey with students’ NAEP assessment scores.

**Discussion**

In recent years NAEP cognitive assessment results have been linked internationally to place NAEP national and state disaggregated performance on an international TIMSS or PISA scale. NCES now is linking the 2011 grade 8 mathematics and science assessments of NAEP and TIMSS so international benchmarks can be reported on NAEP. Potentially, many of the responses to the background questions can also be compared with similar questions asked on international assessments. Examples include time spent on homework, after-school learning, taking algebra in the eighth grade, or teacher preparation to teach math or science.

To make valid international comparisons, NAEP needs to word its questions so that they are very similar or identical to the wording of the comparable questions on international surveys. Comparability of wording will only be achieved through careful question linking.

Exhibit 7 illustrates the potential payoffs that could occur from linking NAEP responses to those on an international assessment measuring with respect student time learning in regular school lessons and out-of-school lessons compared with high-scoring Japan and Korea.
Almost 30% of U.S. age-15 students spend less than 2 hr. in a math class per week compared with less than 10% of Japanese students and 5 percent of Korean students. Moreover, those students with the lowest scores receive the least math instructional help in-school.

Eighty percent of U.S. age-15 students spend no time learning math in formal afterschool instruction compared with only a quarter of Japanese or Korean students.

It would be valuable for individual states to be able to compare their students’ math instructional time in-school and out-of-school with those of the Asian performers, but NAEP collects very little information about learning time. For example, it asks only about number of days a week in math instruction and not about number of hours and there is no information about time spent in math or other subjects after school. Had NAEP spelled out a basic theoretical framework identifying clusters of questions about time measurement (recommendation 1c) NAEP might have been more likely to align its questions to compare states with the interesting PISA national results.

Recommendation 2f. Build on current NCES cognitive interview techniques by using cognitive laboratories, such as small standing panels, to field test questions to establish their validity and reliability.

Discussion

NCES conducted cognitive laboratory investigations of the responses of students and teachers to questions from the 1996 and 1998 background questionnaires (Levine, Huberman, and Buckner, 2002). Cognitive interviews are an approach “to assess how respondents comprehend survey items and what strategies they use to devise answers.”

The 1990’s studies identified a number of general types of item problems:
• Behavioral frequency discrepancies. These items ask about how frequently a student or teacher engages in specific activities or practices. The average level of agreement between fourth grade students and their teachers on items that used a four-point rating scale was only 38 percent; for eighth grade students and their teachers, the level of agreement was still only 51 percent. Guessing would yield agreement of 25 percent.

• Time frame discrepancies. Differences between teachers and students in the period over which behavior is estimated were common. Teachers would generally think about the current year and students about a very immediate near-term period. Also, when teachers were asked about the frequency of a behavior such as how often a particular science topic was taught, teacher’s responses applied to only when science is taught. Thus the response option, “Almost every day,” was explicitly interpreted as “Almost every day that science is taught.”

• Comprehension discrepancies. Different respondents may interpret items differently. When teachers responded to a question about frequency of a behavior with “students in your class,” some teachers would answer about the typical student and others would respond if any one student exhibited that behavior.

• List format discrepancies: Loss of context. On a long list of items, students or teachers might forget the context in which the question was asked. A student might interpret a question about school behavior such as reading and respond with their general reading behavior in or out of school.

NAEP also conducted a cognitive laboratory analysis of the Responses of fourth and eighth graders to questionnaire items and parental assessment (Levine, et.al. 2001).

The Panel believes that cognitive lab interviews are able to detect and prevent many survey design problems. Hence, it recommends that NAEP use cognitive labs more extensively with an accompanying small panel of adult (teacher/principal) and child respondents to validate and improve background questions. In addition, small-scale pilot studies should be used to assess the feasibility, reliability, and external validity of survey items. We recognize that this may increase costs but it would help make the overall NAEP a better source of information.

**Recommendation Area 3. Reform NAEP Sampling to Enhance the Scope of the Background Questions While Maintaining Sampling Accuracy.**

Limitations of time and concerns over data burden severely constrain the depth of the student background questions. As a result, NAEP often lacks the richness in its background questions that would enable it to replicate the constructs such as those PISA creates from lengthy multiple items around different aspects of research-based
frameworks. To further extend the richness of its data sets, PISA also enhances its basic student and principal questionnaires with optional supplemental questionnaires. NAEP should consider expanding the depth of its questions through a variety of strategies including spiral sampling (currently already under consideration by NAEP), expanded questionnaire time and rotating background questions across samples.

Recommendation 3a. Support NCES’s exploration of a spiral sample methodology to expand the scope of background questions, while assessing the possible loss in the representativeness of disaggregated data.

- Spiraling questions so that no student takes the full set of background questions would allow NAEP to expand the scope of its background items. The current 10-minute limit for the student questionnaire severely constrains the current scope and depth of the student questionnaires. By contrast PISA is able to support richer construct development with its 30-minute student questionnaire.
- In assessing questionnaire spiraling, it is important to consider how it would reduce NAEP’s ability to provide statistically-accurate state-by-state or urban district information, especially if broken out for different student sub-groups.

Discussion

The Panel supports exploring the proposed spiral sampling of questionnaire items in order to implement improvements in student questionnaire scope and depth. As noted, one such improvement would be to enable greater in-depth questioning through clusters of items that measure different aspects of research-based topic frameworks.

However, the Panel urges NCES to quantify how item spiraling will reduce NAEP’s ability to disaggregate state or urban district responses for specific population groups. For example, will background questions be available in sufficient sample size for all population groups for which cognitive student achievement data are reported?

Illustrating this point is an analysis of whether a state has changed its grade-8 access of students to a course in algebra during the two-year interval between successive NAEP assessments. It turns out that Alabama raised the percentage of its students in schools offering grade-8 algebra by 6 percentage points during the two years and Arizona decreased it by 5-percentage points. These changes are sizeable for two years, yet neither change was statistically significant. A spiral sampling approach would further reduce the odds of obtaining statistical significance.

Recommendation 3b. Consider other item-sampling reforms to obtain the needed questionnaire time including lengthening the student survey; establishing a 4-year interval between administration of some background questions; and pooling item responses across survey administrations.

- The ten-minute target length for responses to the student questionnaire does not seem grounded in empirical experience and NAEP would do well to consider the
merits and feasibility of a lengthier questionnaire. TIMSS grade 4 and 8 student questionnaires are targeted for 30 minutes at each grade and do not appear to suffer from high non-response rates.\(^6\)

- Some background questions with slow-moving trends may be adequately monitored through repeating survey questions at four-year intervals.
- Pooling item responses across successive surveys may also be a permissible strategy to expand the sample provided that response changes are sufficiently slow moving.

**Discussion**

These sample reforms could expand the number of background items surveyed over a multi-year period, while maintaining accurate State-by-state reporting of background questions. However, each involves its own tradeoffs in terms of questionnaire time and the availability of items on any one survey. The panel requests that NCES examine and report to NAGB the comparative strengths and weaknesses of different approaches to expanding questionnaire items.

**Recommendation Area 4. Reinstitute the Analysis and Regular Reporting of NAEP Background Questions.**

Rich responses to relevant background questions would mean little if NAEP continues its present practice of including very few findings from the background questionnaires in its reports. The main exception is the reporting of achievement by the congressionally required student subgroups. For other background information, the only recourse for a potential user to these data is to conduct one’s own analyses using the NAEP Data Explorer. As a practical matter, this is an option that only professional researchers (and few others) will have the time and skills to undertake.

This set of recommendations would bolster the analysis and reporting of the background questions by means of separate publications, online tables, and improvements to the Data Explorer. The recommendations also include a caution to not repeat the mistakes of the past by excessive reporting of causal interpretations of point in-time data.

---

\(^6\) *TIMSS 2011 Assessment Design* (p126) describes expected student testing time at grade 4 of 72 minutes for the student achievement booklet and 30 minutes for the student questionnaire. The grade-8 times are 90 minutes for the student achievement booklet and 30 minutes for the student questionnaire.
Recommendation 4a. Prepare special reports highlighting the background question findings.

- The special reports would provide interested readers with key findings derived from the background questions. These special reports could be prepared and released either with the achievement report or during the interval between assessment administrations. The Panel recommends NAEP consider two initial special reports, one organized around learning opportunities in school and a second around learning opportunities and conditions out of school. A third report that explores benchmarking to find correlates of high-performing states and districts should also be considered.
- These synthesis reports would also provide a way to assess the information value of current and past questionnaire items.

Discussion

Special reports would provide access to the background questions in manageable-size documents that don’t overwhelm the reader. An example of a NAEP special report is *The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8*, which is Part II of the National Indian Education Study of 2009. Part II complements the Part I report on NAEP assessment results for American Indian students by providing information about students, their families and communities, and their school experiences.

More generally TIMSS and PISA illustrate two approaches to developing topics for the special reports. TIMSS includes individual chapters organized around different questionnaire topics:

- Students’ Backgrounds and Attitudes Towards Science
- The Science Curriculum
- Teachers of Science
- Classroom Characteristics and Instruction
- School Contexts for Science Learning and Instruction

The 2009 PISA has published a series of special reports, synthesizing lessons learned to improve academic achievement:

- *Overcoming Social Background: Equity in Learning Opportunities and Outcomes* looks at how successful education systems moderate the impact of social background and immigrant status on student and school performance.

The Panel recommends that NAEP give priority to preparing two initial special reports using current data.

- The first report would focus on learning opportunities and conditions in school including examining characteristics of teachers, curriculum and instruction and the distribution of these characteristics among schools with students of various racial and socioeconomic concentrations.
- The second report would explore the characteristics of learning opportunities after-school and in the home, again comparing students from different economic and social backgrounds.

These reports would help inform future background variable data collections by identifying data of the greatest value in what currently is collected.

Other future NAEP reports could take advantage of NAEP’s special data collections. One might examine the characteristics of high-performing states or jurisdictions. Another would explore the extensive NAEP question sets on technology use in instruction.

**Recommendation 4b. Prepare an online compendium of key background indicators for States and participating urban districts.**

**Discussion**

The state-by-state or urban district compendium would take advantage of NAEP’s unique capacity to report a consistent series of state and urban district background data over time. The Panel heard an example of such a report incorporating NAEP data in the STEM area that is being prepared by the nonprofit organization Change the Equation.

Exhibit 8 illustrates for the 22 districts participating in the 2011 Trial Urban Assessments a hypothetical mock-up of background question responses focused around grade 8 and mathematics. A few findings from the urban district data in Exhibit 8 illustrate the potential value of indicator comparisons:

- The systems with the highest percentage of students absent 5 or more days were Detroit, Milwaukee, DC and Cleveland, which were also places with lower student scores.
- For grade 8 students taking algebra, the highest scoring districts of Austin and Charlotte had relatively low rates of absenteeism.

---

7 From *Change the Equation,* a non-profit, non-partisan coalition of more than 100 CEOs who are committed to bringing high-quality Science, Technology, Engineering, and Mathematics (STEM) learning to every U.S. child.
• Although urban school systems have somewhat higher rates of students participating in math at an afterschool tutoring or school program, only Atlanta had at least half the students avail themselves of afterschool assistance.

• Urban districts for the most part have above national-average percentages of staff teaching math with a major, minor or special emphasis in mathematics.

• Access to the Internet at home is widespread among urban areas making school support for learning at home more feasible than might be generally believed.

### Exhibit 8. Illustrative Table of Background Question Indicators With a Grade 8 Math Focus:
School Districts Participating in the 2011 Trial Urban Development Assessment

| Jurisdiction          | Grade 8 All Students | Eligible for National School Lunch | Grade 6 Students Absent 6 or more days last month | Grade 6 Students in Algebra | Grade 6 Students 6 or more Hours of Math Per Week | Grade 6 Students 1 Hour or More Math Homework | Grade 6 Does Math At An Afterschool Program | Grade 6 Enrolled In Math Through Alternative Certification | Grade 8 Teacher Has Math Major/Minor/Special Emphasis | Grade 8 Participates In Math At School | Grade 8 Students At Math Ability | Grade 8 All Students In Math Class | Grade 8 Computer Accessible to Teachers and Students |
|-----------------------|----------------------|-------------------------------------|-----------------------------------------------|----------------------------|--------------------------------|--------------------------------|-----------------------------------------------|-------------------------------------------------|---------------------------------------------------|--------------------------------|---------------------------|--------------------------------|-----------------------------|--------------------------------------------------|
| National              | 284                  | 44                                  | 7                                            | 42                          | 37                            | 17                             | 21                             | 17                             | 38                             | 17                         | 75                        | 45                         | 84                          | 284                                |
| Albuquerque           | 275                  | 60                                  | 8                                            | 57                          | 55                            | 20                             | 27                             | 33                             | 52                             | 32                         | 60                        | 19                         | 77                          | 275                                |
| Atlanta               | 268                  | 62                                  | 6                                            | 75                          | 50                            | 38                             | 57                             | 57                             | 95                             | 84                         | 61                        | 24                         | 96                          | 268                                |
| Austin                | 267                  | 69                                  | 8                                            | 61                          | 57                            | 30                             | 42                             | 37                             | 57                             | 58                         | 53                        | 52                         | 89                          | 267                                |
| Baltimore City        | 261                  | 85                                  | 0                                            | 46                          | 95                            | 41                             | 28                             | 24                             | 76                             | 53                         | 85                        | 71                         | 89                          | 261                                |
| Boston                | 282                  | 76                                  | 6                                            | 78                          | 59                            | 30                             | 13                             | 12                             | 61                             | 47                         | 56                         |
| Charlotte             | 265                  | 52                                  | 6                                            | 87                          | 18                            | 29                             | 44                             | 47                             | 33                             | 85                         | 76                         | 70                         | 76                          | 265                                |
| Chicago               | 270                  | 84                                  | 4                                            | 67                          | 67                            | 37                             | 23                             | 29                             | 84                             | 44                         | 56                        | 30                         | 89                          | 270                                |
| Cleveland            | 256                  | 100                                 | 11                                           | 89                          | 33                            | 25                             | 6                              | 58                             | 14                             | 51                         | 44                        | 16                         | 96                          | 256                                |
| Dallas                | 274                  | 60                                  | 7                                            | 48                          | 57                            | 37                             | 30                             | 61                             | 56                             | 16                         | 53                        | 57                         | 57                          | 274                                |
| Detroit               | 248                  | 73                                  | 17                                           | 81                          | 46                            | 37                             | 11                             | 83                             | 16                             | 53                         | 65                        | 51                         | 51                          | 248                                |
| District of Columbus  | 255                  | 70                                  | 12                                           | 65                          | 29                            | 30                             | 57                             | 46                             | 23                             | 26                        | 86                         |
| Fresno               | 256                  | 68                                  | 10                                           | 51                          | 32                            | 11                             | 28                             | 6                              | 57                             | 23                         | 91                        | 75                        | 59                          | 256                                |
| Hillsborough County  | 282                  | 64                                  | 8                                            | 67                          | 59                            | 30                             | 13                             | 22                             | 40                             | 35                         | 95                        | 86                         |
| Houston               | 279                  | 75                                  | 6                                            | 83                          | 20                            | 37                             | 56                             | 53                             | 25                             | 84                         | 58                        | 68                         | 68                          | 279                                |
| Jefferson County (KY)| 274                  | 80                                  | 7                                            | 68                          | 46                            | 14                             | 20                             | 21                             | 34                             | 36                         | 77                        | 80                         | 80                          | 274                                |
| Los Angeles           | 261                  | 82                                  | 6                                            | 67                          | 44                            | 40                             | 27                             | 59                             | 67                             | 37                         | 75                        | 52                         | 74                          | 261                                |
| Miami-Dade           | 272                  | 72                                  | 6                                            | 67                          | 43                            | 47                             | 26                             | 38                             | 72                             | 25                         | 90                        | 13                         | 88                          | 272                                |
| Milwaukee            | 254                  | 81                                  | 13                                           | 76                          | 43                            | 31                             | 37                             | 74                             | 82                             | 29                         | 66                        | 16                         | 86                          | 254                                |
| New York City        | 272                  | 87                                  | 10                                           | 83                          | 35                            | 35                             | 65                             | 36                             | 60                             | 63                         | 79                         |
| Philadelphia         | 285                  | 85                                  | 10                                           | 90                          | 34                            | 27                             | 24                             | 54                             | 37                             | 50                         | 76                        | 86                         |
| San Diego            | 276                  | 60                                  | 8                                            | 48                          | 13                            | 27                             | 11                             | 40                             | 17                             | 78                        | 72                       | 80                         | 80                          | 276                                |

Source: NAEP Data Explorer

An actual set of NAEP urban or state indicators should be carefully developed to include the most informative research-based responses and would summarize other subjects and grades.

The Panel also recommends considering a larger online compendium of national, state or urban background question results be prepared and structured to easily find questions of interest around a topic. The typical educator or policymaker, who would benefit from the findings contained in the background questions, lacks the time to understand and delve into the questionnaires through the NAEP Data Explorer.

To facilitate online access to prepared tables of questions, the user might be given options to select: (a) questions based on a Google-type question search (b) questions as they appear on the student, teacher or school questionnaires; or (c) questions grouped by topic.
and grade. Once the questions are selected, tables at the different system levels would be automatically generated and viewed.

**Recommendation 4c. NAEP’s reports should not indicate causal interpretations using the background questions. However, the NAEP data offer some unique advantages for generating relationships and hypotheses about factors that may be associated with performance and these findings should guide more rigorous in-depth follow-on analyses.**

First, NAEP’s performance reporting by subject, population group or jurisdiction is often the primary source of objective national performance data overtime. These data naturally raise questions about the underlying factors that produce the high and low performance. However, the Panel concludes, as have other NAGB panels before it, that NAEP should not publish causal interpretations of the factors determining performance differences based on the NAEP data.

Second, it is important to differentiate NAEP’s use of rigorous external research to identify, measure and report on background variables that support or work against achievement (Barton, 2002). In such instances, NAEP is not generating the findings from its cross-sectional data, but instead drawing upon an external evidentiary research base for the questions selected. Examples would be the degree to which lower income or lower performing students have access to at least equal levels among opportunity-to-learn variables such as certified teachers or instructional time. Another example would be to compare high and low performers on such factors as alignment of instruction with standards that are systemically related to achievement.

**Recommendation 4d: NAEP should encourage others to conduct exploratory studies of the NAEP background variables.**

- This may be through initiating small-grant competitions for researchers to analyze NAEP background-question data or by partnering or supporting others to conduct their own analyses of the background variables.
- These grants would provide funds for researchers to explore interesting and potentially policy-relevant topics and methodologies.
- The independent reports supported through the external grants could use the background question data to inform national education policy debates without any direct NAEP organizational involvement and oversight over the findings. The external grantees might also explore issues and topics where analysts might employ NAEP data to explore correlations or associations.
- There is precedent for NAEP to support mini-grant competitions of this kind.

**Discussion**

Other statistical agencies routinely support in-depth analyses of their statistical data. For example, the Bureau of Labor Statistics (BLS) has its own employment research and
program development staff to conduct original research using BLS data. The ASA/NSF/Research Fellow program is jointly supported by American Statistical Association and The National Science Foundation with participation of the U.S. Census Bureau, and the Bureau of Economic Analysis. This program jointly supports a Federal Statistics Fellowship program bringing academic researchers to work with statisticians and social scientists in the three federal agencies for up to one year.

NAEP should consider launching a similar program through small grants ($10,000-$50,000) competitively given to independently conduct research using NAEP data including the background questions. The focus of this research would be primarily on measurement and other statistical issues to improve the election and quality of the background variables.

The Panel also suggests that NAEP consider various strategies for encouraging and supporting outside researchers to conduct analyses of the NAEP data. NCES may want to work cooperatively with other organizations and foundations in these efforts. For example, NCES partially supported with foundations the widely cited research by Grissmer (2000) to analyze the state-level NAEP repeated time series achievement and background questions to examine the impact of systemic reform on improved achievement.

**Recommendation 4e. Further improve the powerful online NAEP tools for data analysis.**

- NAEP should follow the PISA model of including with each published table a link to its online downloadable spreadsheet that may be analyzed though software such as Excel.
- Extend the Data Explorer to facilitate the manipulation and analyses of the background questions by themselves without the achievement results. Extending software to build-in multivariate analyses should be considered.

**Discussion**

NAEP should follow the PISA model of including with each published table a link to its online downloadable spreadsheet that is analyzable though software such as Excel. Each NAEP table and chart contains useful breakouts of the overall assessment and background data, which have been extracted and organized to focus on particular topics. Analysts and researchers may want to build off these tables to add more data series, conduct descriptive statistical analyses or pull apart and regroup the data to emphasize different points. Currently, NAEP offers no direct means to work off of the tables and charts in the reports other than to reenter the data by hand or to try and recreate them using the NAEP Data Explorer.

The Panel urges NAEP reporting to follow the lead of PISA by attaching a “statlink” to a downloadable excel file of the data in the table so that the user is able to access directly the data content without burdensome data reentry. Exhibit 9 shows how statlink was used to highlight the U.S. score compared with Singapore. The published PISA chart was
The Panel further recommends that NAEP strengthen the Data Explorer to facilitate the manipulation and analyses of the background questions by themselves without the achievement results. Extending software to build-in multivariate analyses should be considered.

While the NAEP data explorer is a typically excellent and easy to use tool when analyzing achievement results, analysis of the non-cognitive background variables can be quite challenging even for data experts. Several problems occur:

- Finding the question of interest in the Data Explorer is made more difficult by not having an alphabetic listing of question topics. A direct link from a question in the published student, school or teacher questionnaire to that question in the Data Explorer would also be helpful.
- The Data Explorer is designed to use the background questions as categories by which to classify student achievement scores (e.g., by whether a student participates in school-lunch) and not to independently analyze the background question responses themselves.

The following is a real-world example of the challenges that arose in using the Data Explorer to compare how much time teachers in each state spend on math instruction at the fourth grade.
• Step 1. Find whether this question is available on the NAEP Data Explorer.
  – Unfortunately, the Data Explorer does not contain a question search tool to determine if this question is available.
  – Look for “time spent on math instruction” under the curriculum section and find an item for class time spent on different science categories (e.g., earth science), but not for mathematics.
  – Look for “time spent on math instruction” under the “course offerings” section of the Data Explorer and find a question about “4th grade instruction in math” that covers time spent in class, but the latest data are for 1996.
  – Don’t give up, and go to the “classroom management” section of the Data Explorer and find “the 2011 question of interest: Amount of time required for math instruction.” This works but why is the question under classroom management and why is time spent in instruction listed in three different places?

• Step 2. Go to the Data Explorer to print a table displaying the distribution of time each state spends on math instruction at different grades. Instead obtain a table (Exhibit 10) that distributes State assessment scores by time intervals, but does not display the frequencies of the time intervals themselves.

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Average Scale Score</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Alabama</td>
<td>212</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Alaska</td>
<td>232</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>Arizona</td>
<td>226</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Exhibit 10. Normal Data Explorer Display That Uses Background Variables (Time Spent Per Week on Math) As Classifiers To Distribute Achievement

The problem is that Data Explorer has a default that assumes interest in the distribution of assessment findings and not in the distribution of the background variables. The override selection to obtain a straightforward table of the time distribution of math scores is through a little known and not easily found path under the statistics option under edit reports. This permits the user to deselect assessment as the dependent variable and replace with the percentages distribution of the background question (Exhibit 11). This option should be highlighted in the NAEP general instructions and in the edit reports screen that everyone sees.

Finally the Panel understands that that the Data Explorer once had a capability to conduct multivariate analyses, but that is was removed by the NCES Chief Statistician because of concern about potentially disclosing personally identifiable information about sampled students. The Panel understands this concern, but
requests NCES to review the decision to determine whether disclosure safeguards can be built into an online multivariate capability.

**Exhibit 11. Desired NAEP Data Explorer Display That Presents The Distribution of Time Spent On Math Per Week By State**

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Less than 3 hours</th>
<th>3-4.9 hours</th>
<th>5-6.9 hours</th>
<th>7 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard Error</td>
<td>Percentage</td>
<td>Standard Error</td>
</tr>
<tr>
<td>2011</td>
<td>Alabama</td>
<td>4</td>
<td>(1.1)</td>
<td>3</td>
<td>(1.2)</td>
</tr>
<tr>
<td></td>
<td>Alaska</td>
<td>3</td>
<td>(0.5)</td>
<td>2</td>
<td>(0.9)</td>
</tr>
<tr>
<td></td>
<td>Arizona</td>
<td>3</td>
<td>(1.1)</td>
<td>3</td>
<td>(1.1)</td>
</tr>
</tbody>
</table>

NOTE: Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.


5. Implementing the Panel Recommendations

The panel report identifies four areas for improving the usefulness and use of the NAEP Background Questionnaires with respect to question selection, measurement, sampling, and analyses and reporting.

The panel recognizes that the benefits of the recommendations in each area should be balanced against their cost in relation to other expenditures in NAEP’s annual budget of over $130 million. A decision on the merits of each item involves potential tradeoffs that are outside the panel’s mandate and expertise. In considering resource priorities, however, the panel concludes that even though the background variables have been underused in recent years, they could, for a relatively modest expenditure, become the means for greatly increasing the usefulness and impact of NAEP. The panel therefore urges that its recommendations be implemented through:

- Producing *special reports* on the background data that analyze the considerable quantity of data already collected but largely unreported and unanalyzed.
- Moving quickly to initiate a long-term effort to improve the relevance, quality, coherence and usefulness of a *core and rotated set of background variables* while implementing recommended improvements for measurement accuracy and sampling efficiency.
- Further improving the usability of the Data Explorer and other NCES online tools, which are already of high quality.
Recommendation 5a. Exploit existing background data through special reports focused on issues and topics informed by background questions.

Discussion

The proposed special reports in 5a are designed to mine the unexploited investment in the largely unanalyzed background questions. These reports might be modeled on the special publication of background data from the National Indian Education Study of 2009, *Part II: The Educational Experiences of American Indian and Alaska Native Students in Grades 4 and 8*, cited in Recommendation 4a.

The special publications would describe:

- In-school learning opportunities and other educational experiences focusing on data already collected on curriculum, instruction, teachers and other school resources including technology.
- Out-of-school learning opportunities and other educational experiences including after-school and at home.
- The background characteristics of high performing states and school systems contrasted with low-performers. This benchmarking study would be purely descriptive, serving to guide follow-on research to improve understanding of the factors differentiating high and low performing states and districts.

These would be three synthesis reports, drawing on data from NAEP assessments across the curriculum and, where possible, trends over time.

Recommendation 5b. Initiate a set of activities to build clusters of core and second-tier questions around high-priority topics for the 2015 NAEP administration.

Discussion

Given the long lead times for questionnaire development, this effort needs to begin immediately in order to affect the 2015 NAEP reading and mathematics administration. The revised questionnaires would refocus the background questions to identify an expanded first-tier core and second-tier set of rotated question clusters, including a rotated set of policy issues (Strategies 1 and 2, Exhibit 12). As NAEP redefines its question sets, NAEP would improve measures through published evaluations of their validity, reliability and consistency with each major assessment (Strategy 3, Exhibit 12). To find the questionnaire time to develop in-depth question sets, Strategy 4 prepares a NAEP analysis and report on a combination of sampling reforms addressing spiraling questions and extra question time.
## Exhibit 12. Longer-term Background Question Activities / Products

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Recommendation</th>
<th>Activities/Products</th>
</tr>
</thead>
</table>
| 1. Select core and rotated clusters of questions around research-based theoretical frameworks | 1a, 1c         | • Identify 1<sup>st</sup> tier core clusters (student sub-groups, student learning opportunities, student motivation)  
• Identify 2<sup>nd</sup> tier rotated questions  
• Publish background questions with research-based justifications for question clusters |
| 2. Extend NAEP Background Questionnaires to monitor topics of current policy interest | 1b             | • Identify current and future policy issues that are suited for NAEP Background Question (Common Core, Teacher evaluation, online instruction)  
• Propose rotating cycle of 3 major policy areas beginning with 2013 assessment. |
| 3. Launch a process for the continual examination of the validity, reliability, efficiency, and consistency of measures | 2a, 2b, 2c, 1d, 2f | • Report on validity & reliability of SES & responses at different age levels  
• Implement quality review procedures for reliability and consistency of questions.  
• Launch a cognitive laboratory capability with possibly an available small standing supplementary panel. |
| 4. Report on item sampling reforms to incorporate extended question sets and topics including eliminating duplicative and low-priority items | 3a, 3b         | • Report on a strategy to add questions for cluster analyses and policy issues through questionnaire spiraling, alternating questions across assessment administrations, adding extra questionnaire time and eliminating low-priority items, |

### Recommendation 5c. Further improve the usability of the Data Explorer and other NAEP online tools, which are already of high quality.

### Discussion

While the Data Explorer is an excellent tool for online access of NAEP achievement data, addressing weaknesses in the analyses and display of the background data in the Data Explorer and publications would extend the usefulness of NAEP’s current online tools.

- Simplify and clarify how to use the Data Explorer to analyze the distribution of responses on background questions.
- Explore the potential for conducting multivariate analyses through the Data Explorer.
- Build links that allow the data in tables and charts in NAEP publications to transfer to excel spreadsheets for further analyses.
Recommendation 5d. Promote implementation by creating a single Governing Board committee responsible for all background questions; provide adequate resource support, while ensuring efficient resource use; and publicize background question products and findings.

Discussion

To promote implementation of the background question recommendations and make sure change occurs, the panel suggests that NAGB establish a separate standing committee to review all background questions and oversee a multi-year development plan to improve the questions and their use. Currently, the Board’s responsibilities for the background questions are divided between the Assessment Development and the Reporting and Dissemination Committees. A unified standing committee should regularly monitor and report on implementation of the panel’s recommendations by NCES and Governing Board staff.

The panel further recommends that a review be conducted of the resources needed in terms of time, money and personnel to implement the recommendations in this report. One approach to the problem may be to reduce costs in certain areas. For example, efforts should be made to eliminate lower-priority activities, such as the duplicative collection of racial data and the disproportionate number of questions asked in areas such as technology. Another approach should be to make a clear and powerful case for the usefulness of having a coherent set of relevant and valid background variables to help explain NAEP results and to take this case to the Department of Education, the Office of Management and Budget (OMB), and Congress.

In conclusion, the NAEP background questions are a unique national information resource. The Governing Board and NCES have a responsibility to develop this resource to better understand academic achievement and the contexts in which it occurs and, hopefully, to help spur educational improvement.
REFERENCES


Noel-Miller and Hauser -- NAEP-SES Project Progress Report (August 2011)


MEMBERSHIP AND AFFILIATIONS

- Marshall S. Smith (Chair)
  Visiting Scholar, Carnegie Foundation for the Advancement of Teaching
  Former U.S. Under Secretary of Education
  Palo Alto, CA

- Naomi Chudowsky
  Co-Director
  Caldera Research, LLC
  Bend, OR

- Alan Ginsburg (Secretary)
  Education Consultant
  Former Director of Policy and Program Evaluation
  U.S. Department of Education
  Washington, DC

- Robert Hauser
  Executive Director
  Division of Behavioral and Social Sciences and Education
  National Research Council
  Washington, DC

- Jennifer Jennings
  Assistant Professor of Sociology
  New York University
  New York, NY

- Sharon Lewis
  Research Director
  Council of the Great City Schools
  Washington, DC
Policymakers Weigh Gathering More Data for NAEP

Goal is to improve understanding of performance

By Erik W. Robelen

As many experts raise questions about the future of "the nation's report card," the governing board for the assessment program is exploring changes aimed at leveraging the achievement data to better inform education policy and practice.

The core idea, outlined in a report to the board, is to expand and make far greater use of the background information collected when the National Assessment of Educational Progress is given. In doing so, the report suggests, NAEP could identify factors that may differentiate high-performing states and urban districts from low performers.

The effort, it says, would parallel the extensive reporting of background variables in global assessment systems, such as the Program for International Student Achievement, or PISA.

The report was released just weeks after the Obama administration proposed a fiscal 2013 budget that would cut the NAEP budget by $6 million, while funding a pilot program of state participation in PISA.

"Currently, the NAEP background questions are a potentially important but largely underused national resource," says the report by a six-member expert panel commissioned by the National Assessment Governing Board, or NAGB, which sets policy for the testing program.

"These data could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education and the context for understanding achievement and its improvement."

In addition, the report says NAEP background questions could help track policy trends, such as implementation of the Common Core State Standards or new teacher-evaluation systems.

The report, presented this month to NAGB at a meeting in New Orleans, was apparently well-received by many board members, including the chairman, former Massachusetts Commissioner of Education David P. Driscoll. But some of the ideas are generating push back from current and former federal officials.

"NAGB has a tool that they want to use for everything," said Mark S. Schneider, a former commissioner of the National Center for Education Statistics, the arm of the U.S. Department of Education that administers the test. He argues that NAEP should stick to its core strengths, namely measuring student achievement and serving as a benchmark for state assessments.

"I find this just a distraction," Mr. Schneider said of the proposed plan.

Causation vs. Correlation

Although the report emphasizes the importance of not letting correlations between math achievement and rates of absenteeism, for instance, be confused for causation, Mr. Schneider argues that such distinctions would be lost on the public and risk damaging NAEP's reputation.
"They will make statements that will inevitably push the boundaries, and you will end up with questionable reports, in my opinion," said Mr. Schneider, who is now a vice president of the Washington-based American Institutes for Research. Other concerns raised about the proposals are the cost involved, especially given the president's proposed cut to NAEP, and what some experts say may be resistance to the federal government's collection and reporting of more information on students, given privacy concerns.

The new report, commissioned by NAGB, notes that complementing the NAEP tests is a "rich collection" of background questions regularly asked of students, teachers, and schools. But the collection and the public reporting of such information have been significantly scaled back over the past decade, the report says.

"NAEP should restore and improve upon its earlier practice of making much greater use of background data," the report says, "but do so in a more sound and research-supported way."

It offers recommendations in four areas related to the background questions: asking "important questions," improving the accuracy of measures, strengthening sampling efficiency, and reinstituting what it calls "meaningful analysis and reporting."

It's the fourth area, analysis and reporting, that is proving especially controversial.

Marshall S. "Mike" Smith, a co-author of the report and a former U.S. undersecretary of education in the Clinton administration, notes that the report comes at a time when NAEP's long-term relevance is at issue. He cites the work to develop common assessments across states in English/language arts and mathematics, as well as the growing prominence of international exams, like PISA.

"The future of NAEP is somewhat in doubt," Mr. Smith said.

PISA's use of extensive background questions, he said, has enabled it to have wide influence.

"They've built narratives around the assessments: Why are there differences among countries" in achievement, he said. "We can't do that with NAEP. We're not able to construct plausible scenarios or narratives about why there are different achievement levels among states. And we've seen that can be a powerful mechanism for motivating reform."

Mr. Driscoll, the chairman of NAGB, said the next step is for board staff members to draft recommendations on how the proposed changes could be implemented.

"I have challenged the board to think about how NAEP and NAGB can make a difference and have an impact," he said. "There is some very valuable information that we can lay out ... that would be instructive for all of us."

The report makes clear that NAEP should not be used to assert causes for variation in student achievement, but that a series of "descriptive findings" could be illustrative and help "generate hypotheses" for further study. For example, it might highlight differences in access to 8th grade algebra courses or to a teacher who majored in math.

"A valid concern over causal interpretations has led to a serious and unjustified overreaction," the report says.
But some observers see reason for concern.

"It's a mistake to present results that are purely descriptive," said Grover J. "Russ" Whitehurst, a senior fellow at the Brookings Institution in Washington who was the director of the federal Institute of Education Sciences under President George W. Bush. "It is misleading, and it doesn't make any difference if you have a footnote saying these results should not be considered causally."

Jack Buckley, the current NCES commissioner, expressed reservations about some of the suggestions, especially in the analysis and reporting of the background data.

"The panel is looking toward PISA as an exemplar," he said. "Folks at [the Organization for Economic Cooperation and Development, which administers PISA] write these papers and get a broad audience, but it's not always clear that the data can support the conclusions they reach about what works."

Mr. Buckley said he understands NAGB's desire to be "policy-relevant," but he cautioned that "we have to carefully determine what is the best data source for measuring different things."

Mr. Driscoll said he's keenly aware of not going too far with how the background data are used.

"I agree ... that we have to be careful about the causal effects," he said. "I think we've gone too far in one direction to de-emphasize the background questions, and the danger is to go too far in the other direction."
Overview
As part of the Board’s continuing outreach efforts, the Governing Board contracted with the Council of Chief State School Officers (CCSSO) in September 2007 to form a Task Force charged with providing state feedback and recommendations to the Board on NAEP policy areas and projects. The Task Force consists of 12 high-level state education agency staff members who were chosen based on expertise and interest in assessment, and geographic representation of the nation. Task Force members include:

- 1 chief state school officer
- 5 deputy superintendents
- 3 associate superintendents of accountability and assessment
- 3 public information officers

Schedule of Task Force Meetings
The Task Force convenes for two in-person meetings and four WebEx meetings annually.

Policy Issues
During the Task Force’s 29 meetings to date, they have addressed a number of key policy issues:

- NAEP reporting process
- Inclusion and accommodations
- NAEP schedule of assessments
- Reading trend line
- NAEP race/ethnicity reporting
- Common Core State Standards and Assessments
- Misuse and misinterpretation of NAEP data
- International benchmarking
- Board initiatives on raising achievement and closing gaps
- NAEP 12th grade preparedness

On each issue, Task Force members provided substantive input on these NAEP topics and made significant contributions in a variety of related areas. Task Force Chair Pat Wright will also provide a presentation to the Board on November 30, 2012 to provide an overview of the Task Force’s recent discussions.

Outreach
Beyond the Task Force meetings, members have addressed their peers on the group’s purpose and activities to date. Such venues have included briefings to state chiefs at CCSSO’s Legislative Conferences and to state assessment directors at meetings of the Education Information Management Advisory Consortium (EIMAC). Additionally, there have been panel presentations on the Task Force at the annual CCSSO National Conference on Student Assessment.
Policy Task Force Members (2012-2013)

**Patricia Wright, Task Force Chair**  
Superintendent of Public Instruction  
Virginia Department of Education  
Richmond, VA

**Deborah Sigman, Task Force Vice Chair**  
Deputy Superintendent  
California Department of Education  
Sacramento, CA

**David V. Abbott**  
Deputy Commissioner/General Counsel  
Rhode Island Department of Elementary and Secondary Education  
Providence, RI

**Liza Cordeiro**  
Executive Director, Office of Communications  
West Virginia Department of Education  
Charleston, WV

**Tom Foster**  
Director of Assessment  
Kansas State Department of Education  
Topeka, KS

**Pete Goldschmidt**  
Director of Assessment  
New Mexico Public Education Department  
Santa Fe, NM

**Lisa Y. Gross**  
Director, Division of Communications and Community Engagement  
Kentucky Department of Education  
Frankfort, KY

**Dan Hupp**  
Director of Student Assessment  
Maine Department of Education  
Augusta, ME

**Susie Morrison**  
Deputy Superintendent/Chief of Staff  
Illinois State Board of Education  
Springfield, IL

**Nate Olson**  
Communications Manager  
Washington Office of Superintendent of Public Instruction  
Olympia, WA

**Joel Thornton**  
Chief of Staff  
Georgia Department of Education  
Atlanta, GA

**Joyce Zurkowski**  
Director of Student Assessment  
Colorado Department of Education  
Denver, CO
National Assessment Governing Board
Council of Chief State School Officers
Policy Task Force

SUMMARY OF WORK TO DATE
DECEMBER 2007 - NOVEMBER 2012

The following charts summarize Task Force input on topics addressed since the Task Force first met in December 2007. In many areas, the Task Force has provided important input for ongoing projects, draft policies, or other documents as noted in the “Follow-up Activities” column. For areas where additional follow-up activities are possible, an asterisk (*) signals an opportunity for the Board to consider the Task Force’s suggestions.

This document represents an ongoing tracking of issues and topics discussed by the Task Force across eleven separate areas denoted as Topic #1 through Topic #11. Updates to this document since November 2011, are denoted in yellow highlighting. Topics with new Task Force discussion points are:

- Topic #4: Grade 12 NAEP
- Topic #6: NAEP Background Questions
- Topic #10: Board Initiatives on Raising Achievement and Closing Gaps
- Topic #11: Future of NAEP Initiatives

The last oral briefing from the Task Force was delivered at the November 2011 Board meeting. Task Force Chair Pat Wright will deliver the annual Task Force briefing on November 30, 2012.
# Topic #1: NAEP Reporting Process


## Task Force Discussion and Input

### Executive summary of NAEP Report Cards
Increase audience attention to the Executive Summary by conveying more clearly the key findings of each Report Card.

### Reporting socioeconomic status data
Some Task Force members discussed that the locality/district type classifications that result from current SES data collection procedures often yield district-level labels that do not match the actual socioeconomic conditions in the district. For example, districts may be classified as suburban even though the schools’ student populations mirror urban populations.

### Shaping development of Report Cards
Identify questions the data should attempt to answer. Use these questions and question types to shape Report Cards.

## Follow-up Activities

Report Cards have:
- Streamlined the overall presentation of findings in the Executive Summary
- Used lists to prominently display key findings
- Improved data displays for a general public audience
- Removed less essential footnotes

**

The Board is monitoring NCES-led efforts to improve socioeconomic status measures. Some of these efforts were piloted in the 2010 assessment administrations. A newly developed white paper on improving SES data collection will be presented at the November 29, 2012 meeting of the Ad Hoc Committee on Background Information.

**

NCES and Board staffs met in late 2009 with representatives of the 11 pilot states to discuss the 2009 report of grade 12 state-level results in reading and mathematics.

NCES and Board staffs are using new NAEP report formats, including *Findings in Brief for 2011*.

---

* Denotes Task Force input for future Board consideration
### Topic #1: NAEP Reporting Process

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>General NAEP talking points as a resource</td>
<td>NCES has been updating NAEP brochures for a general audience. See Topic #8’s follow-up activities for additional details on efforts to address misinterpretation of NAEP data.</td>
</tr>
<tr>
<td>Developed talking points to explain what NAEP is; and develop talking points about specific interpretation challenges in reviewing assessment results. These can be distributed to states, media, and other stakeholders.</td>
<td></td>
</tr>
<tr>
<td>Report-Card-specific talking points</td>
<td>This resource is provided to NAEP State Coordinators; and is also to be included in executive summaries of Report Cards.</td>
</tr>
<tr>
<td>Developed talking points to convey the “common message” in release of specific NAEP Report Cards.</td>
<td>The audit of Board communications conducted by Reingold and the May 2010 Board discussion on future directions for the Board suggest expanded efforts to support Report Card releases.</td>
</tr>
<tr>
<td>Anti-testing sentiment</td>
<td>In November 2010, the Board’s Reporting and Dissemination Committee recommended a communications plan that addresses this issue, and the Board adopted the plan.</td>
</tr>
<tr>
<td>Address anti-testing sentiment, highlighting how districts and schools benefit from NAEP.</td>
<td>The Board’s Reporting and Dissemination Committee recommended a communications plan that incorporates social media tools. The Board’s communication plan was adopted in November 2010.</td>
</tr>
<tr>
<td>Social media</td>
<td></td>
</tr>
<tr>
<td>Recognize that several social media sites support ongoing conversations, and if NAEP moves in this direction, efforts will be needed to maintain a continuous conversation.</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
## Topic #1: NAEP Reporting Process

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prioritizing NAEP audiences</strong></td>
<td>The Board’s Reporting and Dissemination Committee considered a communications plan that outlines NAEP’s target audiences. The Board’s communication plan was adopted in November 2010.</td>
</tr>
<tr>
<td>Prioritize target audiences, highlighting policy and business audiences.</td>
<td>The Board has been increasing outreach to the business community via the NAEP Business Policy Task Force. A business community outreach webinar was held to discuss the grade 12 NAEP results for the 11 states that participated in the grade 12 NAEP state pilot.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accessible informational resources</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider a more accessible format for the NAEP Data Explorer, such as an iPhone app.</td>
<td>NCES and Board staffs are working to develop ideas for apps using NAEP data and items.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Improving outreach to schools</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider the most effective outreach strategy for schools in each state. In some states, the NAEP State Coordinator has the ideal skill set to lead these efforts.</td>
<td>Several of these ideas are being considered by the Board as part of the action proposals addressing Topic #10: Board Initiatives on Raising Achievement and Closing Gaps.</td>
</tr>
<tr>
<td>Carefully craft messages about the intended use and purpose of NAEP.</td>
<td></td>
</tr>
<tr>
<td>Emphasize information beyond NAEP scores (e.g., profiles on what students are able to do and should be able to do).</td>
<td></td>
</tr>
<tr>
<td>Recognize that as more data are made available, there is a greater possibility for misinterpretation and inappropriate use of these data.</td>
<td></td>
</tr>
<tr>
<td>Maintain a balance between making NAEP meaningful as a significant driver of policy, and making NAEP meaningful at the school level.</td>
<td></td>
</tr>
<tr>
<td>Clarify the benefits of NAEP from the school perspective.</td>
<td></td>
</tr>
<tr>
<td>Capitalize on ways to use NAEP items and performance criteria that are consistent with the Common Core State Standards, given the large number of adopting states.</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
## TopiC#1: NAEP Reporting Process

### Task Force Discussion and Input

**Enhancing outreach to states**  
The Task Force provided the following suggestions on state use of NAEP information beyond score reports:
- Strategically engage state or consortia assessment programs to create a shared knowledge base and collaboration around lessons learned (e.g., webinars, users’ groups).
- Leverage NAEP college and job training preparedness efforts to support assessment consortia efforts on college and career readiness (e.g., setting a common college placement test score).
- Introduce NAEP data tools to educators during pre-service as a resource for teaching data use.

### Follow-up Activities

*  
Several of these ideas are being considered by the Board as part of the action proposals addressing *Topic #10: Board Initiatives on Raising Achievement and Closing Gaps*.

**NAEP and social media**  
Bridge traditional and new social media communications efforts by formatting Twitter submissions in a newsletter format.

### Improving pre-release Report Card briefings for governors and chiefs

- Highlight for chiefs major data changes and unexpected findings.
- Identify for chiefs a small number of policy issues present in several states. Identifying these issues in advance of the briefing and addressing a portion of the briefing to these issues would be useful.
- Highlight notable subscale trends at the state level.
- Continue to host joint pre-releases for NAEP state coordinators and public information officers; these are viewed as the most useful convenings.
- Support NAEP knowledge in new chiefs through NAEP Ambassadors.

### Providing embargo access to new media outlets

- Move slowly on this topic. As more media outlets are given access to NAEP results, issues will continue to arise.
- Continue to provide early access to states and districts before NAEP data are public in order to give adequate preparation time for media inquiries.
- Share with states how other federal agencies approach releasing embargoed information to the media.

* Denotes Task Force input for future Board consideration
**TOPIC#1: NAEP REPORTING PROCESS**

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
</table>
| **Considering a policy on embargo practices**  
- Formalize current embargo practices through a written policy, with expanded access provided to certain new audiences.  
- Prioritize access to media outlets that report the news rather than those that provide opinions to keep the focus on accurate factual representation of data.  
- Clarify in the embargo process that readership is a criterion for determining whether the source should have early access.  
- Consult states regarding organizations or individuals who request embargo access and who lack media qualifications. If there is an established relationship between the state and the requestor, this may help with embargo access determinations. | The Board reviewed the NAEP embargo process and commissioned research on other organizations’ embargo policies to inform this review. This process resulted in a narrow set of NAEP embargo guidelines for traditional news sources and reporters on assignment. This set of embargo guidelines will be examined after several releases to determine whether broader access should be granted to non-traditional news sources. |

* Denotes Task Force input for future Board consideration
**TOPIC #2: NAEP INCLUSION AND ACCOMMODATIONS**


<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-Up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Misconceptions of accommodations’ usage</strong>&lt;br&gt;Address media questions on whether states are attempting to influence NAEP results by providing greater numbers of accommodations.</td>
<td>In recent Report Cards:&lt;br&gt;Provided more context on use of accommodations in a national assessment setting, which are generally consistent with state assessment practices, including a special Report Card section on inclusion and accommodations under the new policy on inclusion adopted by the Board on March 6, 2010.</td>
</tr>
<tr>
<td><strong>State demographics</strong>&lt;br&gt;Address variations in exclusions and accommodations due solely to different population characteristics in a state.</td>
<td>In recent Report Cards:&lt;br&gt;Improved explanation of likely reasons for state variations in proportion of special needs students, (especially English language learners), as well as state policy differences. These factors contribute to differences in exclusion and accommodation rates on NAEP.</td>
</tr>
<tr>
<td><strong>Basis for variance in exclusion rates</strong>&lt;br&gt;Expand the interpretation of results section in Report Cards to explain the basis for differences among states – the NAEP policy on accommodations and inclusion and state policies. This section should support states in communicating this issue to the press and districts.</td>
<td>*&lt;br&gt;See a summary of efforts to improve explanatory notes above.</td>
</tr>
<tr>
<td><strong>Expertise related to SD and ELL students</strong>&lt;br&gt;Include broad expertise relevant to students with disabilities and English language learners in the charge to the Board’s Ad Hoc Committee.</td>
<td>Members appointed to the Ad Hoc Committee had broad expertise in teaching, research, and policy related to SD and ELL students.&lt;br&gt;The Board widened the expertise brought to bear on this important issue by convening Expert Panels to provide further recommendations on the issues specific to English language learners and students with disabilities.</td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
### Topic#2: NAEP Inclusion and Accommodations

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Within-year and across-year variations</em>&lt;br&gt;Ensure that the Ad Hoc Committee addresses the variations in exclusion rates&lt;br&gt;within states over time, as well as variations among states in a given year.*</td>
<td>In the policy adopted by the Board on March 6, 2010:&lt;br&gt;A policy goal of high inclusion is emphasized—95 percent or higher for all students. Below an 85-percent participation rate for SD students or ELL students, reporting will prominently designate these participation rates in NAEP Report Cards.</td>
</tr>
<tr>
<td><em>Determining inclusion for each student</em>&lt;br&gt;  ▪ Add a NAEP appendix or a checkbox to individual education programs (IEPs) in case the student is sampled for NAEP, to indicate whether the student should participate in NAEP and how. States are concerned about consistency of decision protocols to exclude students at the school level, both from building-level staff and NAEP contractor staff. A coordinated effort is needed.*</td>
<td>Based on the Board-commissioned research to identify model rules for uniform national criteria, the Expert Panel recommendations, and the public comments received, the final policy developed by the Ad Hoc Committee has determined that that modifying IEP forms should not be pursued at this time.</td>
</tr>
<tr>
<td>▪ Create a guidance document to support state development of IEP templates.</td>
<td></td>
</tr>
<tr>
<td><em>The current decision trees</em>&lt;br&gt;Some Task Force members suggested that the Board standardize NAEP inclusion because it is a national test administration, and standardization supports NAEP’s credibility. Other Task Force members disagreed, citing interruptions in the day-to-day accommodations the student already receives and different state definitions of English language learners.&lt;br&gt;  ▪ Start with a least invasive approach, and then scale up if needed.&lt;br&gt;  ▪ Examine how current policy can align with Title I.&lt;br&gt;  ▪ Examine how prospective changes will be communicated to IEP teams.*</td>
<td>In the policy adopted by the Board:&lt;br&gt;The current decision tree is to be replaced with a new decision tree that standardizes participation in NAEP. An operational definition for ELL is provided for NAEP, and a language screener is supported as a future research and development effort.</td>
</tr>
<tr>
<td>In considering the possibility of uniform national rules:&lt;br&gt;  ▪ Uniform national rules may have unintended negative consequences if more students opt out. Participation and inclusion rates may decrease.&lt;br&gt;  ▪ Consistency in inclusion is important to maximize comparability. Using the language screener would allow case-by-case determinations for each student.&lt;br&gt;  ▪ Policy guidance in defining ELL subgroups could provide comparability.</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
<table>
<thead>
<tr>
<th><strong>Task Force Discussion and Input</strong></th>
<th><strong>Follow-up Activities</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feasibility of new accommodations</strong> Investigate the feasibility of new accommodations to increase inclusion. Task Force members discussed that it is not clear from research whether reading aloud and permitting calculators threatens construct validity.</td>
<td><strong>∗</strong> In the policy adopted by the Board: The guiding principles indicate a high priority for maximum inclusion of students while also maintaining the validity and comparability of collected data. Research is also a key component of the policy. The Governing Board has received presentations from NCES on various studies that address increased accessibility of NAEP. The Board will receive ongoing updates on this topic.</td>
</tr>
</tbody>
</table>
| **Prominence of exclusion rate data**  
- Add an indicator for level of exclusion adjacent to the NAEP Report Card achievement level state-by-state bar chart to increase attention to the inclusion issue and to provide clearer information. Footnotes and appendices can be easily overlooked.  
- Use the next administration of NAEP to encourage maximum inclusion of SDs and ELLs without flagging jurisdictions whose inclusion rates do not meet the 95% and 85% goals.  
- Identify states’ distance from the 85% inclusion goal to motivate improvement. | In the policy adopted by the Board: Below an 85-percent participation rate for SD students or ELL students, reporting will prominently designate these participation rates in NAEP Report Cards. |
| **Guidance for states** Provide compelling guidance on this issue, similar to how NCLB regulations prompted states to align with 1% and 2% waivers for students with disabilities. | In the policy adopted by the Board: Resources that clearly outline the purpose and value of NAEP and of full student participation in the assessment are integral to encouraging high participation rates. |

* Denotes Task Force input for future Board consideration
### TOPIC#2: NAEP INCLUSION AND ACCOMMODATIONS

<table>
<thead>
<tr>
<th><strong>Task Force Discussion and Input</strong></th>
<th><strong>Follow-up Activities</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Considerations offered for the Ad Hoc Committee’s deliberation</strong></td>
<td>In the policy adopted by the Board:</td>
</tr>
<tr>
<td>- What is NAEP’s goal regarding inclusion? Does NAEP aim to include all students? A mission statement on this issue is one way to clearly communicate NAEP’s goal.</td>
<td>The guiding principles indicate a high priority for maximum inclusion of students while also maintaining the validity and comparability of collected data. Research is also a key component of the policy.</td>
</tr>
<tr>
<td>- States vary greatly in the type of accommodations allowed. States with many accommodations are perceived as trying to influence their NAEP results.</td>
<td>Comments from both the Task Force and EIMAC were collected and considered.</td>
</tr>
<tr>
<td>- For comments on prospective policy options, the Education Information and Management Advisory Consortium (EIMAC) may be an avenue for discussion in order to get a more representative response from states.</td>
<td></td>
</tr>
<tr>
<td>- In considering research on accommodations, there is a solid body of knowledge on accommodations, especially since NCLB was enacted. This should be used as a resource as the Board moves forward.</td>
<td></td>
</tr>
<tr>
<td><strong>Policy suggestions for the Ad Hoc Committee</strong></td>
<td></td>
</tr>
<tr>
<td>- Adopt a guiding principle that includes language about maximizing meaningful participation in order to provide the most accurate assessment of student performance.</td>
<td>In the policy adopted by the Board:</td>
</tr>
<tr>
<td>- Focus on maintaining the purpose of NAEP.</td>
<td>As noted above, the guiding principles indicate a high priority for maximum inclusion of students while also maintaining the validity and comparability of collected data. Also, accommodations and modifications are distinguished.</td>
</tr>
<tr>
<td>- Encourage NAEP to accommodate to students.</td>
<td></td>
</tr>
<tr>
<td>- Distinguish modifications, which change what is being measured, from accommodations.</td>
<td></td>
</tr>
<tr>
<td><strong>Feedback on preliminary recommendations of SD Expert Panel</strong></td>
<td>In the policy adopted by the Board:</td>
</tr>
<tr>
<td>- Clarify whether the 95% participation rate mentioned in the fourth recommendation includes or excludes students with significant cognitive disabilities.</td>
<td>The participation rates and the reporting of IEP and 504 students are clarified in the policy. A list of NAEP-appropriate accommodations as well as non-allowed accommodations is to be provided to schools.</td>
</tr>
<tr>
<td>- Clarify language about reporting NAEP results separately for IEP and 504 students. The Task Force noted that data variability within and across states regarding 504 students may make it difficult to disaggregate these data.</td>
<td></td>
</tr>
<tr>
<td>- List accommodations that are not permitted by NAEP instead of those that are permitted by NAEP. Specifying accommodations permitted by NAEP may inadvertently limit states from employing acceptable accommodations.</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
## TOPIC#2: NAEP INCLUSION AND ACCOMMODATIONS

### Task Force Discussion and Input

#### Feedback on preliminary recommendations of ELL Expert Panel
- Disaggregating data on ELLs, including by English language proficiency, will create a significant data burden on under-resourced state education agencies. Additionally, the data requested may be out of date, and both English proficiency assessment cut scores and proficiency levels vary across states.
- Instead of disaggregating students by proficiency level, consider comparing ELLs and former ELLs. This would be less problematic in terms of the data burden placed on states, and it would still distinguish students along the development continuum.

#### Feedback on final recommendations of the Expert Panels
- While endorsing the policy principles, the Task Force cautions against adding new requirements for states, such as the collection of additional data elements.
- Task Force members appreciate the emphasis placed on fairness to students, equity across states, and maximum inclusion.

### Follow-up Activities

#### In the policy adopted by the Board:
- Support is provided for data displays of ELLs and former ELLs.

#### In outreach efforts with stakeholders:
- Data collection and other implementation issues are being addressed.

### Outreach to support implementation

- Use NAEP ambassador meetings as an opportunity to gain high-level support for field-level implementation.
- Ensure that communication efforts do not overlap with peak test release time in the states.
- Send out early, succinct communication points to schools to begin the preparation phase.
- Collaborate with assessment teams to ensure all rules are being followed as closely as possible.
- Develop a PowerPoint and post it on the web to allow users to tailor it to their own special needs. Include a brief timeline providing an historical context for NAEP policy changes.

#### Various outreach efforts have been implemented with workshops held for states and districts participating in the Trial Urban District Assessment (TUDA) on implementing the Board’s new SD/ELL policy.

### Reporting NAEP data under the new inclusion policy

- The margin of error should be generous enough to avoid penalizing states that have the same inclusion rate but have different sample sizes. If it is possible that different states could have the same participation/inclusion rates with one being flagged and the other not being flagged, then communication will be needed to help the public understand the differences.
- Provide general descriptive information in the reports about state and NAEP allowable accommodations to explain why students are excluded.
- Display the percentages of students with disabilities and English language learners in addition to the participation rates.
- Develop a communication plan with advance notice to states and talking points to probe thinking before the pre-release workshop. Include public information officers in the pre-release workshop and identify issues that are expected to be confusing to media and the public.

#### The first round of NAEP reports under the new inclusion policy were released on November 1, 2011. These reports included a special Report Card section on inclusion and accommodations under the new policy.

---

* Denotes Task Force input for future Board consideration
## TopiC#3: NAEP SChedule of Assessments

*Addressed in December 13, 2007 in-person meeting and April 22, 2011 WebEx.*

### Task Force Discussion and Input

<table>
<thead>
<tr>
<th>Key subjects and grades for states</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consider focusing on grades 4 and 8 reading and mathematics if Congress does not increase funding for NAEP. NAEP grades 4 and 8 are most valuable in state policy making. NAEP should also assess state-level science and writing.</td>
<td>Congress appropriated a $10 million increase for NAEP in Fiscal Year 2008. Therefore, cuts to the NAEP schedule of assessment were not needed for Fiscal Year 2008. In May 2009, the Board decided to administer the 2011 NAEP Writing assessment at the national level only in both grades 8 and 12 for this first-ever computer-based NAEP assessment.</td>
</tr>
</tbody>
</table>

**Grade 12 NAEP**

Several Task Force members said that grade 12 should not be tested in NAEP, partly because state standards on secondary school specialty subjects diverge and the challenge of motivating students in grade 12 is formidable. These Task Force members supported the use of currently existing assessment instruments and other indicators of college preparation to identify whether an information base already exists for the progress of 12th graders.

Other Task Force members indicated some support for testing at grade 12, highlighting that 12th graders might be more motivated to take a writing assessment than a multiple-choice test, partly because a writing assessment offers them an opportunity to express themselves. Also, grade 12 is important to providing a comprehensive view of assessment. Some of these Task Force members supported assessment of subjects on a cohort-basis so that the same cohort of students would be tested at grades 4, 8, and 12, and progress would be observed for each cohort.

In May 2010, the Board made several changes to the NAEP assessment schedule. One change was to increase the frequency of the grade 12 NAEP reading and mathematics assessments while providing for continued voluntary state-level participation. Both of these changes are also aligned with the Board’s preparedness reporting initiative.

---

* Denotes Task Force input for future Board consideration
# Topic#3: NAEP Schedule of Assessments

## Task Force Discussion and Input

*Frameworks as a resource for states*

Consider not only the benefit of the test data but the benefit of the NAEP frameworks when considering the schedule of future assessments. The Task Force acknowledged that NAEP frameworks are especially helpful to states—they are “keystone” documents. Many states use NAEP frameworks as a resource when revising their elementary and middle school standards.

## Follow-up Activities

* The Board is examining ways to increase dissemination of NAEP frameworks through the Web and other outreach activities.

The new Governing Board communications contractor Reingold has suggested several strategies for making NAEP frameworks more accessible to a wider audience.

The Board has begun to create interactive web-based versions of the NAEP frameworks. The first interactive framework, for the Technology and Engineering Literacy (TEL) assessment, was completed in February 2011.

## Considerations on Computer-Based Grade 4 Writing

- Although some Task Force members noted that NAEP computer-based writing assessment at 4th grade may signal where we need to be as a nation, other Task Force members provided the following considerations for implementing a valid computer-based writing assessment at grade 4:
  - Lack of experience. Students vary in their classroom experience with computers at grade 4. Several states do not begin computer-based writing assessment until grade 5. Some members voiced support for NAEP computer-based writing at grade 4 despite variability in students’ experience.
  - Developmental concerns. Some teachers are concerned that prior to grade 5, students may lack the developmental ability to compose writing on a computer.
  - Capacity. State technology capacity remains a concern. Some states are transitioning to grade 4 computer-based assessments, but this is very uneven across the states.

* Denotes Task Force input for future Board consideration
**Topic#3: NAEP Schedule of Assessments**

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning the NAEP Schedule through 2022</td>
<td></td>
</tr>
</tbody>
</table>
| ▪ *ESEA subjects.* Focusing NAEP resources on subjects mandated to be tested in the reauthorization of the Elementary and Secondary Education Act will help to align with national priorities. For example, if science is a higher priority in the reauthorization, then science would appear more frequently on the NAEP schedule of assessments. | *
| ▪ *NAEP reading and mathematics.* Maintaining NAEP’s audit function by testing a variety of subjects as distinct from the Common Core State Standards may be useful to states, but some Task Force members questioned the need for future NAEP testing in reading and math. Criteria and considerations for priorities should include the Common Core and where NAEP can provide the best information to states. | Further action on the schedule may be needed in early 2013, depending on the NAEP budget situation. |
| ▪ *Subjects other than NAEP reading and math.* Given future alignment among states and between consortia assessments and NAEP, there will be a heightened need for NAEP data in subjects not typically assessed by states, such as economics, civics, and the arts. | |
| ▪ *Common Core-NAEP alignment.* If revision to NAEP frameworks is planned, then align the NAEP frameworks to the Common Core State Standards on a more aggressive schedule in order to implement aligned tests earlier than proposed on the draft NAEP assessment schedule. | |
| ▪ *State readiness for computer-based delivery.* Obtain a status summary on whether and how states are administering computer-based assessments. There is a need for caution and flexibility in the schedule given the magnitude of changes and questions about capacity. | |

* Denotes Task Force input for future Board consideration
**TOPIC #4: GRADE 12 NAEP**


<table>
<thead>
<tr>
<th>TASK FORCE DISCUSSION AND INPUT</th>
<th>FOLLOW-UP ACTIVITIES</th>
</tr>
</thead>
</table>
| *Facilitating transparency of the process*  
Clarify purpose and timelines for reporting preparedness. | The issuance of the Technical Panel Final Report addresses this information. Task Force members also recommended research priorities, which were taken into account when the Board adopted the Program of Preparedness Research at the March 2009 quarterly Board meeting. |
| *Using multiple tests*  
Several Task Force members noted that states appreciate the ability to use multiple existing tests and minimize the testing burden. | Based on the discussions of the Technical Panel, the Board is pursuing various statistical relationship studies. Some of these studies examine NAEP performance in relation to performance on other assessments widely used as indicators of preparedness. |
| *Addressing the motivation of 12th graders*  
Some Task Force members noted that 12th graders would not be motivated unless stakes were attached to NAEP. Motivation of 12th graders taking NAEP remains a concern of the Task Force.  
Address the firm preconception that grade 12 students will not be motivated. Even if participation levels are high, it will take more than compelling data to gain support. | Several of the studies in the Board’s Program of Preparedness Research relate NAEP performance to performance on other assessments. These analyses may provide a rich opportunity to examine motivation concerns for grade 12 NAEP examinees.  
The Board received an embargoed briefing in May 2009 and November 2009 regarding participation data and results of NAEP initiatives to increase participation and motivation. Another briefing on grade 12 participation and engagement was provided at the November 2010 Board meeting. |

* Denotes Task Force input for future Board consideration
## TOPIC#4: GRADE 12 NAEP

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Reporting results from the state-level grade 12 NAEP pilot</em>&lt;br&gt;Some Task Force members noted that the grade 12 state pilot should be reported in a similar fashion as other NAEP assessments. Other Task Force members suggested that the pilot nature of the project be stressed. Acknowledge limitations in generalizing findings from an 11-state pilot.</td>
<td>NCES and Governing Board staff met with pilot states to discuss reporting issues and gather recommendations for 2009 reporting of grade 12 state-level results. The Board has adopted a resolution that calls for the first round of preparedness research findings to be reported separately from the standard NAEP Report Card data.</td>
</tr>
</tbody>
</table>

| Pilot status of the grade 12 NAEP state-level assessment<br>Provide the opportunity in the future for the Task Force to review data and preparedness research findings to further identify whether the “pilot” status of the state-level grade 12 NAEP assessment is appropriate. | * |

| Benchmarking performance at grade 12<br>Considering that most state assessments are directed at the 10th or 11th grade level, some Task Force members questioned the value of benchmarking performance at a 12th grade level. | * |

| Prospective information to collect from examinees<br>Some Task Force members expressed interest in revising the student questionnaire, expanding the sample to allow for disaggregation based on student postsecondary options to further contextualize results, and setting achievement levels regarding preparedness. | The Board had already approved the 2009 background questionnaire, but it does include questions related to educational aspirations and postsecondary plans. The 2009 Grade 12 NAEP Report Card presents data regarding student aspirations. The Board convened an expert panel on background questions, and an Ad Hoc Committee has been established to explore enhancements to NAEP background information. |

* Denotes Task Force input for future Board consideration
## Topic#4: Grade 12 NAEP

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preparedness definitions</strong>&lt;br&gt;Clarify the meaning of preparedness and its relationship to states’ and other organizations’ definitions of preparedness and readiness. Some members expressed concern that NAEP’s definition could lead to confusion, similar to the different definitions of proficient.</td>
<td>*&lt;br&gt;The Technical Panel Final Report outlines more explicitly the meaning of preparedness in the NAEP context. At the November 2008 Board meeting, COSDAM affirmed the importance of the Panel’s work in this regard.&lt;br&gt;In March 2009, the Board adopted a working definition of preparedness in the NAEP context to be refined during the course of the Program of Preparedness Research.</td>
</tr>
<tr>
<td><strong>Engaging multiple stakeholder groups</strong>&lt;br&gt;Recognize the potential for the visibility of NAEP preparedness to encourage conversations between K-12, the business sector, and higher education institutions regarding the requirements for success after high school. This could be an opportunity to engage stakeholders.</td>
<td>The Technical Panel Final Report calls for subject matter experts to represent various stakeholder groups. At the November 2008 Board meeting, COSDAM affirmed the importance of this recommendation.&lt;br&gt;The Board has engaged a commission comprised of Board members and other individuals with experience in government, industry, and education. The commission’s focus is to communicate with a wide range of audiences regarding information and plans for reporting grade 12 NAEP results in terms of preparedness.</td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
**TOPIC#4: GRADE 12 NAEP**

<table>
<thead>
<tr>
<th>TASK FORCE DISCUSSION AND INPUT</th>
<th>FOLLOW-UP ACTIVITIES</th>
</tr>
</thead>
</table>
| *Implications for state accountability systems*  
Anticipate the implications of NAEP preparedness reporting on state accountability systems, state high school testing, state P-16 conversations (including the establishment of statewide college placement cut scores), other organizations’ definitions of readiness/preparedness (ACT, states, Achieve), state graduation rates, international benchmarking, state legislative school improvement funding, and state reporting on individual students. What will the NAEP preparedness indicators mean in this context, and how will they add value? Clarify what the preparedness indicators do and do not indicate. | *  
The Technical Panel Final Report has addressed some of these issues and has outlined scenarios to more clearly indicate what NAEP preparedness indicators mean and what they do not mean.  
In 2012, the Board has discussed extensively how to communicate research findings from the Board’s Program of Preparedness Research. |
| *Distinguishing 12th grade preparedness from 12th grade proficiency*  
NAEP proficiency and preparedness results will be compared, and their meanings should be clear. Determine the value of continuing to report proficiency and what value NAEP preparedness can and should provide to states; consider using preparedness in place of proficiency. Achieve has asked states to equate proficiency with readiness for credit-bearing coursework. | *  
In May 2009 and August 2009, there were joint sessions for the Board’s COSDAM and R&D committees to discuss both the technical and reporting perspectives of prospective preparedness statements in NAEP Report Cards. The clarity of achievement levels and preparedness indicators was a key consideration in the joint discussion.  
The Board has engaged a commission comprised of Board members and individuals with experience in government, industry, and education. The commission’s focus is to communicate with a wide range of audiences regarding information and plans for reporting grade 12 NAEP results in terms of preparedness. |
| *Research projects*  
Utilize careful standard-setting processes, post validity studies, and a clear narrow scope when developing the preparedness indicators, in the context of the broad range of evidence needed to determine students’ preparedness. | The Program of Preparedness Research adopted by the Board in March 2009, addresses the importance of post validity studies through studies that examine NAEP data relative to outcome indicators available in longitudinal datasets. |

* Denotes Task Force input for future Board consideration
## Topic#4: Grade 12 NAEP

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
</table>
| Technical documentation as a resource  
Contextualize the results in a comprehensive technical report stressing the pilot status of the state-level results. The report should include important contextual factors such as student participation, motivation, and preparation as well as synthesizing findings from the studies suggested by the Technical Panel on 12th Grade Preparedness Research. | A comprehensive technical report will be released for each phase of the Board’s program of preparedness research. The phase 1 report will be released in late 2012. |

**Implications of changing NAEP to a measure of preparedness**  
The Task Force noted the dramatic change for NAEP from being a measure of progress to a measure of preparedness. NAEP assessments at grades 4, 8, and 12 have not been used to report on preparedness for the next educational stage. Some members cautioned that this may not be an advisable avenue for NAEP, while other members felt that the Governing Board’s focus on academic preparedness, as opposed to readiness, is appropriate.  

**Contextual information to consider with preparedness indicators**  
The high school graduation rate is a preparedness indicator to the public. Differences between NAEP preparedness determinations and graduation rates will be controversial. The National Governors Association compact rate has focused the discussion on all students, not just students who reach the 12th grade. Some Task Force members noted providing graduation rate information may not fit with the focus of NAEP reports.  

**Clarifying the definition of a 12th grade student**  
The Task Force was not clear about whether students are considered 12th graders based on their high school cohort or based on credits earned. Since over-aged and under-credited high school students are disproportionately concentrated in large urban cities, an examination of this issue in Trial Urban District Assessment (TUDA) districts may be helpful.  

* Denotes Task Force input for future Board consideration
## Topic#4: Grade 12 NAEP

### Task Force Discussion and Input

**Considering the Board’s working definition of preparedness**

Task Force members agreed that having separate definitions for college and workplace is worrisome—ideally, students would leave the school system prepared for work and college. Also, Task Force discussion included the idea of tiers (or degrees) of preparedness, which should be clearly communicated in the definition of preparedness.

- Thoroughly consider the implications of these definitions. Given NAEP’s high visibility and the lack of consensus among other preparedness initiatives, these preparedness determinations will receive a lot of attention.
- Provide more clarity in the college preparedness definition by noting it refers to all postsecondary institutions (e.g., four- and two-year colleges).
- Include a statement noting that there are common skills that need to be mastered to enter either college or the workplace.
- Involve the career and technical education community in the development and refinement of the definition.

**How NAEP preparedness can be useful to states**

- Reporting for broader representative groups of students (instead of reporting solely on college-bound students, for example)
- Providing a system evaluation (as opposed to student-level information)
- Reporting preparedness for specified postsecondary education environments rather than treating postsecondary education as monolithic
- Anchoring NAEP cut scores to external reference points
- Serving as a possible analysis tool, e.g., relating preparedness information to subscale performance
- Combining the transcript study with NAEP administration to link course taking with performance.
- Reporting degrees of preparedness rather than using a dichotomous approach

### Follow-up Activities

In March 2009, the Board adopted a working definition of preparedness in the NAEP context to be refined during the course of the Program of Preparedness Research.

* In May 2010, the Board made several changes to the NAEP assessment schedule. One change was to increase the frequency of the grade 12 NAEP reading and mathematics assessments while providing for continued voluntary state-level participation. Both of these changes are also aligned with the Board’s preparedness reporting initiative.

As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed.

**Considering NAEP reading results and preparedness indicators**

A cross grade reading scale would allow appropriate comparisons between 4th and 8th grade data. However, the cross grade scale may be more appropriate at grades 4 and 8 than at grade 12, especially with the new focus on preparedness.

* Denotes Task Force input for future Board consideration
### Topic#4: Grade 12 NAEP

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suggestions regarding grade 12 NAEP preparedness reporting</strong></td>
<td>*</td>
</tr>
<tr>
<td>- Hold pre-release data-free briefings to facilitate better understanding of the results.</td>
<td>The emphasis that the first report is an initial step has been addressed in the outreach conducted by the NAEP 12th Grade Preparedness Commission. Symposia have been held in several states across the country over the past few years.</td>
</tr>
<tr>
<td>- Emphasize that the first report is just one piece of information, and that additional findings will be released as part of an iterative process.</td>
<td>*</td>
</tr>
<tr>
<td>- Incorporate the use of other measures to determine preparedness, e.g., transcript data.</td>
<td>As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. The Board has been holding ongoing conversations with the assessment consortia.</td>
</tr>
</tbody>
</table>

| Key considerations regarding defining grade 12 NAEP preparedness                                | *                                                                                 |
| - Consider whether differences exist between preparedness and proficiency at grade 12.         | As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. |
| - Consider whether this is meant to be a measure of those students in grade 12 or the cohort entering high school at grade 9. | *                                                                                 |
| - Consider effects of the Common Core State Standards Initiative—it’s definition of “readiness” seems to be headed in a different direction compared to the Governing Board’s preparedness initiative. | As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. |
| - Hold joint conversations between NAEP, consortia, and vendors to support comparability. There should be efforts to align definitions of readiness and preparedness. | *                                                                                 |

| Preparedness and readiness terminology                                                          | *                                                                                 |
| The Task Force expressed concern that NAEP and the states may be approaching a potential communication problem similar to the communication challenges caused by the differences between NAEP and state definitions of proficient. NAEP is using the term preparedness to mean the same thing as the term readiness used by the U.S. Department of Education, states, the Common Core State Standards, and the assessment consortia. | As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. |

| Strategies for using NAEP academic preparedness data                                            | *                                                                                 |
| - Focus on the likelihood argument to connote a range of performance, e.g., x% are likely to not need remedial/developmental coursework; avoid the precise argument. | As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. |
| - Frame the results in terms of establishing a baseline of performance.                         | *                                                                                 |
| - Tie the results to the common goal of Common Core readiness.                                 |                                                                                 |
| - Emphasize the definition of preparedness being used.                                          |                                                                                 |
| - Distinguish between the needs of colleges and careers on a continuum, e.g., two-year, four-year, types of careers. |                                                                                 |
| - Include comparisons with tests administered by the business community to strengthen the validation, e.g., industry certification. |                                                                                 |

* Denotes Task Force input for future Board consideration
**TOPIC#4: GRADE 12 NAEP**

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task Force feedback on reporting recommendations</strong></td>
<td>*</td>
</tr>
<tr>
<td>- Expanding NAEP’s reporting role. NAEP’s traditional role of reporting student performance at the national and state levels appears to be expanding to include reporting preparedness.</td>
<td></td>
</tr>
<tr>
<td>- Encouraging student tracking. An unintended consequence of readiness and preparedness reporting may be: reinforcing the tracking of students.</td>
<td></td>
</tr>
<tr>
<td><strong>Task force suggestions for next steps in preparedness research and reporting</strong></td>
<td></td>
</tr>
<tr>
<td>- Use nuance to avoid making statements that appear to relate to individual students.</td>
<td>As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed.</td>
</tr>
<tr>
<td>- Continue to collaborate on aligning consortia readiness and NAEP preparedness to avoid confusion and validity concerns.</td>
<td></td>
</tr>
<tr>
<td>- Consider segmenting the types of colleges and careers that are represented by NAEP performance labels to provide more fine-grained information.</td>
<td></td>
</tr>
<tr>
<td>- Continue studying readiness for entry into job training programs in order to make statements about academic preparedness for this area.</td>
<td></td>
</tr>
<tr>
<td>- Ensure that the career preparedness discussions are focusing on what current employers expect.</td>
<td></td>
</tr>
</tbody>
</table>

**Feedback: Advantages and disadvantages of a 2012 progress report on NAEP preparedness research**

The Task Force valued more heavily the option of releasing the progress report on preparedness in 2012. The option of releasing the NAEP preparedness report after later phases of research are completed was less favored.

- One of the chief advantages to a 2012 release is showing that the Governing Board’s work is relevant to the conversation of what preparedness means. Additionally, this is a critical time to be in the conversation.

- Challenges of releasing the progress report at this time: There are several competing definitions of preparedness and college and career readiness. There is concern about how the average person will understand these multiple definitions. How do they make sense of this without losing faith in the assessment industry? How does someone understand what it means to be prepared for college and career?

* Denotes Task Force input for future Board consideration
### Topic#4: Grade 12 NAEP

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Feedback:</strong> Communication strategies to avoid misinterpretation of the progress report</td>
<td></td>
</tr>
<tr>
<td>▪ Some of the research raises questions about the relevance of certain skills tested by NAEP for students entering job training programs. More information on this feedback may be helpful and the Board should be particularly cautious in how this information is communicated.</td>
<td></td>
</tr>
<tr>
<td>▪ It is important to consider how college readiness is discussed in relation to career readiness.</td>
<td></td>
</tr>
<tr>
<td>▪ There is a need to improve the language regarding: what does preparedness mean and what does readiness mean? These questions should be addressed in the release.</td>
<td></td>
</tr>
<tr>
<td>▪ Discussion among NAEP and the two main assessment consortia (PARCC and Smarter Balanced) should be held soon to address this issue.</td>
<td></td>
</tr>
</tbody>
</table>

| * |
| Initial conversations with the Smarter Balanced leadership have been initiated. CCSSO has also offered to convene the respective leadership teams as a neutral party to facilitate this effort. |

* Denotes Task Force input for future Board consideration
**TOPIC #5: NAEP READING TREND LINE**

*Addressed in February 22, 2008 and March 11, 2008 WebExes; May 29, 2008 in-person meeting.*

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technical decision and process</strong>&lt;br&gt;Determine criteria and process for evaluating feasibility of reading trend, and for reporting 2009 reading results.</td>
<td>COSDAM discussed options for criteria and timelines for decision-making.</td>
</tr>
</tbody>
</table>

**Short-term: Communication Plan for 2009**
- Communicate what has happened early through a communication plan that is tiered to multiple audiences based on expertise and interest.
- Start communication messages with the rationale for the change, the overall issue, and key milestones.
- Follow with the details by repeating interlocking messages as each milestone is reached to reinforce the rationale for the change.
- Do not underestimate how often to repeat the message.
- Frame the message positively. Avoid negative words like “break.” Use positive terms such as “create” or “develop.”
- Be as transparent as possible in communicating studies of content and statistical linking.

The 2009 NAEP Reading Report Card was released. Additional resources were developed to address how trend results were communicated, given the change in frameworks.

**Long-term: Policy Development**
- Recognize implications that the reading trend decision will have on NAEP trends in future subject area assessments.
- Think carefully about the criteria (regardless of subject) for content and statistical linking.
- Consider the 2009 reading trend in the context of other trend areas. Consider conditions that necessitate a new trend and when a new trend is not needed.
- Consider focusing on what type of change was made and where that fits into the hierarchy of possible changes in order to drive decisions made about trend lines; statistical issues should be secondary.
- Consider how future trend line decisions will be affected by the current alignment between state tests and NAEP.

* Denotes Task Force input for future Board consideration

At the November 2009 meeting, the Committee on Standards, Design and Methodology noted several implications for future policy development. Discussions are expected to continue.
Task Force members shared their thoughts on potential implications that may arise as states transition toward the new race/ethnicity categories mandated by the Office of Management and Budget (OMB), noting that it may take years for data collection and related protocols to stabilize at the school-level. The Task Force’s discussions recognized that validity of state-to-state comparisons and reporting of racial/ethnic achievement gaps will likely be challenged by this complex issue. With states implementing the new race/ethnicity categorizations at different times, the Task Force also noted that the lack of a uniform transition will mean a lack of a standard baseline. In 2011, the Board convened an expert panel working group to provide recommendations for how to maximize the use of NAEP background questions. In the Task Force’s discussion of the working group 2012 recommendations report, the most useful variables to states were: career skills, school climate, parental involvement, and student expectations/aspirations.

<table>
<thead>
<tr>
<th>TOPIC #6: NAEP BACKGROUND QUESTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STRATEGIC CONSIDERATIONS FOR THE TRANSITION PROCESS</strong></td>
</tr>
<tr>
<td>- During the transition period, constituent groups may develop a perception that NAEP and states are not being as transparent as possible in communicating about this issue.</td>
</tr>
<tr>
<td>- States that adopt the new race/ethnicity guidelines early may be able to provide guidance about communication.</td>
</tr>
<tr>
<td>- Impact on states will depend on state demographics.</td>
</tr>
</tbody>
</table>

| **SUGGESTIONS FOR NAEP IMPLEMENTATION** |
| - Using data collected from the EDEN system of the U.S. Department of Education, examine data from both new and old systems to allow for comparisons. |
| - Allow many racial/ethnic categories to be reported in NAEP Report Cards, e.g., allowing the possibility of summing to over one hundred percent, if appropriate. |
| - Distinguish data collection and reporting issues. |
| - Involve policy and assessment staff in these ongoing conversations to assess full implications beyond data. |
| - Poll states using the CCSSO network to determine how varied states’ plans are. |

| **REPORTING RACE/ETHNICITY DATA UNDER THE NEW OMB GUIDELINES** |
| - Preserve trend lines by maintaining the old categories in the body of the report and reporting the new categories in the appendices. |
| - Specify the state’s racial/ethnic population and the status of the state’s racial/ethnic data collection when the test was administered, and provide general guidelines on how to interpret the NAEP data given the state’s demographic and data collection context. |

| **FOLLOW-UP ACTIVITIES** |
| - The Board is monitoring changes in race/ethnicity categories and their potential impact on trends. |

* Denotes Task Force input for future Board consideration
**Topic#6: NAEP Background Questions**

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
</table>
| **Implications for reporting background information**  
  - Balance “the amount of story” with the amount of substance in terms of NAEP data. Incorporating analyses of background variables in relation to achievement in the main Report Card may present correlations, which should be carefully explained to avoid causal inferences.  
  - Comparing “Common Core” and “non-Common Core” states may be problematic.  | * |
| **Considerations for item development**  
  - Be mindful of any increase in test length/time.  
  - There may be negative implications of adding background questions for students that are not related to in-school activities.  
  - The overall purpose of background questions should determine the scope of the questions.  | *  
  - The Board convened an expert panel on background questions, and an Ad Hoc Committee has been established to explore enhancements to NAEP background information. |
| **Determining data collection focus areas**  
  - For some states, the state longitudinal data system is the best source of student level demographic information.  
  - The background questions could gather information on career skills to provide insights into how to assess these skills. These data are not generally collected by states, but they would be useful to both states and parents.  | *  
  - The Board convened an expert panel on background questions, and an Ad Hoc Committee has been established to explore enhancements to NAEP background information. |

* Denotes Task Force input for future Board consideration
**Topic#6: NAEP Background Questions**

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task Force feedback on the 2012 expert panel recommendations report</strong></td>
<td>*</td>
</tr>
<tr>
<td>▪ Privacy concerns. There are concerns about the appropriateness of certain questions that could be asked and their purposes. Depending on the question, additional consents may be needed at the school or district level because of state-specific privacy concerns about maintaining student confidentiality.</td>
<td>The Board convened an expert panel on background questions, and an Ad Hoc Committee has been established to explore enhancements to NAEP background information.</td>
</tr>
<tr>
<td>▪ Common uses of NAEP at the state-level. States primarily use NAEP data for additional information to supplement state data. For example, states use their own data for root cause analysis and then use NAEP to help determine if emerging issues are shared across other states or specific to their home state.</td>
<td></td>
</tr>
<tr>
<td>▪ Purpose of the questionnaires</td>
<td>*</td>
</tr>
<tr>
<td>□ The purpose of the background questions and analyses should be to illuminate correlations between NAEP student performance and background questions.</td>
<td></td>
</tr>
<tr>
<td>□ The Task Force would appreciate more detailed information on this topic, such as the target audience for the information from background questions and the intended uses of background questions, including potential decisions these data are expected to inform.</td>
<td></td>
</tr>
<tr>
<td><strong>Task Force suggestions for future questionnaires</strong></td>
<td>*</td>
</tr>
<tr>
<td>▪ Make better use of existing NAEP background data in an accessible and useful format.</td>
<td>The Board convened an expert panel on background questions, and an Ad Hoc Committee has been established to explore enhancements to NAEP background information.</td>
</tr>
<tr>
<td>▪ Improve online NAEP data analysis tools to facilitate use of data by various audiences.</td>
<td></td>
</tr>
<tr>
<td>▪ Consider removing some existing questions to create space for new questions in order to maintain the existing time allotment for background questions.</td>
<td></td>
</tr>
<tr>
<td>▪ Focus some new questions on “career skills” and post-secondary plans. This information would be particularly useful to states (as previously noted).</td>
<td></td>
</tr>
<tr>
<td>▪ Clarify who will use any new information produced by the questionnaires and how the information will be used, to guide which questions should be asked.</td>
<td></td>
</tr>
<tr>
<td>▪ Explore different disaggregations. Aggregated state-level data may obscure meaningful differences in student performance.</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
As new milestones are reached in the Common Core State Standards initiative, Task Force discussions have provided insights about state perspectives on how NAEP’s role may evolve. Several of the Task Force’s recommendations on this topic are reflected in other topics, such as Topic #3 NAEP Schedule of Assessments and Topic #10 Board Initiatives on Raising Achievement and Closing Gaps. This has been an ongoing cross-cutting issue discussed in several areas.

### Topic #7: Common Core State Standards


<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Role of NAEP</strong></td>
<td></td>
</tr>
<tr>
<td>▪ NAEP should maintain its role as an independent monitor of student achievement in the short-term. There is tremendous value in NAEP trends, and NAEP is highly regarded.</td>
<td>In August 2008, August 2009, and May 2010, the Board heard presentations from the organizations spearheading the development of the Common Core State Standards (CCSS). At the November 2010 and August 2011 Board meetings, the Board heard a presentation from the two assessment consortia, whose work builds on the CCSS effort.</td>
</tr>
<tr>
<td>▪ NAEP may be able to serve as an anchor to judge the common core assessments, possibly by releasing a special set of items only to states. States could build assessments to anchor against NAEP as an indication of rigor.</td>
<td>At the November 2009 meeting, the consensus of the Executive Committee was that the Board should continue being proactive in following this initiative as it develops and to be supportive and cooperative in responding to requests from CCSSO and NGA.</td>
</tr>
<tr>
<td>▪ International measures could be the key contribution from NAEP that complements the Common Core.</td>
<td>At the May 2010 Board meeting in Milwaukee, Wisconsin, the Board began discussions about the future of NAEP. This discussion has been a recurring feature of future Board meetings.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NAEP communications and informational resources</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Differentiate NAEP and the Governing Board from the Common Core and other assessment initiatives. It is important to highlight the differences between NAEP and the assessments for the Common Core. If there are no differences, people may lose interest in NAEP.</td>
<td>At the May 2010 Board meeting in Milwaukee, Wisconsin, the Board began discussions to explore how NAEP can enhance the information it provides to the nation. This discussion has been a recurring feature of future Board meetings.</td>
</tr>
<tr>
<td>▪ Efforts should be pursued to avoid confusion between Report Card releases and releases related to the Common Core Standards initiative.</td>
<td></td>
</tr>
<tr>
<td>▪ Create a compare/contrast document that clarifies similarities and intended purposes of all assessments. Focus on comparison as opposed to “alignment.”</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
In discussions of **Topic #1: NAEP Reporting Process**, the Task Force has emphasized the growing issue of misuse and misinterpretation of NAEP data. The growing prominence of this as a standalone specific reporting issue has initiated **Topic #8: Misuse and Misinterpretation of NAEP Data**.

<table>
<thead>
<tr>
<th><strong>TOPIC #8: MISUSE AND MISINTERPRETATION OF NAEP DATA</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Addressed in September 11, 2009, November 3, 2009, and April 6, 2010 WebExes; July 20, 2010 meeting.</strong></td>
</tr>
</tbody>
</table>

In some states, NAEP data are being used to infer average NAEP scores at the school district level and in other states the differing percentage of students performing at Proficient on NAEP and at Proficient on the state exam is used to discredit state assessment programs. States are not seen as credible when responding to these types of critical research. Focused stand-alone materials for stakeholders should be developed using affirmative language to demonstrate how NAEP should and should not be used. The Governing Board should take a more active role in countering misuse of NAEP data.

- Consider developing a policy statement on the appropriate use of NAEP data.
- Create the following proactive products:
  - A statement from the Governing Board regarding recurring misuses.
  - A flyer to illustrate how NAEP data can be used; address how NAEP data should not be used by using affirmative language wherever possible.
  - A template letter from the Chairman of the Governing Board to respond to common misuses or misinterpretations that arise in op-ed pieces. This will be particularly useful in defending against intentional misuses of data.

<table>
<thead>
<tr>
<th><strong>TASK FORCE DISCUSSION AND INPUT</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resources for countering instances of misuse and misinterpretation</td>
</tr>
<tr>
<td>In some states, NAEP data are being used to infer average NAEP scores at the school district level and in other states the differing percentage of students performing at Proficient on NAEP and at Proficient on the state exam is used to discredit state assessment programs. States are not seen as credible when responding to these types of critical research. Focused stand-alone materials for stakeholders should be developed using affirmative language to demonstrate how NAEP should and should not be used. The Governing Board should take a more active role in countering misuse of NAEP data.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>FOLLOW-UP ACTIVITIES</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>* At the August 2009 Board meeting, COSDAM received a briefing from NCES on related efforts they are spearheading.</td>
</tr>
<tr>
<td>At the November 2009 Board meeting, COSDAM was briefed by Task Force member Teri Siskind, who provided a summary of the Task Force’s suggestions for addressing this issue.</td>
</tr>
<tr>
<td>The NAEP Validity Studies (NVS) panel, an expert advisory group to NCES, has initiated efforts to develop materials that address interpretation issues.</td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
**Topic#8: Misuse and Misinterpretation of NAEP Data**

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definitions of proficient</strong></td>
<td></td>
</tr>
<tr>
<td>Clarify the relationship between state and NAEP definitions of proficient as well as the conceptual underpinnings, e.g., the larger content coverage of NAEP assessments.</td>
<td>* In November 2009, the Board’s Committee on Standards, Design and Methodology received a briefing from Task Force member Teri Siskind that highlighted the Task Force’s concern about the usage of proficient. The Board is discussing future use of the term proficient and reauthorization legislation as part of the larger discussion on the future of the Governing Board and NAEP. This discussion began at the May 2010 quarterly Board meeting.</td>
</tr>
<tr>
<td>▪ NAEP performance levels are aspirational goals, developed in the early 1990s.</td>
<td></td>
</tr>
<tr>
<td>▪ State performance levels are accountability determinations developed in terms of grade-level performance as part of No Child Left Behind.</td>
<td></td>
</tr>
<tr>
<td>▪ Options:</td>
<td></td>
</tr>
<tr>
<td>□ Convey degrees of proficiency (e.g., basic proficiency, proficiency, and advanced proficiency). This suggested labeling would be easier for the public to understand in conjunction with states’ reports of proficiency.</td>
<td></td>
</tr>
<tr>
<td>□ Expand the interpretation of results section in Report Cards to include more explanation regarding differences in proficiency definitions.</td>
<td></td>
</tr>
<tr>
<td>▪ Consider whether the Board would like to support removing the term “Proficient” from state performance expectations in the reauthorization of the Elementary and Secondary Education Act (ESEA). The Task Force is ready to provide the Board with advice on this topic, if desired.</td>
<td></td>
</tr>
</tbody>
</table>

**Making data presentations easier to interpret**

Consider the following changes for NAEP reports:

- Translate the effect size for the reader. It is difficult for states to police interpretation of scales.
- Change the vertical scale so that it is not as easily misinterpreted. One possibility may be to include the grade-level of the student assessed in front of his/her score (e.g., 4-350 for fourth graders, 8-350 for eighth graders, and 12-350 for twelfth graders).
- Reconsider the presentation of state rankings. States are ranked higher or lower than each other even if the scale score differences are insignificant or nonexistent.
- Compare states with surrounding states or other demographically similar states. The public may be likely to rank states using the online tool without appropriate context or understanding. In discussing the potential for a mega-states report, Task Force members pointed out the vast demographic differences between the five most populous states and discussed the potential value of analyzing states with the largest populations of certain students (e.g., English language learners (ELLs) or Native American students). This sort of analysis may be more useful instead of focusing on overall student population size.
- Use upcoming Report Card release to issue caveats of what the data signify and how the data can and cannot be interpreted.

* Denotes Task Force input for future Board consideration
### Topic #8: Misuse and Misinterpretation of NAEP Data

#### Task Force Discussion and Input

*Outreach efforts with stakeholders*
Engage the following groups proactively:
- Engage national/AP/wire reporters for pre-release data-free briefings. Reinstate this with the mapping study.
- Convene PIOs for pre-release data-free briefings to help PIOs prepare their state’s reporters.
- Reach out to schools of journalism to raise the profile of these issues among faculty, who can then provide responsive training to their students.
- Engage the Education Writers Association to discuss use and misuse of data.
- Consider focusing on research organizations, e.g., foundations and think tanks.
- Consider state concerns about unscrupulous third party reactions to NAEP results.
- Provide a timely briefing on new Report Card releases for Public Information Officers (PIOs) and other communications stakeholders on the relevant contextual information, so that they are better equipped to deal with media inquiries.

*Supporting appropriate interpretations of NAEP data*
- Produce a brief document or brochure on sampling and incorporate frequent misconceptions.
- Produce more materials that are easy to repackage (plug) into a story, and this will naturally encourage reporters to use the materials.

#### Follow-up Activities

* Governing Board and NCES discussions are underway to consider how outreach can be expanded to support a better understanding of NAEP.

One key objective of the communications plan currently being reviewed by the Board is to strengthen the relevance and use of The Nation’s Report Card, expanding engagement with NAEP data and research and using Report Card releases as a high profile catalyst for continuous outreach that engages and informs audiences throughout the year.

The Board holds meetings with editorial boards for major news outlets around the country. These meetings address NAEP issues to raise awareness and provide clarifications that will improve reporting on NAEP.

The Board’s Executive Director is engaging in more speaking opportunities and presentations to various policy groups. This includes groups such as the American Educational Research Association (AERA) and the CCSSO Chiefs Policy Forum.

* Denotes Task Force input for future Board consideration
**TOPIC #9: INTERNATIONAL BENCHMARKING**

*Addressed in May 28, 2009 (as part of Topic #7: Common Core State Standards), January 25, 2010, and July 20, 2010 in-person meetings.*

With national and state-level support for gathering more information on how U.S. students compare with international peers, the Task Force is discussing important considerations for the Board’s future work in this area.

<table>
<thead>
<tr>
<th><strong>TOPIC#9: INTERNATIONAL BENCHMARKING</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TASK FORCE DISCUSSION AND INPUT</strong></td>
</tr>
<tr>
<td><strong>Considerations for possible future roles of NAEP</strong></td>
</tr>
<tr>
<td>- NAEP can be used as the international benchmark.</td>
</tr>
<tr>
<td>- Chiefs are concerned about overtesting, but there is urgency for international comparisons and a desire to improve efficiency and effectiveness.</td>
</tr>
<tr>
<td>- Embed non-secured NAEP items on state assessments to be used as a set of anchor data to determine alignment capabilities.</td>
</tr>
<tr>
<td>- International measures could be the key contribution from NAEP that complements the Common Core. With this in mind, it may be important for the nation to strongly consider grade 12 TIMSS participation.</td>
</tr>
<tr>
<td><strong>FOLLOW-UP ACTIVITIES</strong></td>
</tr>
<tr>
<td>In November 2009, the Board adopted a resolution supporting international linking projects for NAEP.</td>
</tr>
<tr>
<td>In March 2010 and May 2010, the Board adopted changes to the NAEP Schedule of Assessments to further support international linking projects.</td>
</tr>
<tr>
<td>Results from the 2011 NAEP/TIMSS Linking Study are scheduled for release in early 2013.</td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
As new initiatives are being considered by the Board addressing achievement gaps and the evolving policy context, the Task Force has provided timely suggestions.

**TOPIC #10: BOARD INITIATIVES ON RAISING ACHIEVEMENT AND CLOSING GAPS**


<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Considerations and suggestions for future initiatives in development</strong></td>
<td>*</td>
</tr>
<tr>
<td>- Involve states instead of reaching schools directly without burdening state budgets.</td>
<td>At the May 2010 Board meeting in Milwaukee, Wisconsin, the Board began discussions to explore how NAEP can enhance the information it provides to the nation. This discussion has been a recurring feature of future Board meetings.</td>
</tr>
<tr>
<td>- Identify stakeholder groups that are already engaging parents in community-based efforts to raise awareness of public education issues.</td>
<td></td>
</tr>
<tr>
<td>- Emphasize comparative international standing and achievement gaps to change expectations.</td>
<td></td>
</tr>
<tr>
<td>- Focus reporting on what students can do at each level, and compare these results to skills required by colleges and careers to make the information meaningful and action-oriented.</td>
<td></td>
</tr>
<tr>
<td>- Personalize NAEP by developing a tool that parents can use to identify the questions they should be asking about student performance, and help parents organize around these tools.</td>
<td></td>
</tr>
<tr>
<td>- Feature sample items to demonstrate test rigor, or make sample tests available.</td>
<td></td>
</tr>
<tr>
<td>- Extract and share lessons about teachers or use of time from the NAEP background questionnaire.</td>
<td></td>
</tr>
<tr>
<td>- Reach out to different sources, such as the Medical College Admissions Test (MCAT) and the military, regarding cutting-edge technologies for data capture that could make Artificial Intelligence (AI) scoring more accessible to states.</td>
<td></td>
</tr>
<tr>
<td>- Foster collaboration between NCES and the consortia on AI scoring.</td>
<td></td>
</tr>
</tbody>
</table>

**Valuable Roles for NAEP in the Common Core Era**

- Continue to serve as a valid external auditing tool to gauge the effectiveness of common and individual state assessments.
- Provide supplementary information with respect to consortia assessments.
- Link NAEP content to various international assessments, in addition to the Trends in International Mathematics and Science Study (TIMSS). State budgets do not allow for states to participate directly in international assessments, and these linking studies provide the opportunity for states to obtain feedback on different types of assessments.
- Serve as a resource to guide policy at the national, state, and local levels, instead of focusing on school implementation activities.
- Use NAEP data other than achievement data in a meaningful way that can inform and shape policy (e.g., richer extraction of NAEP background questionnaire data about student characteristics).  

* Denotes Task Force input for future Board consideration
## Topic #10: Board Initiatives on Raising Achievement and Closing Gaps

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Considerations for the role of NAEP</strong></td>
<td></td>
</tr>
<tr>
<td>- <strong>Stretching the intended role of NAEP.</strong> The Task Force is concerned with the Board’s desire</td>
<td></td>
</tr>
<tr>
<td>to make NAEP more relevant, given the intended purpose of NAEP. Moving away from NAEP’s purpose</td>
<td></td>
</tr>
<tr>
<td>may result in complicated messaging and negative media attention for NAEP.</td>
<td></td>
</tr>
<tr>
<td>- <strong>Describing best practices at the state-level.</strong> Promoting “best practices” for states is</td>
<td></td>
</tr>
<tr>
<td>problematic and may be used against some jurisdictions. States are presented in NAEP</td>
<td></td>
</tr>
<tr>
<td>reporting as homogeneous jurisdictions without sensitivity to differences among states.</td>
<td></td>
</tr>
<tr>
<td>- <strong>Highlighting best practices without influencing curricular decisions.</strong> Legislation</td>
<td></td>
</tr>
<tr>
<td>precludes the use of NAEP to influence curricular decisions; there is a thin line between</td>
<td></td>
</tr>
<tr>
<td>influencing curriculum and sharing best practices.</td>
<td></td>
</tr>
<tr>
<td>- <strong>Policy context shifts affecting NAEP’s role.</strong> The reauthorization of the Elementary and</td>
<td></td>
</tr>
<tr>
<td>Secondary Education Act could change the role of NAEP in the context of Common Core and</td>
<td></td>
</tr>
<tr>
<td>consortia assessments.</td>
<td></td>
</tr>
<tr>
<td>- <strong>Losing impact in the context of over-testing sentiment.</strong> There is a sentiment of over-</td>
<td></td>
</tr>
<tr>
<td>testing, and there are several negative reports on student performance, which may be</td>
<td></td>
</tr>
<tr>
<td>overwhelming for the public.</td>
<td></td>
</tr>
<tr>
<td>- <strong>Identifying NAEP’s relevance to parents.</strong> Research findings indicate that international</td>
<td></td>
</tr>
<tr>
<td>comparisons are not resonating with parents.</td>
<td></td>
</tr>
<tr>
<td><strong>Suggestions for using NAEP data</strong></td>
<td></td>
</tr>
<tr>
<td>- Use background information to contextualize what students are learning, as done with the</td>
<td></td>
</tr>
<tr>
<td>Programme for International Student Assessment (PISA).</td>
<td></td>
</tr>
<tr>
<td>- Use NAEP results to identify high-performing student groups and report the results in terms</td>
<td></td>
</tr>
<tr>
<td>of what is working for groups of students.</td>
<td></td>
</tr>
<tr>
<td>- Triangulate NAEP results with results from other large-scale assessment programs external</td>
<td></td>
</tr>
<tr>
<td>to states (e.g., SAT, ACT, TIMSS, PISA, PIRLS) to answer the questions: What are the data</td>
<td></td>
</tr>
<tr>
<td>telling us? How can we inform expectations about rigor?</td>
<td></td>
</tr>
<tr>
<td><strong>Suggestions for sharing NAEP data with new audiences</strong></td>
<td></td>
</tr>
<tr>
<td>- Enhance pre-service teachers' understanding of NAEP by working with national teacher</td>
<td></td>
</tr>
<tr>
<td>preparation organizations to promote NAEP’s value early in teachers' careers and leverage</td>
<td></td>
</tr>
<tr>
<td>the university research-based perspective.</td>
<td></td>
</tr>
<tr>
<td>- Build researcher capacity to appropriately use, interpret, and report on best practices.</td>
<td></td>
</tr>
<tr>
<td>- Focus the parent initiative on the college preparedness discussion, which is relevant to</td>
<td></td>
</tr>
<tr>
<td>parents and connects to the Common Core State Standards.</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration

The Board developed a list of priority activities and action plans at the December 2011 Board meeting.

The Board convened an expert panel on background questions in November 2011 with a report to the Board in March 2012.
**TOPIC #10: BOARD INITIATIVES ON RAISING ACHIEVEMENT AND CLOSING GAPS**

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suggested improvements for planning the parent initiative</strong></td>
<td>*</td>
</tr>
<tr>
<td>▪ Define the audience more clearly. Consider the purpose of the Board’s initiative in informing parents and goal of raising achievement and closing gaps. Finally, determine what action parents are being asked to take; in other words, what does the Board want parents to do with the NAEP results?</td>
<td>The Board used the Task Force’s input to refine the audience and specificity of its final recommendations for implementing the parent initiative.</td>
</tr>
<tr>
<td>▪ Place NAEP in the larger context of state/consortia assessment for coherent communications about the value/future of NAEP and alignment with state/consortia assessment. Focus on building validity/role of NAEP at policymaker level.</td>
<td>*</td>
</tr>
<tr>
<td>▪ Use sample questions to illustrate NAEP; consider how complementary the items are to state assessments. Leverage mass communications and engage the National School Public Relations Association.</td>
<td>NCES has several focus reports in the development stage. The Board is considering additional topics for focus reports.</td>
</tr>
<tr>
<td><strong>Task Force suggestions for focus report development</strong></td>
<td>*</td>
</tr>
<tr>
<td>In developing focus reports, the Task Force provided the following suggestions to the Board:</td>
<td></td>
</tr>
<tr>
<td>▪ Ensure a direct relationship between the topic and NAEP achievement.</td>
<td></td>
</tr>
<tr>
<td>▪ Ensure objectivity of analysis.</td>
<td></td>
</tr>
<tr>
<td>▪ Consider new background questions that will enrich the focus reports. For example, if there will be a focus report on charter schools, then identify background questions that will be relevant to this topic.</td>
<td></td>
</tr>
<tr>
<td><strong>Task Force priority topics for future focus reports</strong></td>
<td>*</td>
</tr>
<tr>
<td>▪ Charter schools: A 10-year report</td>
<td></td>
</tr>
<tr>
<td>▪ Opportunity-to-Learn</td>
<td></td>
</tr>
<tr>
<td>▪ Education policies and instructional practices of high-performing or high-growth states and districts</td>
<td></td>
</tr>
<tr>
<td>▪ Learning in the South</td>
<td></td>
</tr>
<tr>
<td>▪ Other regional reports</td>
<td></td>
</tr>
<tr>
<td>▪ Eighth-grade algebra (access to algebra rigor)</td>
<td></td>
</tr>
</tbody>
</table>

* Denotes Task Force input for future Board consideration
In January 2012, NCES sponsored a summit for a broad range of SEA staff regarding the NAEP’s future role and potential advances for assessment content and delivery. The summit was a follow-up to an August 2011 NCES summit with technology and innovation leaders. Participants in the January 2012 summit provided the following operational and policy suggestions:

- Lead assessment research and development on new item types and new constructs to inform the field.
- Leverage computer-based assessment to learn about student cognition, e.g., by tracking key strokes and how students use editing tools.
- Maintain NAEP’s role as an external indicator, while establishing links to consortia and international assessments. The Task Force was also asked to provide its input on the future potential roles for NAEP.

### Task Force Discussion and Input

#### New constructs
- Investigate research-based questions and measure new constructs. Students need non-academic constructs to be competitive globally.

#### Reporting relative to state/consortia assessments
- Consider how NAEP can complement Common Core and consortia efforts. If NAEP can provide supporting validity evidence for these efforts, this would be a valuable consensus for the field.
- Link the consortia results if cross-consortia performance levels are not comparable.

#### Engaging higher education
- Use linking of 12th grade NAEP preparedness with SAT, ACT, state assessments, and state longitudinal databases to initiate conversations with higher education policymakers.

#### Career readiness
- Address NAEP preparedness research and reporting to development of a career readiness standard. Currently, the NAEP preparedness initiative emphasizes SAT and ACT linking research, but these measures only address college readiness.

#### Possible ways to use NAEP in state accountability
- Offer a secure set of NAEP items for consortia and/or state standard-setting. Mapping consortia cut scores onto the NAEP scale may also be helpful in this regard. Careful implementation of these suggestions will be needed to guard against criticism regarding federal intrusion.

* Denotes Task Force input for future Board consideration
Ethics Division Contact List

Phone: 202-401-8309
Fax: 202-260-5104

Susan Winchell, Assistant General Counsel
Susan.Winchell@ed.gov

Marcella Goodridge, Deputy Assistant General Counsel
Marcella.Keiller@ed.gov

Mailing Address
U.S. Department of Education
Office of the General Counsel Ethics Division
400 Maryland Avenue, SW Suite 6E225
Washington, D.C. 20202-2110

General e-mail or questions: OGC.Ethics@ed.gov
ETHICS PRIMER

FOR

THE NATIONAL ASSESSMENT GOVERNING BOARD

November 2009
Ethics Division
Office of the General Counsel
U.S. Department of Education
### Contents

**EXECUTIVE SUMMARY** ........................................................................................................................................... 3

**ETHICS LAWS AND RULES APPLICABLE TO SGES** ............................................................................................................................ 6

**I. INTRODUCTION** .................................................................................................................................................. 6

**II. YOUR STATUS AS A SPECIAL GOVERNMENT EMPLOYEE** .......................................................................................... 6

   A. What is a “special Government employee”? ........................................................................................................ 6
   B. Do the ethics restrictions apply when I am not working for NAGB? ........................................................................... 6

**III. CONFLICTS OF INTEREST** ........................................................................................................................................ 6

   A. What criminal conflict of interest statutes apply to SGEs? .................................................................................. 6
   B. What financial conflicts of interest may arise for SGEs under section 208? .......................................................... 7
   C. How do I resolve a conflict of interest? ................................................................................................................ 9
   D. What restrictions apply to my representation of third parties under sections 203 and 205? ....................................... 10

**IV. POST-EMPLOYMENT** ........................................................................................................................................ 10

**V. STANDARDS OF ETHICAL CONDUCT AND OTHER ETHICS RULES** ............................................................................. 11

   A. What restrictions apply if I want to engage in fundraising? ........................................................................ 11
   B. What restrictions are there on my acceptance of gifts? .......................................................................................... 11
   C. What restrictions apply if I want to “lobby” Congress? .................................................................................... 12
   D. What does “misuse of position” mean? ................................................................................................................ 12
   E. May I keep my day job and still serve on NAGB? .............................................................................................. 12
   F. Are there any restrictions on my political activities? .......................................................................................... 13
   G. What restrictions do I face if I want to teach, speak, or write on matters that are related to the duties I perform for NAGB? .................................................................................................................. 13
   H. What restrictions apply if my government duties involve the awarding of contracts? ........................................... 14
   I. What restrictions apply to my interaction with foreign entities? ........................................................................... 14
   J. What do I do if I am called to be an expert witness? ............................................................................................ 15
   K. May I keep and use frequent flyer miles that I earn when I am on official NAGB travel? ................................. 15

**VI. CONCLUSION** .................................................................................................................................................. 15
EXECUTIVE SUMMARY

Now that you are a member of the National Assessment Governing Board (“NAGB”) you need to know what ethics laws and rules apply to you. The following is a very brief summary of these rules. For a more detailed discussion of how these rules apply to you, please refer to the attached summary entitled “Ethics Laws and Rules Applicable to SGEs.”

Your Status as a Special Government Employee

You are considered an SGE and not a regular federal employee because NAGB anticipates that you will be serving the federal government through your position for only 130 days or less during any period of 365 consecutive days. Whether or not you are paid by the Board for your service is irrelevant. This summary discusses how the ethics rules apply to SGEs.

Criminal Statutes Apply to Your Activities

Some of the ethics laws that apply to you carry criminal penalties. Below is a brief summary of the most important of these laws.

- The chief conflict of interest law bars you from participating personally and substantially in your capacity as a member of NAGB in any particular matter before the federal government that has a direct and predictable effect on your own financial interests or the financial interests of others with whom you have certain relationships. See 18 U.S.C. Section 208.

- If you find yourself with a financial conflict of interest, you have four options: (1) disqualify yourself (you don’t participate in any way in the matter); (2) resign from the outside entity that is the basis for the conflict; (3) sell or divest the stock or other financial interest that is the basis for the conflict; or (4) request and obtain a statutory waiver.¹

- Two other laws prohibit you from representing a third party, with or without compensation, before any court or agency in connection with any particular matter involving specific parties in which the United States is a party or has a direct and substantial interest and in which you have participated personally and substantially as an SGE. In addition, if you serve the federal government for more than 60 days during the immediately preceding period of 365 consecutive days, these restrictions apply to any matter that is pending with NAGB. But remember that these restrictions do not apply to particular matters of general applicability, such as broadly applicable policies, rulemaking proceedings or legislation, that do not involve specific parties. See 18 U.S.C. Sections 203 and 205.

¹ In rare circumstances, with the concurrence of the U.S. Office of Government Ethics, you may obtain a waiver of the conflict of interest.
• Another criminal law limits some of your activities after your service on NAGB ends. This law prohibits you from representing others in connection with the same particular matter involving specific parties in which you participated personally and substantially during your service to NAGB. This prohibition lasts for your lifetime. See 18 U.S.C. Section 207.

Standards of Ethical Conduct for Employees of the Executive Branch

The Standards of Ethical Conduct for Employees of the Executive Branch (Standards), 5 C.F.R. Part 2635, are regulations that apply both to regular federal government employees and to SGEs. However, a few exceptions exist in the Standards in recognition of the fact that SGEs are working for the government only in a very limited way. A brief synopsis of some of these rules and their exceptions follow.

• **Fundraising:** You may not use your official title, position and authority to engage in fundraising.

• **Gifts:** You may not accept gifts from a “prohibited source” or offered to you because of your official position on NAGB. A prohibited source includes any person: seeking official action from NAGB; doing or seeking to do business with NAGB; conducting activities regulated by NAGB; or having interests that may be substantially affected by your official duties. There are many exceptions to this rule that are discussed in more detail in the accompanying memorandum.

• **Lobbying:** In your role as a member of NAGB, you may not urge others to contact Congress or a state legislature to urge the passage or defeat of legislation. Additional restrictions exist regarding lobbying. You should contact Department of Education’s Ethics Division before engaging in any type of lobbying.

• **Misuse of Position:** You may not use your position on NAGB or nonpublic information gained through your service on NAGB to seek advantage for yourself or others. In addition, you may not use your NAGB title in a manner that makes it appear that NAGB is sanctioning your views, products, services or personal enterprises.

• **Political Activities:** You may not engage in political activity when you are on duty or in a federal government building or car, and you may never use your official title as a member of NAGB in connection with political activities.

• **Teaching, Speaking and Writing:** You may not receive compensation for teaching, speaking or writing if: (1) the invitation was offered to you because of your position on NAGB; (2) the information conveyed by you draws substantially on nonpublic information that you obtained by working on NAGB; (3) the invitation was extended to you by an organization or person who has interests that may be substantially affected by
your performance on NAGB; or (4) the subject of your work deals in a significant way with a matter involving specific parties that you worked on while on NAGB. Again, there are some exceptions to this rule that are outlined in more detail in the accompanying memorandum.

Required Filing of a Financial Disclosure Report By SGEs

As a member of the NAGB, you are required to file a confidential financial disclosure report (also referred to as a “450” Report) when you are first appointed, and annually thereafter if you are reappointed. The purpose of the financial disclosure form is to protect you from inadvertently violating any of the criminal conflict of interest statutes and so that NAGB can know that your advice is free from any real or perceived conflicts of interest.

Please do not rely solely on this “Executive Summary” before undertaking your duties. There are many subtle nuances that are not discussed in this summary that may apply to your specific situation. The attached expanded summary provides additional detail that will help you better understand the ethics rules. Please feel free to call or e-mail Marcella Goodridge in the Ethics Division of the Office of the General Counsel at the U.S. Department of Education at (202) 401-8309, or Marcella.Keiller@ed.gov, for answers to any specific ethics questions that may arise in the course of your service on NAGB.
ETHICS LAWS AND RULES APPLICABLE TO SGES

I. INTRODUCTION

Although the ethics rules are numerous and detailed, a single, simple principle underlies these rules: You should never use your public office for private gain, either for yourself, or for any third party. In addition, you must refrain not only from engaging in any activity that violates the ethics rules, but you must also refrain from any activity that creates the appearance of a violation of any of these rules. The summary below is designed to help you avoid violating any ethics rules covering your activities as a member of NAGB.

II. YOUR STATUS AS A SPECIAL GOVERNMENT EMPLOYEE

A. What is a “special Government employee”?

Because you have been appointed to be a member of the NAGB and you are expected to perform your duties for not more than 130 days during the 365 days subsequent to the date of your appointment, you are, by law, a “special Government employee” (SGE). As an SGE, you are a federal government employee. This means that upon appointment, you assume the responsibilities, obligations, and restrictions that are part of public service. Because SGEs are not full-time employees, several of these restrictions apply only in limited circumstances.

B. Do the ethics restrictions apply when I am not working for NAGB?

Yes, any restrictions concerning your private activities (representational services, expert witness activities, etc.) apply equally on days when you serve the federal government through your position on NAGB and on days when you do not, except with respect to political activity. If you have not provided any services for the federal government for some time, but have not received a termination date for your appointment, you must seek a formal resolution of the matter before engaging in conduct prohibited by the ethics rules.

III. CONFLICTS OF INTEREST

A. What criminal conflict of interest statutes apply to SGES?

While you are employed as an SGE, you need to pay particular attention to four criminal conflict of interest laws found in Chapter 11, Title 18 of the United States Code: 18 U.S.C. Sections 203, 205, 207 and 208. These criminal laws include some special provisions for the treatment of SGEs. A discussion of these laws and certain related requirements found in other laws and regulations follows.
B. What financial conflicts of interest may arise for SGEs under section 208?

Section 208 prohibits you from participating personally and substantially in any particular matter that has a direct and predictable effect on your financial interests, including certain interests of others that are imputed to you under the statute. This means that you may not work on NAGB matters if you have certain connections – through the ownership of stock, through employment, or by virtue of other circumstances – with an organization that has a financial interest in the matter. For example, you may not work at all on a contract competition if you own stock valued at a certain amount in a company competing for the contract. You may not participate in a discussion of whether to modify an existing contract with a company if you work for that company. And, you may not assist in the development of a scope of work for a contract competition if you know that an organization on which you serve on the Board of Directors plans to compete for that contract.

In addition to your own personal financial interests, the financial interests of the following persons or organizations are imputed to you and also disqualify you from participating in a particular matter:

1. your spouse;
2. your minor child;
3. your general partner;
4. an organization for which you serve as an officer, director, trustee, general partner or employee; and
5. any prospective employer.

Example 1 You are on the governing board of ABC, a nonprofit organization. ABC’s financial interests are imputed to you under the statute. This means that for the purpose of determining whether you have a conflict of interest, ABC’s financial interests are treated as if they were your own. Accordingly, you may not participate in any NAGB matter in which ABC has a financial interest. Similarly, if you were in the process of discussing employment with ABC, you would be barred from participating in any NAGB matter affecting the financial interests of ABC.

Example 2 You are on the governing board of ABC (or employed by ABC, own stock in ABC, seeking employment with ABC, etc). You are asked to participate in the process of reviewing and scoring contract proposals for a contract competition for a NAGB project. Fifteen organizations have submitted a bid. When you open the proposal from one organization, you...
note that ABC’s name is one of the organizations that has submitted a bid. Or, perhaps ABC is listed as a subcontractor in one of the proposals. This contract competition is a “particular matter” that will have a “direct and predictable effect” upon the financial interests of ABC. In other words, as a result of the contract competition, ABC will either gain business or not, and this decision will affect ABC financially – either negatively or positively. The amount of financial interest is not relevant – as long as ABC’s finances will be affected, unless a regulatory exemption or waiver permits you to do so, you may not work on this competition. And, because each proposal is competing against all of the others, your evaluation of competing proposals will affect the chances ABC has of winning the contract. Accordingly, you may not review any of the proposals.

You must recuse yourself from a matter as soon as you realize that you have a conflict. If, for example, you notice that you have a conflict when you are in the middle of reviewing contract proposals, you put the proposal back in its envelope and call up an NAGB staff member and let that person know that you think that you are disqualified from working on the competition. If there is any question, you should contact the U.S. Department of Education Office of the General Counsel’s Ethics Division for guidance. Once you have determined that you may not work on this matter, send the proposal back to NAGB staff.

You are permitted to participate in a particular matter affecting one campus of a multi-campus institution of higher education, where the disqualifying interest arises from your employment with a separate campus of the same institution, provided that you have no multi-campus responsibilities at the institution. If you are employed with a large university with multiple campuses and you do not have any multi-campus responsibilities, you may participate in official matters—such as grants, contracts, applications, and other particular matters—that affect the financial interests of another campus in the same university system where you are employed. Below are some examples of how section 208 may apply to your activities.

**Example 3** You are employed as a professor at the University of California-Berkeley. NAGB is planning to evaluate the impact of computer-based testing on students with disabilities and English language learners. UC-Berkeley’s science and technology department has submitted a bid. NAGB’s actions will have a direct and predictable effect on the university’s financial interest. Therefore, you may not participate in any way on this matter.

**Example 4** You are employed as a researcher at the University of California-Berkeley. NAGB is planning to evaluate the impact of computer-based testing on students with disabilities and English language learners. The University of California-Los Angeles (UCLA) has submitted a bid to be the contractor for NAGB’s evaluation. You may participate in this matter because it will not have a direct and predictable effect on either your financial interests or UC-Berkeley’s.
C. How do I resolve a conflict of interest?

1. Disqualification

A common method of resolving a conflict of interest is to disqualify yourself from participating in the matter.

Example 5 You are serving on NAGB’s Ad Hoc Committee that will examine issues related to computer-based testing for students with disabilities and English language learners, including developing a study of computer-based testing methodologies. The Request for Proposals has been disseminated. One of the bids submitted is from ABC Corporation (ABC). You own $20,000 worth of stock in ABC. You must advise the U.S. Department of Education Office of the General Counsel’s Ethics Division that you own stock in ABC and you will not be able to participate in any way in the entire contract competition. If ABC is awarded the contract, you will also need to disqualify yourself from the entire matter.

2. Divestiture

Divestiture of a disqualifying interest (usually through the sale of stock) is another remedy available to avoid a potential violation of section 208. SGEs are not eligible for a Certificate of Divestiture (CD). A CD is a tax benefit that allows the deferral or nonrecognition of capital gain where an employee divests a financial interest in order to comply with conflict of interest requirements. Unfortunately, Congress specifically excluded SGEs from eligibility to receive CDs. 26 U.S.C. § 1043(b)(1)(A).

3. Resignation

On some very rare occasions when none of the aforementioned options are available or feasible, an SGE may need to resign from participating in an outside activity with an entity if his or her official activities as an SGE have a direct and predictable effect on the financial interest of that entity creating an irreconcilable conflict.

4. Waiver or Authorization

Another remedy to avoid a conflicting financial interest is to request and obtain a statutory waiver by contacting the Department of Education’s Ethics Division (an authorization is similar to a waiver, but only applies to non-statutory conflicts of interest - what are often referred to as “appearances of a conflict”). You may be granted a waiver only if your financial interest is not so substantial as to be deemed to be likely to affect the integrity of your services.
**Example 6** In the scenario described in Examples 1 and 2 above, you are granted a waiver permitting you to participate in a general policy matter that affects ABC’s financial interests as long as the matter affects all similarly situated entities in the same manner. But you would remain disqualified from participating in a matter that specifically involves ABC, which in this case means the entire contract competition.

**D. What restrictions apply to my representation of third parties under sections 203 and 205?**

With regard to particular matters in which you have participated personally and substantially while serving NAGB, you are prohibited from representing a third party on those particular matters, with or without compensation, before any court or agency, when the United States is a party or has a direct and substantial interest in the matter. See 18 U.S.C. Sections 203 and 205.

In addition, if you serve the federal government for more than 60 days during the immediately preceding period of 365 consecutive days, you are prohibited from representing a third party on any matter involving specific parties pending before NAGB, even if your work at NAGB did not involve these matters. These restrictions do not apply to particular matters of general applicability, such as broadly applicable policies, rulemaking procedures or legislation that does not involve specific parties.

**IV. POST-EMPLOYMENT**

After your appointment terminates at NAGB, you need to pay particular attention to one more criminal statute that subjects you to restrictions regarding certain matters that you may have worked on as a member of NAGB. Pursuant to 18 U.S.C. Section 207, you may never represent any third party, other than in the performance of your official government duties, in connection with the same particular matter involving specific parties in which you participated personally and substantially as a member of NAGB. This is a lifetime prohibition. For example, if you participated in a NAGB discussion concerning a contract to State University, you may never represent State University with respect to that same contract before any official of the Executive Branch of the federal government and you may never represent State University with respect to that contract in any federal court.

Further, if you serve on NAGB more than sixty days and are compensated above a certain level, you may be subject to a one-year “cooling-off” period during which you would be barred from representing before NAGB certain third parties in connection with any matter. There are some exceptions to this law as well, and you should contact the Department of Education’s Ethics Division for guidance.
V. STANDARDS OF ETHICAL CONDUCT AND OTHER ETHICS RULES

The Standards of Ethical Conduct for Employees of the Executive Branch (Standards), 5 C.F.R. Part 2635, are regulations that apply both to regular federal government employees and to SGEs. Although you are treated generally the same as regular employees under the Standards, a few exceptions do exist for SGEs in recognition of the fact that SGEs are working for the government only in a very limited way. In addition, there are other rules that govern your conduct as an SGE, including the Hatch Act, anti-lobbying rules, the Federal Acquisition Regulation, and rules about accepting gifts and compensation from foreign governments. A brief synopsis of some of these rules follows.

A. What restrictions apply if I want to engage in fundraising?

You may not use your NAGB title, position or authority to solicit funds for any organization. In addition, you may not personally solicit funds or other support from persons whose interests may be affected substantially by the performance or nonperformance of your official duties.

B. What restrictions are there on my acceptance of gifts?

You are prohibited from accepting gifts (almost anything of monetary value) from a “prohibited source” or gifts given because of your official position as a member of NAGB, unless a specific exception applies. The definition of “prohibited source” includes any person:

- seeking official action from NAGB;
- doing or seeking to do business with NAGB; or
- having interests that may be substantially affected by your official duties at NAGB.

The definition also includes organizations the majority of whose members fall within any of these categories. You may accept various benefits resulting from your outside business or employment activities, if a reasonable person would conclude that such benefits are not offered or enhanced because of your official position. The most commonly applicable exceptions to the gift rule allow you to accept:

- Modest items of food other than a meal, such as coffee, soft drinks, or donuts;
- Most plaques, certificates and trophies;
- Discounts available to all Government employees;
- Anything for which you pay market value;
- Gifts valued at $20 or less per occasion, totaling no more than $50 in a calendar year from any one source;
- Gifts clearly motivated by friendship or family relationship;
- Gifts resulting from your outside business activities, including those of your spouse; and
- Free attendance or meal which is provided by:
1. the sponsor of the event for the day on which you are speaking at the event, or for a widely-attended gathering of mutual interest to a number of parties when the necessary determination of agency interest has been made; or

2. someone other than the sponsor of a widely-attended gathering of mutual interest to a number of parties when more than 100 people are expected to attend, the aggregate value of the gift is under $335, and the necessary determination of agency interest has been made.

C. What restrictions apply if I want to “lobby” Congress?

NAGB and its members are permitted to communicate directly with Congress in their official capacity on matters that are related to legislation or appropriations deemed necessary to conduct NAGB’s “public business” (i.e., the NAGB’s statutory functions and responsibilities). However, the Anti-Lobbying Act, 18 U.S.C. Section 1913, prohibits you, in your official capacity at NAGB, from engaging in “grass-roots lobbying” (i.e., directly or indirectly suggesting or requesting that others contact Congress or a state legislature to urge the passage or defeat of proposed or pending legislation), even if it is related to the NAGB’s public business. The Anti-Lobbying Act also requires that any permissible direct communications with Congress in your official capacity at NAGB be made only through official channels.

None of these restrictions prohibit you from lobbying members of Congress or state legislatures, or urging others to do so, on your own time in your personal capacity. If you lobby Congress or state legislatures in your personal capacity, and the issue is related to NAGB’s business, you should make it clear that you are not representing NAGB and not acting in your official capacity as a member. Also, please note that when you are lobbying as a private citizen, you are not permitted to use government resources or equipment (including, but not limited to, computers, telephones, fax machines, copy machines, stationery), or seek assistance from NAGB staff.

D. What does “misuse of position” mean?

You may not use your position on NAGB to seek advantage for yourself or others. You also may not use nonpublic information gained through your service at NAGB to seek advantage for yourself or others. Finally, you may not use your NAGB title in a manner that makes it appear that the NAGB is sanctioning your views, products, services or personal enterprises. Of course, you may list your membership on NAGB on your curriculum vitae, but you may never use your status as an NAGB member to advertise or promote your personal activities. Please seek advice from the Department of Education Office of the General Counsel’s Ethics Division if you have any questions in this area.

E. May I keep my day job and still serve on NAGB?

Yes, you may continue to collect your regular salary from an outside employer for days on which
you are providing services to the federal government (whether your federal government service is paid or unpaid). However, if you have another consultant or advisory position with NAGB or any other federal department or agency, you may not receive per diem or salary from NAGB for the same day for services performed for the two positions.

F. Are there any restrictions on my political activities?

You may not engage in any political activities while you are on duty (i.e., performing government services) or when you are in a government building or vehicle. Although you are not subject to any restrictions on your political activities when you are not performing government services, you may never use your official title as a member of NAGB in connection with any political activities.

G. What restrictions do I face if I want to teach, speak, or write on matters that are related to the duties I perform for NAGB?

You may not receive compensation for teaching, speaking, or writing if:

- the activity is performed as part of your official duties (e.g., a speech on behalf of NAGB);

- the invitation to engage in the activity was extended primarily because of your official position at NAGB, rather than expertise in the subject matter;

- the invitation or offer of compensation was extended to you by someone with interests that may be affected substantially by your duties;

- the information conveyed through the activity draws substantially on nonpublic information obtained through your service at NAGB; or

- the activity deals, in significant part, with a matter involving specific parties to which you are currently assigned or had been assigned during your current NAGB appointment.

Notwithstanding the restrictions in bold type you may accept compensation for teaching a course requiring multiple presentations offered as part of: (a) the regularly established curriculum of various specified types of educational institutions; or (b) educational or training programs sponsored and funded by federal, State, or local government. However, if you teach at an educational institution, you must not participate in any NAGB matters that involve that institution.
H. What restrictions apply if my government duties involve the awarding of contracts?

If you are involved in the awarding of any contracts, please seek advice from the Ethics Division. There are special provisions that cover your involvement in the awarding of contracts. For example, you may not accept compensation as an employee, officer, director, or consultant of a contractor within the one-year period after leaving Government service where you participated in certain procurement matters pertaining to that contractor. In addition, if you disclose certain information pertaining to Federal procurements that you obtained during your service on a committee, you may face sanctions, including criminal penalties.

I. What restrictions apply to my interaction with foreign entities?

The emoluments clause of the U.S. Constitution prohibits you from receiving any emolument, office or title of any kind from a foreign government, including political subdivisions of a foreign government. An emolument is compensation received by virtue of holding an office or having employment with a foreign government and includes, for example, salary, honoraria, transportation, per diem allowances, household goods, shipment costs, and housing allowances. This clause has been interpreted to be broader than the traditional notion of employment and includes, for example, income received through a partnership when an identifiable portion of the partnership draw can be attributed to the partnership’s fees from such foreign government. This provision has particular relevance to positions with foreign universities that are government-operated, as opposed to private institutions. United States Constitution, art. I § 9, cl. 8. There are also statutory provisions restricting acceptance of gifts from foreign governments. 5 U.S.C. § 7342. You should seek advice from the Ethics Division regarding the details about these restrictions. Additionally, a criminal statute bars employment or consultation with a foreign entity for the purpose of providing foreign agent representation or lobbying. 18 U.S.C. § 219.

The ban on participating in foreign agent activities covered by the Foreign Agents Registration Act (FARA) prohibits representation of foreign governments or foreign political parties before the United States Government, as well as a number of other activities conducted within the United States on behalf of such entities. There are certain FARA exceptions related to trade or commerce, legal representation, humanitarian fundraising, and religious, scholastic, or scientific pursuits. The Lobbying Disclosure Act of 1995 requires certain covered Federal officials who serve as agents of foreign principals (other than foreign governments or foreign political parties) to register if they work on behalf of foreign corporations, associations, or other organizations.

Finally, certain restrictions apply after your position with NAGB terminates. Specifically, 18 U.S.C. § 207 includes restrictions on former employees who participated in trade or treaty negotiations on behalf of the United States (18 U.S.C. § 207(b)) and on former senior employees who wish to represent, or aid or advise in the representation of, a foreign entity with the intent to influence a decision of a Federal employee or agency (18 U.S.C. § 207(f)).
J. What do I do if I am called to be an expert witness?

Government employees generally may not participate as an expert witness, with or without compensation, other than on behalf of the United States, in any proceeding before a federal court or agency in which the United States is a party or has a direct and substantial interest. This restriction applies to most SGEs only if the SGE actually participated officially in the same proceeding or in the particular matter that is the subject of the proceeding. If you are appointed by the President, serve on a commission established by statute, or serve (or are expected to serve) for more than 60 days in a period of 365 days, the restriction on expert service also applies to any proceeding in which NAGB is a party or has a direct and substantial interest.

K. May I keep and use frequent flyer miles that I earn when I am on official NAGB travel?

Yes, you may use frequent flyer miles or other airline awards or promotions accumulated on official NAGB travel for your own personal use.

VI. CONCLUSION

We understand that these laws are complex and may not be intuitive. Again, we caution you that this summary is merely an introduction to the ethics laws and rules that apply to you. You should always feel free to contact the Department of Education Office of the General Counsel’s Ethics Division with any questions or concerns.

Marcella Goodridge Keiller, Attorney
U.S. Department of Education
Office of the General Counsel
400 Maryland Avenue, S.W., Room 6E237
Washington, D.C. 20202-2110
(202) 401-8309
(202) 260-5104 (fax)

Marcella.Keiller@ed.gov
Exploratory Analysis of NAEP Data
Prepared for the National Assessment Governing Board

Who Attends Charter Schools and How Are Those Students Doing?

December 2012

By Naomi Chudowsky and Alan Ginsburg
ABSTRACT

This report examines what the National Assessment of Educational Progress (NAEP) can tell us about charter school enrollment and student performance compared to that of regular public schools. The study uses NAEP reading and mathematics data from 2011 and the earlier years when charter school data first became available (2003 for grade 4; 2005 for grade 8). The study finds that while charter schools still account for a relatively small percentage of overall public school enrollment (about 3%), this percentage is growing rapidly, particularly in large cities. A sizable jump was found in charter school enrollment for the black student subgroup.

In terms of achievement, there is a consistent pattern of higher average NAEP scores for regular public schools than for charters when we look at the nation as a whole. However, the closer we focus in on large cities, where most charter schools are located, the more the picture changes in favor of charter schools. In all large cities combined, student achievement is roughly even overall, but the black and Hispanic subgroups show higher scores in charter schools. When we examined four urban areas specifically (DC, Atlanta, Chicago, Milwaukee), students in charter schools significantly outperformed their peers in regular public schools in many of the subjects/grades analyzed. Charter school class sizes are smaller, and there is some evidence that charter school students receive more instructional time in some subjects. The report recommends making some changes to the NAEP Data Explorer interface and including data for research purposes in the Trial Urban District Assessment (TUDA) database that is representative of all charter schools located within each participating urban district.
EXECUTIVE SUMMARY

Charter schools are independent public schools that operate with a large measure of autonomy from local districts. Students attend charter schools by choice. The schools themselves are exempt from many state and local rules but are subject to accountability standards established in their charters.

There has been considerable controversy over charter schools and whether they foster higher student achievement. In 2005, the National Center for Education Statistics (NCES) issued a study of America’s charter schools based on the 2003 NAEP reading and mathematics assessments for 4th grade.¹ The study found that students in charter schools were generally lagging behind their counterparts in regular public schools. But the differences were statistically significant only for math, not for reading. Given the growth of charter schools since 2003, it is important to take a fresh look at what NAEP can tell us about charter school enrollment and student performance, based on 2011 data.

This report examines five questions using NAEP data:

- Who attends charter schools? How has charter school enrollment changed over the past decade?

- Does student achievement differ between charter schools and regular public school students? Has that changed since 2003?

- How do certain subgroups, such as low-income, black, and Hispanic students, perform in charter schools compared to regular schools?

- How does the performance of charter schools compare to that of regular public schools located within the same city?

- Are there important differences between charter and regular schools that are captured by the NAEP background data?

NAEP collects a rich set of student, teacher, and school responses to background questionnaires (referred to as “background variables”) that can help in understanding the context for NAEP achievement results and give insights into how to improve them. We can use NAEP to address the charter school questions above because one of the background variables that has been collected since 2003 from each school in the NAEP sample is whether it is a charter or regular public school.

¹ For the 2003 charter school study, NAEP oversampled charter schools at grade 4 to make sure that the sample of charter school students was large enough to conduct the necessary analyses. In 2005, NAEP started assessing larger numbers of students and from that time forward, the NAEP sample included a representative sample of charter schools without having to oversample.
NAEP background questions are a potentially important but largely underused national resource (Smith et al., 2012). This is one of several exploratory analyses, commissioned by the National Assessment Governing Board (NAGB) to illustrate the usefulness of NAEP background data for helping to understand important education issues. However, there are limitations to this study. One cannot conclude from the results the causal effects of attending a charter school on student achievement. The study provides descriptive and correlational evidence and by itself does not prove causation.

Who Attends Charter Schools

The National Center for Education Statistics (NCES, 2010) reports that from 2000 to 2010,\(^2\) the number of students enrolled in public charter schools more than quadrupled--from 0.3 million to 1.6 million students. During this period, the percentage of all public schools that were charter schools increased from 2 to 5 percent, numbering 5,000 schools in 2010. At that time, about 55% of charter schools were located in cities, 21% were in suburban areas, 8% were in towns, and 16% were in rural areas. In some cities, charter schools serve a large proportion of students. For instance, in New Orleans more than half of all public school students attend charter schools, and in the District of Columbia, more than a third.

Questions have been raised about whether charter schools are appropriately serving students with disabilities. In response, the U.S. Government Accountability Office (GAO, 2012) conducted a study using federal data and found that charter schools enrolled a lower percentage of students with disabilities than traditional public schools. In school year 2009-2010, they found that approximately 11% of students enrolled in traditional public schools were students with disabilities compared to about 8% of students enrolled in charter schools.

This study, using NAEP 2011 data, finds that charter schools are playing an increasingly important role in American education, particularly in large cities and for certain subgroups. Charter school data go back to 2003 in grade 4; 2005 in grade 8; and are only available for 2009 in grade 12.

The main findings related to charter school enrollment are:

- Nationally, students attending charter schools account for a small slice of overall public school enrollment: about 3% in 2011 at grades 4 and 8. Still, this represents a significant increase, compared to 1% in 2003/05.

- Higher percentages of students living in large cities attend charter schools. For example, in grade 4, charter schools enrolled 3% of all large city students in 2003 and this grew to 6% in 2011.

- While all of the subgroups analyzed showed significant increases in charter school enrollment since 2003, the most notable jump was for the black subgroup. For example,

\(^2\) 2010 is the most current year for which NCES charter school enrollment data are available.
in the large cities, the percentage of grade 4 black students attending charter schools grew from 4% in 2003 to 12% in 2011. The percentage of black low-income students attending charter schools in the large cities is roughly similar to that of black students in general.

- There are also differences in the student compositions of charter and regular public schools. Both nationally and in the large cities, charter student bodies include a significantly larger proportion of black students than regular public schools. Regular schools have significantly larger proportions of white students.

- In grade 4, a larger proportion of the charter school student body was low-income in 2011, compared to regular schools. A similar pattern can be seen at grade 8, but the difference between charter and regular schools is not significant.

- In the large cities, in grade 4, regular schools had a significantly larger percentage of Hispanic students in their student bodies in 2011. In grade 8 the same was true in 2005, but by 2011, the Hispanic composition of the two types of schools became more even.

- Also in the large cities in 2011, regular schools had significantly higher rates of enrollment for students with disabilities at grade 8, and English language learners at grade 4.

Achievement of Charter School Students

An earlier NCES study (2005) comparing student achievement in charter and non-charter public schools using 2003 NAEP data, found that charter school performance lagged behind that of regular schools in 4th grade math, and was about the same in 4th grade reading. That study noted that when comparing the performance of charter and regular public school students, it is important to compare students who share a common characteristic. For example, in math, 4th grade charter school students as a whole did not perform as well as their public school counterparts. However, the math performance of white, black, and Hispanic 4th graders in charter schools was not measurably different from the performance of 4th graders with similar racial/ethnic backgrounds in other public schools.

More recently, Betts and Tang (2011) conducted a meta-analysis of several rigorous studies that used either experimental (lottery) or student-level growth based methods to infer the causal impact of attending a charter school on student performance. From this set of studies, Betts and Tang concluded that charter elementary schools, on average, outperformed traditional public schools in reading and in math, and charter middle schools were doing better in math. Their analysis found no substantial difference in math or reading achievement between traditional and charter high schools or in reading achievement at the middle school level.

The main findings from our analysis of NAEP achievement data are:
- At the national level, there is a consistent pattern of higher average NAEP scores for regular public schools than for charter schools. This pattern is apparent in all grade/subjects analyzed: grades 4, 8 and 12 in reading, math, and science.

- NAEP scores in grades 4 and 8 reading and math have increased between 2003/05 and 2011, in both regular public and charter schools, with larger gains for charter schools. The gains for regular public schools tend to be statistically significant, while a similar amount of growth for charter schools does not, probably because of the small charter school sample size.

- Focusing on large cities, average NAEP scores for charter and regular schools were similar, both in 2003/05 and 2011. The only significant difference was in grade 4 math in 2003, in favor of regular public schools. By 2011 this difference had disappeared.

- However, the findings tend to favor charter schools when one focuses on black, Hispanic, and low-income students within the large cities. In many subject/grade combinations students in these subgroups in charter schools performed significantly better in 2011 than those in regular public schools. By contrast, in 2003/2005 these subgroups performed similarly in charter and regular schools, and in one case (low-income students in grade 4 math), the regular schools were ahead.

- The performance of black low-income students attending charter schools in large cities is particularly striking. This group has shown a large increase in scores. In 2011 their achievement was significantly higher than that of similar students in regular large-city schools in grade 8 reading and grades 4 and 8 math.

- In the large cities, the only significant subgroup findings in favor of regular schools in 2011 were for Asians (in grade 4 math) and whites (in grade 4 reading).

- When we look more closely at a few large urban districts, the 2011 results clearly favor charter schools. In the four cities where NAEP data permitted comparisons (DC, Atlanta, Chicago and Milwaukee), students in charter schools significantly outperformed their peers in regular public schools in many of the subjects/grades analyzed. In those four districts, there are no subjects/grades where regular schools significantly outperformed charter schools.

Why do we find that charter schools underperform regular public schools nationally, but outperform them in large cities? One possible explanation is that at the national level, NAEP results for regular public schools include a wide range of schools, including many high-performing suburban schools in high-income communities. If the purpose is to compare charter and regular schools, the fairest way is to compare them within similar locations. The location where charter school enrollment is most concentrated is the large cities. As noted above, 6% of large city grade 4 students and 8% of large city grade 8 students attend charter schools, compared to 3% at each grade nationally. Making comparisons within large cities increases the likelihood that the regular and charter schools are serving similar populations of students.
**Other Factors Related to Instruction**

There are hundreds of variables collected from students, teachers, and school administrators each time NAEP is administered and very little of those data get used by researchers. For this exploratory study, we compared charter and regular public schools on a few policy-relevant factors (time spent on core subjects, class size, and teacher qualifications), to give a flavor of the kinds of information that are potentially available. These analyses were run for grades 4 and 8 in large cities.

The main findings are that in the large cities:

- Grade 8 charter school students get significantly more time per week of language arts instruction. Grade 8 charter school students may also be getting more instruction in math, although the finding is not significant.
- Class sizes tend to be larger in regular schools than in charter schools.
- Significantly more teachers in regular public schools have a major in the subject they teach.
- More grade 8 math teachers in charter schools entered the teaching profession through an alternate certification program.
- Teachers in regular schools have more years of teaching experience than teachers in charter schools.

**Strengthening NAEP Data and the Data Explorer**

The main way that members of the public and education researchers access NAEP data is through the online *NAEP Data Explorer*. It is a powerful tool and, with practice, researchers can conduct a wide variety of analyses. However, the interface is not as user-friendly as it could be. One of the purposes of this exploratory study was to identify ways that the NAEP background data could be improved. We offer these suggestions based on our experience:

- Refine the online Data Explorer to make it as intuitive and efficient possible.
- Find a better way to search the variables and filter those available in a given year.
- Take stock of background variables that may no longer be relevant and consider reinstating some that are still of interest.

Charter schools are likely to continue to grow in importance. They surely will be the focus of future research. These schools are serving ever-larger numbers and percentages of students—
particularly minority and low-income students. The large urban districts are where studies of charter schools should be focused. So finally we suggest:

- Find a way to include results for all charter schools within the geographic boundaries of an urban district in the TUDA dataset, for research purposes.
Who Attends Charter Schools and How Are Those Students Doing?

DATA AND FINDINGS

I. Who Attends Charter Schools?

Proportion of Public School Students Attending Charter Schools

How large of a role do charter schools play in American public education? The percentages of all public school students attending charter schools are shown in table 1a. These figures are based on the NAEP mathematics sample in each grade. Charter school data go back to 2003 in grade 4; 2005 in grade 8; and are only available for 2009 in grade 12.

The main findings are:

- Students attending charter schools account for a small slice of overall public school enrollment: just 3% in 2011. However, the percentage of public school students attending charter schools has increased significantly over the past decade. In grade 4, the percentage rose from 1% in 2003 to 3% in 2011, and the same was true in grade 8 from 2005 to 2011.

- While all of the subgroups analyzed showed significant increases in charter school enrollment, the most notable jump was for the black subgroup. In grade 4, the percentage of black public school students attending charter schools increased from 2% in 2003 to 7% in 2011. In grade 8, the percentage of black students enrolled in charter schools rose from 3% in 2005 to 6% in 2011.
Table 1a. National percentage of public school students attending charter schools, by grade and subgroup, for earliest year in which charter school data are available and 2011 (based on NAEP mathematics sample)

<table>
<thead>
<tr>
<th></th>
<th>National Grade 4</th>
<th>National Grade 8</th>
<th>National Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Students</td>
<td>1</td>
<td>3*</td>
<td>1</td>
</tr>
<tr>
<td>White</td>
<td>1</td>
<td>2*</td>
<td>1</td>
</tr>
<tr>
<td>Black</td>
<td>2</td>
<td>7*</td>
<td>3</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1</td>
<td>3*</td>
<td>1</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>1</td>
<td>2*</td>
<td>1</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>1</td>
<td>2*</td>
<td>2</td>
</tr>
<tr>
<td>English language learner</td>
<td>1</td>
<td>2*</td>
<td>2</td>
</tr>
<tr>
<td>Low-income (free/reduced price lunch)</td>
<td>1</td>
<td>3*</td>
<td>2</td>
</tr>
</tbody>
</table>

*The difference between base year and 2011 is statistically significant at the .05 level.

Most charter schools are located in urban areas, so it is not surprising that charter schools serve larger proportions of students in large cities than is the case nationally, as shown in table 1b. Grade 12 results are not shown because they cannot be disaggregated for large cities. For this analysis, we further disaggregated the data for the black low-income group.

The main findings regarding charter school enrollment in the large cities are:

- The percentage of public school students attending charter schools in the large cities has grown significantly. In grade 4, charter schools enrolled 3% of all large city students in 2003 and this grew to 6% in 2011. In grade 8, charter schools enrolled 5% of students in 2005 which grew to 8% in 2011.

- There has been an especially large increase in the percentage of black students attending charter schools in large cities. In 2003, 4% of black public schools students in large cities attended charter schools; in 2011 that figure rose to 12%. In grade 8 the percentage went from 8% in 2005 to 13% in 2011.

- The percentage of black low-income students attending charter schools in large cities are roughly similar to that of black students in general. Again, there was significant growth in charter school enrollment for this subgroup between 2003 and 2011.
Table 1b. Percentage of large city students attending charter schools, by grade and subgroup, for earliest year in which charter school data are available and 2011 (based on NAEP mathematics sample)

<table>
<thead>
<tr>
<th></th>
<th>Large Cities Grade 4</th>
<th>Large Cities Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Students</td>
<td>3</td>
<td>6*</td>
</tr>
<tr>
<td>White</td>
<td>2</td>
<td>5*</td>
</tr>
<tr>
<td>Black</td>
<td>4</td>
<td>12*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
<td>4*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>#</td>
<td>2*</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>2</td>
<td>5*</td>
</tr>
<tr>
<td>English language learner</td>
<td>2</td>
<td>3*</td>
</tr>
<tr>
<td>Low-income (free/reduced price lunch)</td>
<td>2</td>
<td>6*</td>
</tr>
<tr>
<td>Black low-income</td>
<td>3</td>
<td>11*</td>
</tr>
</tbody>
</table>

* The difference between base year and 2011 is statistically significant at the .05 level.
# Rounds to 0

Student Composition of Charter Schools Compared to Regular Public Schools

Tables 1c and 1d provide another perspective on who attends charter schools. These tables show differences in the student composition of charter and regular public schools, addressing questions such as: Is the student body at charter schools made up of larger or smaller percentages of low-income students than at regular public schools?

The main findings from table 1c are:

- Nationally, charter student bodies include significantly larger proportions of black students than regular public schools. Regular public schools have larger proportions of white students. This was true in grades 4 and 8 in 2003/05, and was still the case in 2011.

- In grade 4, charter student bodies included a significantly larger percentage of low-income students in 2011. A similar pattern can be seen at grade 8, but the difference between charter and regular schools is not significant.
Table 1c. At the national level, percentages of students attending charter and regular schools that fall into various subgroups (based on NAEP mathematics sample)

<table>
<thead>
<tr>
<th></th>
<th>National Grade 4</th>
<th></th>
<th>National Grade 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular public</td>
<td>Charter</td>
<td>Regular public</td>
</tr>
<tr>
<td>White</td>
<td>45</td>
<td>58*</td>
<td>35</td>
<td>53*</td>
</tr>
<tr>
<td>Black</td>
<td>31*</td>
<td>17</td>
<td>37*</td>
<td>15</td>
</tr>
<tr>
<td>Hispanic</td>
<td>20</td>
<td>19</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>2</td>
<td>4*</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>8</td>
<td>11*</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>English language learners</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Low-income (free/ reduced price lunch)</td>
<td>42</td>
<td>44</td>
<td>59*</td>
<td>52</td>
</tr>
</tbody>
</table>

*The difference between the percentages for charter and regular public schools is statistically significant at the .05 level.

The findings for large cities, displayed in table 1d, are:

- As with the national picture, large city charter schools have larger proportions of black students, although at grade 8 in 2011, the difference is not statistically significant.

- In grade 4, regular schools in the large cities had a significantly larger percentage of Hispanic students in their student bodies in 2011. In grade 8 the same was true in 2005, but by 2011, the Hispanic composition of the two types of schools became more even.

- In the large cities, regular public schools had a higher rate of enrollment for students with disabilities in 2011, and at grade 8, the difference was significant.

- In large cities, regular public schools had significantly larger proportions of English language learners at grade 4 in 2003 and 2011.
Table 1d. In large cities, percentages of students attending charter and regular schools that fall into various subgroups (based on NAEP mathematics sample)

<table>
<thead>
<tr>
<th></th>
<th>Large Cities Grade 4</th>
<th></th>
<th>Large Cities Grade 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular public</td>
<td>Charter</td>
<td>Regular public</td>
</tr>
<tr>
<td>White</td>
<td>17</td>
<td>22</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Black</td>
<td>51*</td>
<td>33</td>
<td>53*</td>
<td>25</td>
</tr>
<tr>
<td>Hispanic</td>
<td>29</td>
<td>36</td>
<td>27</td>
<td>45*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>1</td>
<td>7*</td>
<td>2</td>
<td>8*</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>English language learner</td>
<td>13</td>
<td>20*</td>
<td>12</td>
<td>22*</td>
</tr>
<tr>
<td>Low-income (free/ reduced price lunch)</td>
<td>67</td>
<td>69</td>
<td>73</td>
<td>74</td>
</tr>
<tr>
<td>Black low-income</td>
<td>35</td>
<td>28</td>
<td>41*</td>
<td>22</td>
</tr>
</tbody>
</table>

*The difference between the percentages for charter and regular public schools is statistically significant at the .05 level.
II. Student Achievement in Charter and Regular Public Schools

This section begins with a comparison of NAEP achievement in charter and regular public schools at the national level. We then focus in on performance in the large cities, the location where charter school enrollment is most concentrated. Finally, we compare achievement in charter and regular schools in a few particular urban districts where NAEP data permitted comparisons.

Achievement for All Students

Are students in charter schools reaching higher levels of achievement than students in regular public schools? Table 2a presents the results for students overall in grades 4 and 8, and table 2b shows the results for grade 12. The main findings:

- At the national level, there is a consistent pattern of higher average NAEP scores for regular public schools than for charter schools. This pattern is apparent in all subject/grade combinations and years analyzed. In some cases, such as grade 4 math, the regular schools’ scores are significantly higher.

- Grade 12 math and reading scores were significantly higher in regular public schools than in charter schools in 2009, the only year for which data are available.

- In science, regular schools also posted significantly higher scores at all grade levels in 2009.

- NAEP scores in grades 4 and 8 reading and math have increased between 2003/05 and 2011, in both regular public and charter schools, with larger gains for charter schools. Not shown in table 2a is that the gains for regular public schools tend to be statistically significant, while a similar amount of growth for charter schools does not, probably because of the small charter school sample size.

Table 2a. National NAEP achievement in charter and regular public schools, for earliest year in which charter school data are available and 2011

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular Public</td>
<td>Charter</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>212</td>
<td>217</td>
<td>218</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>255</td>
<td>260*</td>
<td>261</td>
</tr>
<tr>
<td>Grade 4 Math</td>
<td>228</td>
<td>234*</td>
<td>237</td>
</tr>
<tr>
<td>Grade 8 Math</td>
<td>268</td>
<td>278*</td>
<td>281</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.
Table 2b. National grade 12 and science results for charter and regular public schools, 2009

<table>
<thead>
<tr>
<th></th>
<th>National 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
</tr>
<tr>
<td>Grade 12 Reading</td>
<td>276</td>
</tr>
<tr>
<td>Grade 12 Math</td>
<td>138</td>
</tr>
<tr>
<td>Grade 4 Science</td>
<td>139</td>
</tr>
<tr>
<td>Grade 8 Science</td>
<td>141</td>
</tr>
<tr>
<td>Grade 12 Science</td>
<td>137</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.

Focusing on the results for large cities (grade 12 data are not available), tables 2c and 2d show that:

- Average NAEP scores for charter and regular public schools in large cities generally similar, both in the base year and 2011. The only significant difference was in grade 4 math in 2003, in favor of regular public schools. By 2011, that difference disappeared.
- Both charter and regular schools showed growth over the years in all subjects/grades analyzed.

Table 2c. Large city NAEP achievement in charter and regular public schools, for earliest year in which charter school data are available and 2011

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular Public</td>
<td>Charter</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>201</td>
<td>204</td>
<td>210</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>251</td>
<td>250</td>
<td>254</td>
</tr>
<tr>
<td>Grade 4 Math</td>
<td>216</td>
<td>224*</td>
<td>232</td>
</tr>
<tr>
<td>Grade 8 Math</td>
<td>264</td>
<td>265</td>
<td>275</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.

Table 2d. Large city NAEP science results, 2009

<table>
<thead>
<tr>
<th></th>
<th>Large Cities 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
</tr>
<tr>
<td>Grade 4 Science</td>
<td>130</td>
</tr>
<tr>
<td>Grade 8 Science</td>
<td>132</td>
</tr>
</tbody>
</table>
Achievement by Subgroup

Up to this point, results have been in favor of regular public schools. But as discussed in part I of this report, the populations of students who attend charter and regular public schools differ. Therefore, it is important, when comparing achievement across these two types of schools, to compare like groups to one another. Then one can address question such as: In large cities, are low-income students who attend charter schools performing better than their low-income peers in regular public schools?

Tables 2e and 2f display achievement broken down by subgroup for reading and math in grades 4 and 8, the subjects/grades with the most complete data. The main findings for subgroups at the national level are:

- In several subject/grade combinations, a subgroup performed significantly better in regular public schools in 2003/05; but by 2011 that subgroup performed similarly in both charter and regular schools. This was true of low-income students in grade 4 reading and math; and black students and students with disabilities in grade 8 math.

- Hispanic students used to perform similarly in charter and regular schools in 2003/05; but in 2011 they performed significantly better in grade 4 reading and grade 8 math in charter schools. The same pattern emerged for ELLs in grade 4 reading.

- Asian students were the one subgroup that performed significantly better in regular public schools in 2011, in grades 4 and 8 math.

- In both charter and regular schools, almost all subgroups showed growth in scores from 2003/05 to 2011, in all subject/grade combinations. There were just a few instances where the English language learner or Asian subgroups showed declines.
Table 2e. National subgroup achievement in charter and regular public schools, for earliest year in which charter school data are available and 2011

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular Public</td>
<td>Charter</td>
<td>Regular Public</td>
</tr>
<tr>
<td><strong>Grade 4 Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>227</td>
<td>227</td>
<td>232</td>
<td>230</td>
</tr>
<tr>
<td>Black</td>
<td>195</td>
<td>197</td>
<td>206</td>
<td>205</td>
</tr>
<tr>
<td>Hispanic</td>
<td>201</td>
<td>199</td>
<td>212*</td>
<td>205</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>224</td>
<td>225</td>
<td>226</td>
<td>234</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>180</td>
<td>184</td>
<td>184</td>
<td>186</td>
</tr>
<tr>
<td>English language learners</td>
<td>183</td>
<td>186</td>
<td>195*</td>
<td>188</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>195</td>
<td>201*</td>
<td>208</td>
<td>207</td>
</tr>
<tr>
<td><strong>Grade 8 Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>271</td>
<td>269</td>
<td>274</td>
<td>272</td>
</tr>
<tr>
<td>Black</td>
<td>239</td>
<td>242</td>
<td>251</td>
<td>247</td>
</tr>
<tr>
<td>Hispanic</td>
<td>252</td>
<td>245</td>
<td>255</td>
<td>251</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>268</td>
<td>270</td>
<td>272</td>
<td>275</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>228</td>
<td>226</td>
<td>235</td>
<td>230</td>
</tr>
<tr>
<td>English language learners</td>
<td>235</td>
<td>224</td>
<td>231</td>
<td>223</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>243</td>
<td>247</td>
<td>253</td>
<td>251</td>
</tr>
<tr>
<td><strong>Grade 4 Math</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>242</td>
<td>243</td>
<td>251</td>
<td>249</td>
</tr>
<tr>
<td>Black</td>
<td>214</td>
<td>216</td>
<td>224</td>
<td>224</td>
</tr>
<tr>
<td>Hispanic</td>
<td>219</td>
<td>221</td>
<td>234</td>
<td>229</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>248</td>
<td>246</td>
<td>243</td>
<td>256*</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>209</td>
<td>214</td>
<td>218</td>
<td>218</td>
</tr>
<tr>
<td>English language learners</td>
<td>211</td>
<td>214</td>
<td>222</td>
<td>219</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>216</td>
<td>222*</td>
<td>229</td>
<td>229</td>
</tr>
<tr>
<td><strong>Grade 8 Math</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>284</td>
<td>288</td>
<td>296</td>
<td>293</td>
</tr>
<tr>
<td>Black</td>
<td>248</td>
<td>254*</td>
<td>266</td>
<td>262</td>
</tr>
<tr>
<td>Hispanic</td>
<td>270</td>
<td>261</td>
<td>277*</td>
<td>269</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>287</td>
<td>295</td>
<td>287</td>
<td>303*</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>235</td>
<td>244*</td>
<td>254</td>
<td>249</td>
</tr>
<tr>
<td>English language learners</td>
<td>251</td>
<td>244</td>
<td>254</td>
<td>243</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>255</td>
<td>262</td>
<td>271</td>
<td>269</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.
The findings tend to favor charter schools the more we focus in on large cities and subgroups within the large cities. The main findings from Table 2f are:

- In 2011, the Hispanic subgroup performed significantly higher in charter schools than in regular schools. This was true in grade 4 reading, and in grades 4 and 8 math.

- Black students in large cities performed similarly in charter and regular schools in the base year of our analysis (2003/05). But in 2011, achievement for black students was significantly higher in charter schools in grade 8 reading and grade 4 math.

  - A similar pattern was found for the black-low-income subgroup. In 2011, achievement for this group was significantly higher in charter schools in grade 8 reading, and grades 4 and 8 math. The black low-income group in charter schools showed a striking increase in scores between 2003/05 and 2011.

  - The only significant findings in favor of regular schools in 2011 were for the Asian subgroup (grade 4 math) and the white subgroups (grade 4 reading).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular</td>
<td>Charter</td>
</tr>
<tr>
<td><strong>Grade 4 Reading</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>220</td>
<td>226</td>
<td>225</td>
</tr>
<tr>
<td>Black</td>
<td>193</td>
<td>193</td>
<td>206</td>
</tr>
<tr>
<td>Hispanic</td>
<td>202</td>
<td>197</td>
<td>210*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>‡</td>
<td>223</td>
<td>211</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>175</td>
<td>175</td>
<td>177</td>
</tr>
<tr>
<td>English language learners</td>
<td>181</td>
<td>184</td>
<td>189</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>192</td>
<td>197</td>
<td>205</td>
</tr>
<tr>
<td>Black low-income</td>
<td>189</td>
<td>191</td>
<td>203</td>
</tr>
<tr>
<td><strong>Grade 8 Reading</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>269</td>
<td>270</td>
<td>273</td>
</tr>
<tr>
<td>Black</td>
<td>241</td>
<td>239</td>
<td>250*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>253</td>
<td>243</td>
<td>251</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>‡</td>
<td>266</td>
<td>269</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>239</td>
<td>213</td>
<td>226</td>
</tr>
<tr>
<td>English language learners</td>
<td>239</td>
<td>220</td>
<td>227*</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>243</td>
<td>243</td>
<td>250</td>
</tr>
<tr>
<td>Black low-income</td>
<td>233</td>
<td>237</td>
<td>250*</td>
</tr>
<tr>
<td><strong>Grade 4 Math</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>230</td>
<td>243*</td>
<td>249</td>
</tr>
<tr>
<td>Black</td>
<td>210</td>
<td>212</td>
<td>227*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>217</td>
<td>219</td>
<td>232*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>‡</td>
<td>246</td>
<td>231</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>199</td>
<td>204</td>
<td>205</td>
</tr>
<tr>
<td>English language learners</td>
<td>209</td>
<td>211</td>
<td>219</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>212</td>
<td>217*</td>
<td>228</td>
</tr>
<tr>
<td>Black low-income</td>
<td>207</td>
<td>210</td>
<td>225*</td>
</tr>
<tr>
<td><strong>Grade 8 Math</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>286</td>
<td>288</td>
<td>292</td>
</tr>
<tr>
<td>Black</td>
<td>250</td>
<td>250</td>
<td>267</td>
</tr>
<tr>
<td>Hispanic</td>
<td>272</td>
<td>258</td>
<td>278*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>295</td>
<td>289</td>
<td>281</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td>233</td>
<td>230</td>
<td>247</td>
</tr>
<tr>
<td>English language learners</td>
<td>250*</td>
<td>238</td>
<td>256</td>
</tr>
<tr>
<td>Low-income</td>
<td>256</td>
<td>256</td>
<td>272*</td>
</tr>
<tr>
<td>Black low-income</td>
<td>244</td>
<td>246</td>
<td>265*</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.
Achievement for Selected Urban Districts

NAEP reports results for a number of large school districts, referred to as the Trial Urban District Assessment (TUDA). Due to a policy change in 2009, charter schools within the geographic boundaries of an urban district, but operated independently of the district and not included in the district’s Adequate Yearly Progress (AYP) report under NCLB, are no longer included in the NAEP results for TUDA districts. School districts vary in terms of whether the charter schools within their boundaries are included in the district’s AYP results.

Table 2g presents results for several districts with charter school data available. Results are shown for all students and the low-income subgroup. Other subgroups (e.g., black, Hispanic) are not included because samples were not large enough to meet NAEP reporting standards.

The findings are:

- Within the four cities analyzed, the pattern is clear that students in charter schools significantly outperform their peers in regular public schools.

- There was no subject/grade in any of the cities analyzed where regular schools scored significantly higher than charter schools.

Why do charter schools underperform regular public schools nationally, but outperform them in large cities? At the national level, NAEP results for regular public schools include a wide range of schools, including many high-performing suburban schools in high-income communities. If the purpose is to compare charter and regular schools, the fairest way is to compare them within similar locations. As noted above, charter school enrollment is most concentrated in large cities: 6% of large city grade 4 students and 8% of large city grade 8 students attend charter schools, compared to 3% at each grade nationally. Making comparisons within large cities increases the likelihood that regular and charter schools are serving similar populations of students.

Table 2g. For selected urban districts, subgroup achievement in charter schools (CH) and regular public schools (RP)

<table>
<thead>
<tr>
<th></th>
<th>DC</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CH</td>
<td>RP</td>
<td>CH</td>
<td>RP</td>
<td>CH</td>
<td>RP</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All students</td>
<td>200</td>
<td>201</td>
<td>216</td>
<td>211</td>
<td>214*</td>
<td>202</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>195*</td>
<td>188</td>
<td>207</td>
<td>200</td>
<td>213*</td>
<td>199</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All students</td>
<td>249*</td>
<td>237</td>
<td>264*</td>
<td>252</td>
<td>254</td>
<td>253</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>245*</td>
<td>228</td>
<td>262*</td>
<td>246</td>
<td>257*</td>
<td>249</td>
</tr>
<tr>
<td>Grade 4 Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All students</td>
<td>222</td>
<td>222</td>
<td>233*</td>
<td>228</td>
<td>235*</td>
<td>223</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>218*</td>
<td>211</td>
<td>223</td>
<td>218</td>
<td>234*</td>
<td>220</td>
</tr>
<tr>
<td>Grade 8 Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All students</td>
<td>267*</td>
<td>255</td>
<td>274*</td>
<td>265</td>
<td>269</td>
<td>270</td>
</tr>
<tr>
<td>Low-income (free/reduced lunch)</td>
<td>262*</td>
<td>246</td>
<td>270*</td>
<td>259</td>
<td>270</td>
<td>267</td>
</tr>
</tbody>
</table>

*The difference between charter and regular public schools is statistically significant at the 0.05 level.
III. Other School Factors

In this section we compare charter and regular public schools on several school factors other than student achievement: time spent on core subjects, class size, and teacher qualifications. The NAEP background data also include many questions about instructional content and methods that are more subject-specific and are not addressed in this report.

These analyses were conducted for large cities rather than at the national level because, as demonstrated in earlier parts of this report, that is where the majority of charter schools are located and it is more meaningful to compare charter and regular schools serving similar populations of students in similar locations. We do not present results for particular school districts, as we did in part II, because of insufficient data at the district level.

**Time Spent on Core Subjects**

Do charter and regular schools located in large cities differ in the amount of time that they spend on core subjects? **Table 3a** displays the 2011 results for large cities:

- A significantly higher percentage of regular school teachers reported spending less than five hours a week on grade 8 language arts. Overall, it appears that charter school students are getting more instruction in grade 8 language arts.

- A significantly higher percentage of regular school teachers reported spending 5-7 hours a week on grade 4 math. Higher percentages of charter school than regular teachers reported spending more than 7 hours a week in that subject but the difference is not significant.

- There is suggestive evidence (large differences but not statistically significant) that charter school students are getting more instruction in grade 8 math.

**Table 3a. In large cities, percentages of teachers reporting various amounts of time spent per week on language arts and math instruction, 2011**

<table>
<thead>
<tr>
<th></th>
<th>Less than 5 hrs/week</th>
<th>5-6.9 hrs/week</th>
<th>7 or more hrs/week</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular Public</td>
<td>Charter</td>
</tr>
<tr>
<td>Grade 4 Language Arts</td>
<td>11</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Grade 8 Language Arts</td>
<td>19</td>
<td>35*</td>
<td>47</td>
</tr>
<tr>
<td>Grade 4 Math</td>
<td>13</td>
<td>9</td>
<td>39</td>
</tr>
<tr>
<td>Grade 8 Math</td>
<td>35</td>
<td>46</td>
<td>35</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.
Class Size

The NAEP background questionnaire asks teachers how many students are in the class being assessed. Table 3b shows that:

- Class sizes tend to be larger in regular schools than in charter schools. Significantly more regular school teachers reported having more than 26 students in their classes in grade 4 language arts and grade 8 math.

Table 3b. Percentages of teachers in large cities reporting various numbers of students in their classes, 2011

<table>
<thead>
<tr>
<th></th>
<th>18 or fewer students</th>
<th>19-25 students</th>
<th>26 or more students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular Public</td>
<td>Charter</td>
</tr>
<tr>
<td>Grade 4 Language Arts</td>
<td>24</td>
<td>11</td>
<td>40</td>
</tr>
<tr>
<td>Grade 4 Math</td>
<td>20</td>
<td>12</td>
<td>42</td>
</tr>
<tr>
<td>Grade 8 Language Arts</td>
<td>10</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>Grade 8 Math</td>
<td>12</td>
<td>7</td>
<td>51</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.

Teacher Qualifications

Table 3c presents findings regarding the qualifications of math teachers, and table 3d presents findings for English/language arts teachers:

- In large cities, significantly more students in regular public schools have teachers with a major in the subject being taught, whether it is math (grade 4 and 8) or English/language arts (grade 8).

- Significantly more grade 8 students in charter schools are taught by math teachers who entered the teaching profession through an alternate certification program.

- Significantly more students in regular public schools have teachers with more than 10 years of teaching experience than those in charter schools. Significantly more charter school students are taught by teachers with less than 4 years of teaching experience. This is true at grades 4 and 8, for both math and reading.
Table 3c. Percentages of students in large cities whose teachers have various qualifications, based on NAEP math teacher questionnaire, 2011

<table>
<thead>
<tr>
<th></th>
<th>Grade 4 Math</th>
<th></th>
<th>Grade 8 Math</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular</td>
<td>Charter</td>
<td>Regular</td>
</tr>
<tr>
<td>Alternate certification program</td>
<td>18</td>
<td>18</td>
<td>47*</td>
<td>30</td>
</tr>
<tr>
<td>Certified by NBPTS</td>
<td>12</td>
<td>13</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Major in math</td>
<td>4</td>
<td>9*</td>
<td>11</td>
<td>21*</td>
</tr>
<tr>
<td>Major in math education</td>
<td>16</td>
<td>21</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>Years taught</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-4 years</td>
<td>47*</td>
<td>23</td>
<td>43*</td>
<td>19</td>
</tr>
<tr>
<td>5-9 years</td>
<td>33</td>
<td>33</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>10-19 years</td>
<td>16</td>
<td>30*</td>
<td>21</td>
<td>33*</td>
</tr>
<tr>
<td>20+ years</td>
<td>3</td>
<td>14*</td>
<td>5</td>
<td>18*</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.

Table 3d. Percentages of students in large cities whose teachers have various qualifications, based on NAEP reading teacher questionnaire, 2011

<table>
<thead>
<tr>
<th></th>
<th>Grade 4 Reading</th>
<th></th>
<th>Grade 8 Reading</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Charter</td>
<td>Regular</td>
<td>Charter</td>
<td>Regular</td>
</tr>
<tr>
<td>Alternate certification program</td>
<td>23</td>
<td>16</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>Certified by NBPTS</td>
<td>12</td>
<td>13</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Major in English</td>
<td>6</td>
<td>6</td>
<td>39</td>
<td>45</td>
</tr>
<tr>
<td>Major in reading language arts</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>20*</td>
</tr>
<tr>
<td>Years taught</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-4 years</td>
<td>37*</td>
<td>16</td>
<td>43*</td>
<td>18</td>
</tr>
<tr>
<td>5-9 years</td>
<td>33</td>
<td>33</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td>10-19 years</td>
<td>16</td>
<td>30*</td>
<td>13</td>
<td>35*</td>
</tr>
<tr>
<td>20+ years</td>
<td>3</td>
<td>14*</td>
<td>9</td>
<td>20*</td>
</tr>
</tbody>
</table>

*The difference between charter schools and regular public schools is statistically significant at the .05 level.
IV. Strengthening the NAEP Data and the Data Explorer

One of the purposes of this exploratory study was to identify ways that the NAEP background data could be made more useful. So in this final section, we offer some suggestions based on our experience using the NAEP data, not only on this project, but in our past research.

The main way that researchers access the NAEP data are through the online NAEP Data Explorer. It is a powerful tool and, with practice, researchers can conduct a wide variety of analyses. However, the interface is not as user-friendly as it could be. Many of the actions necessary to conduct certain analyses are not intuitive, and it is only after trial-and-error (or with the help of a more-experienced colleague) that one discovers how to retrieve the data or carry out the desired analyses. It would be worthwhile to take the steps necessary to refine the Data Explorer, so that research projects could be conducted as intuitively and efficiently as possible. Perhaps this could be done by gathering some experienced users of the Data Explorer together with programmers who develop the website, to have a brainstorming session on ways to improve the interface. We are starting out with a good tool, so there is no need to reinvent it—simply to refine it and work through some of the kinks that exist.

One major difficulty for the user is sorting through the hundreds of background variables listed in the Data Explorer to find the few of interest. The majority of background questions listed in the Data Explorer were asked in earlier years but are no longer asked. So the researcher must scroll through hundreds of variables which are no longer relevant or for which data are no longer available. It would be helpful if, once the researcher checks the years of interest, only those variables available for that year (or years) show up on the screen.

The background variables are also not organized as well as they could be. The responses from the school, teacher, and student-level questionnaires are often mixed together. If one is interested in time spent on instruction in a certain subject, it is necessary to scroll through dozens of variables to find that question, unless one happens to know that it is under the heading of “Instructional Content and Practice” and then under the subheading “Classroom Management” (not at all intuitive). One option might be to add a search bar to the program, so that the user could type in a key word and find relevant variables that way.

We realize that NCES regularly reviews the background variables, but it still seems that there is some redundancy and areas where useful information is missing. Some questions may be out of date, such as whether a student has a set of encyclopedias at home. Other background variables seem overly detailed, such as some of those in the area of teacher professional development. There are also old questions that could be reinstated, such as those regarding school climate—school safety, parent involvement, gang activity, discipline—that are still clearly relevant and of interest to education stakeholders.

Finally, we ran into a specific problem when trying to research charter schools, related to the TUDA data. Due to a policy change in 2009, charter schools within the geographic boundaries of an urban district, but operated independently of the district and not included in the district’s AYP report under NCLB, are no longer included in the NAEP results for TUDA districts. School districts vary in terms of whether the charter schools within their boundaries are included in the
district’s AYP results. While we understand why districts would not want charter schools included in their results, it makes it impossible to compare charter and non-charter performance within most major cities where large numbers of charter schools are located. In this study we found a clear pattern of charter schools outperforming regular schools in the four cities that had sufficient data. It would be desirable to find a way to include an option for including non-district-governed charter schools in an urban district’s results, solely for research purposes (not for general TUDA reporting). After all, charter schools are likely to continue to grow in importance and be the focus of research. These schools are serving ever-larger numbers and percentages of students—particularly minority and low income students—and the large urban districts are where they should be studied.
REFERENCES


TIME FOR LEARNING:
An Exploratory Analysis
of NAEP Data

PREPARED FOR THE NATIONAL
ASSESSMENT GOVERNING BOARD

By Alan Ginsburg and Naomi Chudowsky

December 2012

The authors wish to thank Lawrence Feinberg, of the Governing Board staff, for his very helpful comments and editorial suggestions.
ABSTRACT

This report uses NAEP background data to track time and learning since the mid-1990s in three areas: student absenteeism; classroom instructional time in mathematics, reading, music and the visual arts; and homework time expected by teachers.

Key report findings are:

- **Students with higher rates of *monthly absenteeism* score disproportionately at the Basic or below-Basic levels of NAEP achievement for grades 4, 8 and 12. About one-quarter of below-Basic students were absent three days or more a month in 2011, which translates into missing more than five weeks of school over a year. By contrast only one-in-ten Advanced students were absent three or more days a month.**

  Given the strong association between student achievement and absenteeism, it is sensible for schools to focus on improving the attendance of lower-achieving students with high absenteeism rates as part of their efforts to boost academic achievement. However, the NAEP data show that there was little or no change in the percentage of students absent 3 or more days between 1994 and 2011.

- **Average weekly *instructional time* is greater in reading than in mathematics. Instructional time in both subjects declines markedly from grade 4 to grade 8.**
  
  - Mathematics and reading instructional time has increased at both grades 4 and 8 since the mid-1990s, but 40 to 50 percent of grade 8 students still spend less than five hours per week on these two core subjects.
  - At grade 8, over half the below-Basic students on NAEP achievement levels spend less than five hours a week (i.e., less than an hour a day) on mathematics instruction; about 40 percent of these lowest-achievers spend less than an hour a day on instruction in reading-language arts.

- **The *frequency of instruction in music and the visual arts*—when measured by the number of times these subjects are taught at grade 8 each week—did not decline between 1994 and 2008, as some education experts have suggested.**

- **More *homework time* is expected by teachers at grade 8 than at grade 4, but the amounts have not changed markedly between the mid-1990s and 2011 despite the pressures from No Child Left Behind. Black and Hispanic students are expected to spend somewhat more time on homework than Whites—perhaps a response by teachers to lower average achievement—but American Indian students, also a lower scoring group, are not given more homework than White students.**
It is recommended that the National Assessment Governing Board (NAGB) consider further exploratory analyses. The priority should be to report time use for individual states and urban districts participating in NAEP and for additional subjects, especially science. The additional reports could form part of a series, possibly entitled *NAEP Portraits of American Education*, which would include reports based on other background variables as well. To provide data for a comprehensive analysis of students’ time for learning, NAGB should consider extending the background questionnaires to cover the length of the school day, the length of the school year, and learning-related activities beyond the regular school day, both formal and informal. Consistency of wording with the major international assessments of PIRLS, PISA, and TIMSS should also be explored.
EXECUTIVE SUMMARY

The NAEP student background variables provide a rich source of information to track time and learning in U.S. schools. Based on the NAEP data, this report examines the time students use for learning from three perspectives: (1) days absent from school; (2) classroom weekly instructional time in mathematics, reading, music and the visual arts; and (3) the amount of time teachers expect students to spend each day on homework.

Research indicates that the amount of time spent on learning is related to student achievement provided that the time is used to provide high-quality instruction (Aronson, et.al., 1999; Silva, 2007). The time-outcome relationship is especially strong for at-risk students (National Center on Time and Learning, 2011).

NAEP is a unique national resource in its availability of data since the mid-1990s on student achievement and time use for learning (Smith, et.al, 2012). This period covers changes that span two major federal education reforms: systemic reform under President Clinton, introduced in the 1994 Improving America Schools Act; and school accountability for proficiency under President Bush’s 2001 No Child Left Behind Act.

The NAEP data have limitations as well as strengths. One drawback is that the NAEP background questionnaires do not ask about length of the school day or the number of days in the school year, so that a comprehensive picture of students’ time for learning in school is not available. A second limitation is that the wording of questions on the NAEP background questionnaires frequently shifted between the mid-1990s and 2011, which limits precise tracking of how time use has changed.

Days Absent

Teachers can’t effectively teach students who are frequently absent from school. NAEP questionnaires ask students in grades 4, 8 and 12 about their days absent in the prior month with responses ranging from no days absent to five or more. To place responses in context, students absent three or more days in the prior month will, at that rate, miss at least five weeks of instruction in a school year.

Higher absenteeism and lower NAEP achievement scores are closely associated at grades 4, 8 and 12. Typical of the findings, the grade 8 results (Exhibit A) indicate that:

- Fifty-six percent of Advanced-achieving students had perfect attendance during the prior month compared with only 39 percent of those below Basic.

---

1 The analyses presented in this report were made possible by the very useful NAEP Data Explorer.
• Conversely, 28 percent of the below-Basic students were absent three or more days in the prior month compared with only 10 percent of those at Advanced.

**Exhibit A. Percentages for grade 8 students at different NAEP reading achievement levels by days absent from school the prior month: 2011**

<table>
<thead>
<tr>
<th>Days absent last month</th>
<th>below Basic</th>
<th>at Basic</th>
<th>at Proficient</th>
<th>at Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>39</td>
<td>33</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>1-2 days</td>
<td>45</td>
<td>36</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>3-4 days</td>
<td>56</td>
<td>34</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>5 or more days</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given the strong association between student achievement and absenteeism, it is sensible for schools to focus on improving the attendance of lower-achieving students with high absenteeism rates as part of their efforts to boost academic achievement. Certainly, if children are absent from school, they may well find it more difficult to learn what is taught in their classes.

The NAEP data show, however, that there was little or no change in the percentage of students absent 3 or more days between 1994 and 2011. Despite some improvement at grade 12, about the same proportion—around one-fifth—has had such high absentee rates for almost two decades. One possible cause of the failure to make substantial improvement may be that few states track absenteeism in their public reporting of school data. A recent study found only six states that do so (Balfanz & Byrnes, 2012). The old adage “what gets measured gets done” seems to hold here. Tracking excessive absenteeism could be an important step in leading schools to focus on the problem.

The NAEP data on days absent from school in the prior month also show:

• About one-fifth of the students at grades 4, 8 and 12 were absent three or more days in the prior month (equivalent to five weeks a year).
• Perfect attendance in the prior month declines from 51 percent at grade 4 to 38 percent at grade 12, although it is not clear why students at grade 4 should be less sick.
• Private school students are somewhat less likely to be absent three or more days in the prior month than students in public school (about a 5 percentage point differential at grade 8).
• American Indian students have higher absenteeism rates than any other racial/ethnic group. For example, at eighth grade, 31 percent of American Indian students report having been absent three days or more in previous month, compared with about 20 percent of whites, blacks, and Hispanics, and only 11 percent of Asian/Pacific Islanders.
• Students eligible for free or reduced-price lunch, who are from lower-income families, are 7 percentage points more likely to be absent at least 3 days a month at grades 4 and 8 and 3 percentage points more likely at grade 12.

Instructional Time in Mathematics and Reading

Research indicates that the amount of time students spend in instruction, along with the quality of instructional time, exerts more influence on learning than the length of the school day (Silva, 2007). Given the pressures of testing to demonstrate adequate yearly progress in mathematics and reading, there is interest in knowing whether instructional time in these subjects is extended for at-risk populations and whether it has increased since the mid-1990s.

Instructional time in mathematics. This is particularly important for student learning as mathematics, unlike reading, is learned primarily at school. In examining students’ time for learning mathematics, mathematics instruction is compensatory with respect to student achievement if lower achievers spend more time on mathematics each week than higher achievers.
Exhibit B shows that the distribution of mathematics instructional hours is compensatory with respect to providing greater mathematics instruction to students at lower NAEP achievement levels in grades 4 and 8. Nevertheless, a majority of grade 8 students who score below Basic, the lowest NAEP achievement category, are receiving less than an hour a day of mathematics instruction.

- More than half (55 percent) of the below-Basic students spend less than an hour a day (5 hours a week) on mathematics instruction. Among students at Advanced 73 percent spend less than an hour a day in math classes.

- At grade 4, lower achievers are more likely to receive some additional mathematics instruction. Eighty-seven percent are spending more than five hours a week on math, including 31 percent that spend 7 hours a week or more. However, the differences across achievement levels are not large.

Other significant findings about mathematics instructional time include:

- The typical (modal) time spent in mathematics each week declines significantly between grades 4 and 8. At 4th grade, 59 percent of the students spend 5-6.9 hours per week in mathematics instruction, but at 8th grade 63 percent of students spend less than five hours a week on mathematics.
• Time spent on mathematics clearly has increased since the mid-1990s, but changes in the wording of background questions have limited NAEP’s ability to report precise comparisons over time.
  o At grade 4, in 1996 about 34 percent of U.S. students spent less than four hours on mathematics each week. By 2011 only 12 percent of students spent less than five hours a week on math.
  o At grade 8, in 1996 only 33 percent of students spent four or more hours in mathematics instruction; by 2003 that had increased to 51 percent. In 2005, 31 percent of students spent five or more hours in mathematics (at least an average of an hour a day); by 2011 that had increased to 37 percent.

• Instructional time is also compensatory with respect to Black, Hispanic, and American Indian students, three student groups with lower average achievement than Whites and Asians. At grades 4 and 8, the proportion of these minority-group students spending seven hours a week or more on math instruction is between 6-12 percentage points higher than for Whites and Asians.

• Public schools provide more time for mathematics instruction than do private schools: 31 percent of the grade 4 students in public school spend seven or more hours a week on mathematics compared with only seven percent in private school, a difference of 24 percentage points. However, at grade 8 the difference narrows to only seven percentage points.

• A greater percentage of students eligible for school lunch receive 7 or more hours of mathematics instruction each week – 9 percentage points more at grade 4 and 7 points more at grade 8.

Instructional time in reading. The 2011 NAEP questionnaire asks teachers of reading about time spent on instruction in language arts – formal reading, grammar and writing. Before 2005, the NAEP questionnaires asked only for the time spent on formal reading, which excluded writing and grammar. Thus, the time comparisons NAEP can make to current practice go back only six years.
Exhibit C shows reading-language arts instructional time for students by NAEP achievement levels. It indicates that while the distribution of weekly time spent on reading is compensatory with respect to NAEP achievement levels, nonetheless:

- Time spent on reading per week declines sharply between grades 4 and 8. About 47 percent of grade 4 students on average across all achievement levels receive 10 hours or more hours of reading instruction a week—at least two hours per day. An equal 47 percent of grade 8 students on average across all achievement levels spend less than 5 hours a week on reading-language arts instruction—less than an hour a day.

- At grade 8, nearly four-in-ten below-Basic students spend less than five hours a week on instruction in reading-language arts. This is less than an hour a day for the lowest reading group. At grade 4, about 12 percent of below-basic students spend less than 5 hours a week on reading compared to 9 percent of advanced students.

Other findings on instructional time in reading-language arts include:

- The change in time spent on reading instruction at grade 4 was minimal between 2005 and 2011. At grade 8, where NAEP has a longer time series, there was a
modest increase from 2002 to 2011—about 10 percentage points in students receiving more than an hour of instruction a day.

- At grade 4 the patterns of time spent on reading instruction are similar for all racial/ethnic groups except Hispanics, who receive significantly more time than whites. At grade 8, blacks and Hispanics—but not American Indians—spend more time on reading instruction than do whites and Asians.

- At grade 4 almost half of public school students receive more than 10 hours of reading-language arts instruction per week, compared with just 22 percent of those in private school. The public-private school differences are much less at eighth grade.

- The pattern of reading instruction for school-lunch eligible students at grade 4 is similar to that for students from higher-income families, but at grade 8 school-lunch eligible students are 9 percentage points more likely to receive 7 or more hours per week of reading instruction.

**Frequency of Music and Visual Arts Instruction**

Music and visual arts are essential elements of a K-12 curriculum. Because these subjects are not covered by federal requirements for annual testing, there are concerns they may be de-emphasized in the current curriculum. Although some retrospective studies suggest a decline in music and visual arts exposure since No Child Left Behind was enacted in 2001, a recent NCES Fast Response Survey (Parsad & Spiegelman, 2012) found that schools offering at least some music and the visual arts had not decreased between 1998-2010.
The NAEP data findings are similar, as shown in Exhibit D. Moreover, the NAEP data go beyond the information from the Fast Response Survey in that they compare the frequency of instructional offerings during the week, not just whether music and the visual arts are offered.

- The frequencies of weekly offerings in music and the visual arts have not declined between 1994 and 2008 at grade 8 (the level surveyed by NAEP).
- In fact, NAEP reports an increase from 49 percent to 57 percent in the proportion of eighth grade students receiving instruction in music at least 3 times a week.

The data also show that exposure to music and art is similar across racial/ethnic groups. Public school students receive more instruction in the arts than their peers in private school. School-lunch eligible and ineligible students experience similar patterns of frequency of music exposure, but 8 percentage points fewer school-lunch eligible students receive at least some classes in visual art.

**Expected Homework Time**

Given the pressures to meet No Child Left Behind improvement requirements, it might seem likely that the time teachers expect students to spend on homework would increase over the last decade. NAEP asked similar questions from 1996 to 2011 about homework time in mathematics, but it has included no questions about homework time in reading.
The results shown in Exhibit E indicate that the change in homework time between 1996 and 2011 in mathematics differed at grades 4 and 8:

- At grade 4 the expected time students would spend on mathematics homework was quite similar between 1996 and 2011. Most students were expected to spend 15-30 minutes daily on mathematics homework in both years.
- By contrast, grade 8 students with heavy mathematics homework assignments—one hour or more a day—rose from 4 to 17 percent.

The 2011 results show two other differences in expected mathematics homework:

- At both grades 4 and 8, private school students are expected to do somewhat more homework each day than public school students. The difference at grade 4 is small, but at grade 8, about 25 percent of private school students are expected to spend an hour or more a day on math homework compared with only 16 percent of public school students.
- Across racial/ethnic groups, American Indians, despite low average achievement, have the lowest amount of homework time at grades 4 and 8 along with Whites (whose average achievement is considerably higher).
- More grade 4 students eligible for school lunch receive higher amounts of homework than students who are ineligible, but at grade 8 the pattern of expected homework time is similar for both groups.

**Recommendations for Further Analyses and Improvements in the Data**

This report has examined at the national level three important aspects of student time and learning: absenteeism; instructional time in mathematics, reading, music and visual art; and the time expected for homework. Recommendations in three areas are proposed for consideration.

1. Two additional analyses on time for learning would be very useful:

   - *Track instructional time for additional academic subjects, possibly to be presented in separate reports.* Science would be a high priority for an additional report because of wide public interest in this field and the federal requirement for annual student assessments.

   - *Disaggregate and display selected key indicators of time use for individual states and participating urban districts* so these jurisdictions can compare their use of time for learning. Such reporting should be part of an online compendium of key
background indicators for states and participating urban districts, as proposed in recommendation 4b of the Expert Panel Report to NAGB on Strengthening NAEP Background Questions (Smith et al., 2012).

2. The Governing Board should consider establishing a series of NAEP Portraits of American Education. The series could be initiated by the time and learning analyses along with the exploratory study of NAEP data on charter schools, which has also been prepared for the Governing Board. The series could then cover other sets of background variables such as those for teachers, curriculum, technology, private schools and students’ out-of-school learning.

3. Our experiences in preparing this report have prompted several methodological suggestions to improve future analyses of NAEP background information:

- **Consider adding questions to the school questionnaire on the length of the school day and school year so that a more comprehensive picture of student time for learning in school is available.** In particular, there is considerable interest in learning more about the length of the school day. For example, the National Center on Time and Learning (2012) in partnership with the Ford Foundation has announced a five-state time collaborative to support expanded learning time in schools. Yet, currently basic data are not available on the distribution of the length of the school day across American schools.

- **Consider adding to the teacher or school survey a question on whether academic performance is used to assign students to more instructional time in a particular subject.** This question would address an important possible reason for instructional time differences.

- **Consider adding questions to the student questionnaire on time spent in out-of-school learning situations**—including formal classes or tutoring, visits to museums and historic sites, cultural programs, and online activities related to learning. While currently after-school learning situations consist mostly of formal tutoring and other organized activities, the rapid expansion of online learning will make it increasingly important to have a good picture of how students spend their time learning outside the regular school day. Indeed, educators are widely discussing the flipped classroom—primarily delivering instruction electronically and doing additional tutoring or homework activities in school. NAEP could provide important information on these out-of-school learning trends.

- When considering the wording of the time-for-learning questions, **examine consistency with the wording of similar questions on the major international assessments** of PIRLS, PISA and TIMSS. Also, consider consistency in wording over time, as proposed in recommendation 1d of the Expert Panel Report on Background Questions (Smith et al., 2012). It might also be desirable to have
consistent wording of the questions in different subjects. For example, the questions about music and the visual arts have asked about the frequency of instruction each week, but not about the amount of instructional time, which is asked—with different time intervals—about instruction in mathematics and reading. Consistent wording would improve time-use comparisons across subjects. However, some of these goals may be conflicting, and careful judgments should be made about which type of consistency is most important.

- *Further improve the powerful online tools for NAEP data analysis.* (Recommendation 4e of the Expert Panel Report.) In particular, software should be extended to build in the capability for multivariate analysis.
TIME FOR LEARNING

Introduction

This report explores the NAEP background variables for students’ time for learning covering the period from the mid 1990’s to 2011. Student variables describing educational time-use are examined in three topic areas of: student days absent from school; classroom instructional time in mathematics, reading, music and the visual arts; and amount of time teachers expect students to spend on homework.

Research consistently finds that exposure to high-quality instructional time that engages students in learning improves student achievement (Aronson, et.al., 1999; Silva, 2007). Thus, time students spend in instruction and on homework, along with the quality of that instructional time and homework, are key elements of students’ opportunity-to-learn to achieve to high academic standards. This connection between time and learning is particularly strong for students who are most at-risk of school failure (Dobie and Fryer 2011; National Center on Time and Learning, 2011).

The National Center on Time and Learning (http://www.timeandlearning.org/) clarifies the underlying theory of why greater amounts of time use, provided that it is of high-quality, can improve student learning include

- “Longer classes allow teachers to cover: more material and examine topics in greater depth; build-in more project-based and hands-on learning; individualize and differentiate instruction; and answer students’ questions.
- Setting aside whole periods each day to focus on small-group instruction to address and overcome student learning deficits.
- With more time, schools do not have to cut back class time in science, social studies, music, art and physical education in order to give more time to the heavily tested subjects of English Language Arts and math.”

This report takes advantage of NAEP’s unique national resource as a repeated national survey on student achievement and student background variables (Smith et al., 2012). Using the student background variables, the report will explore days absent from school and classroom instructional time for students at different NAEP achievement levels, attending public or private schools, and by race/ethnicity or school lunch eligibility (which is an indicator of low-income status).

Along with information on time use for reading and mathematics, NAEP is also unique in its collection of national information on other academic subjects. These include music and the visual arts, which some educators believe are being neglected as schools focus on the state-assessed subjects of mathematics and reading.
The NAEP data are also a unique national resource in that they administer the background questions on time use since the mid-1990s. Over this period, the NAEP student time and learning data span the introduction of two major pieces of legislation: systemic reform under President Clinton’s 1994 Improving the America Schools Act (103 Congress), and school accountability for student proficiency required under President Bush’s 2001 No child Left Behind Act (PL 107-110).

An important technical challenge in analyzing the NAEP data over time is that questions with different wording have been asked about time use on the same topic in different years. Whatever the rationale for individual question wording in a particular year, the changes in wording have disrupted the time series. As a result, the value of NAEP has been diminished for making consistent comparisons about student time use since the mid-1990s.

This report employs several techniques for analyzing the available data over time, so that the information value of these historical data is not lost. One approach is to consolidate responses presented in several shorter time-reporting intervals (e.g., less than 2 hours, 2 to less than 4 hours of reading) into longer intervals that better match-up across different time periods (e.g., less than 4 hours of reading). A second approach is to separately report how students’ time-use has changed only across the years for which the questions have common wording.

The following sections of this exploratory report describe the NAEP data for four sets of variables related to students’ time and learning:

1. Days absent from school;
2. Classroom instructional time in federally-required assessed subjects of reading and mathematics;
3. Frequency of classroom instruction in music and the visual arts, which are not covered under federal assessment requirements; and
4. Teacher expected homework time.

Each set of students’ time-related variables for school are described in terms of their 2011 national values for all students on the most recent NAEP assessment and by how the 2011 values compare with the values from earlier years. The 2011 values for the time variables are then disaggregated to display their distribution by student characteristics including students’ NAEP achievement levels, public or private school enrollment, race/ethnicity and school-lunch eligibility.

It is important to note that this report is descriptive and does not make direct causal connections between the time-use variables and student outcomes. Instead, causal interpretations on the importance of the observed time differences should come from information developed through rigorous causal research employing experimental or quasi-experimental methods.
Days Absent From School

A solid body of research has identified harmful consequences associated with decreased school attendance (Gottfried). Students who are excessively absent are exposed to less classroom instruction and performance declines on exams in the same year (Chen & Stevenson, 1995; Nichols, 2003). Consistently low attendance over several years in the early grades is associated with later problems of non-promotion and dropping out (Neild & Balfanz, 2006). Poor attendance is also associated with increased alienation among peers (Gottfried, Finn, Johnson), harmful behaviors including illegal drug use (Wang, Blomberg & Li, 2005), and greater unemployment (Alexander, et.al, 1997).

Between 1994 and 2011, NAEP asked two almost identical student-reported questions about the frequency of their days absent from school:

- For 2002-2011: How many days were you absent from school last month?
  Responses: None, 1-2 days, 3-4 days, 5-10 days, More than 10 days
- For 1994-2000: How many days of school did you miss last month?
  Responses: None, 1-2 days, 3-4 days, 5-10 days, More than 10 days

We take both questions as equivalent with the slight difference in wording not meaningfully different.

<table>
<thead>
<tr>
<th>Table 1a. Percentages of students by days absent from school the last month at grades 4, 8 and 12: Reading 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days absent</td>
</tr>
<tr>
<td>Percentage</td>
</tr>
<tr>
<td>Grade 4</td>
</tr>
<tr>
<td>Grade 8</td>
</tr>
<tr>
<td>Grade 12</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
1/ Grade 12 is 2009

Looking at days absent across all students nationally in 2011, (Table 1a), perfect attendance for the month is highest at grade 4 and declines successively at grades 8 and 12. However, the proportions of students with higher rates of absenteeism (3 or more days a month or the equivalent of 5 weeks a year) remain similar across the grades:

- Perfect attendance declines over the grades from 51 percent of the students at grade 4 to 38 percent at grade 12.
- However, across all grades, about one-fifth of the students missed at least 3 days a month (3-4 or 5 or more days a month). Specifically, the higher rates of absenteeism of at least 3 days a month are 19 percent at grade 4, 19 percent at
grade 8 and 23 percent at grade 12, a difference of only about 4 percent of the students.

**Table 1b. Changes in the percentages of students by days absent from school the last month, grades 4, 8 and 12: Reading 1994-2011**

<table>
<thead>
<tr>
<th>Grade 4</th>
<th>None</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5 or more days</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Grade 8</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>-3</td>
</tr>
<tr>
<td>Grade 12 1/</td>
<td>4</td>
<td>2</td>
<td>-2</td>
<td>-2</td>
</tr>
</tbody>
</table>

*Source: NAEP Data Explorer 1/ Grade 12 is 1994-2009*

The trends in Table 1b suggest that the frequency of monthly days absent has changed very little between 1994 and 2011.

- The maximum percentage-point change in the proportion of students absent for different amounts of days between 1994 and 2009 is an increase of only 4 percentage points for students with no days absent in the prior month at grade 12. Most other changes are 2 percentage points or less.

**Table 1c. Percentages of students at NAEP reading achievement levels by days absent from school in the last month for grades 4, 8 and 12: 2011**

<table>
<thead>
<tr>
<th>NAEP Reading Achievement Levels</th>
<th>None</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5 or more days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>below Basic</td>
<td>45*</td>
<td>30</td>
<td>14*</td>
<td>11*</td>
</tr>
<tr>
<td>at Basic</td>
<td>51*</td>
<td>30</td>
<td>12*</td>
<td>6*</td>
</tr>
<tr>
<td>at Proficient</td>
<td>55*</td>
<td>30</td>
<td>10*</td>
<td>5*</td>
</tr>
<tr>
<td>at Advanced</td>
<td>58</td>
<td>29</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Grade 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>below Basic</td>
<td>39*</td>
<td>33</td>
<td>17*</td>
<td>11*</td>
</tr>
<tr>
<td>at Basic</td>
<td>45*</td>
<td>36</td>
<td>13*</td>
<td>6*</td>
</tr>
<tr>
<td>at Proficient</td>
<td>50*</td>
<td>36</td>
<td>10*</td>
<td>4*</td>
</tr>
<tr>
<td>at Advanced</td>
<td>56</td>
<td>34</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Grade 12 1/</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>below Basic</td>
<td>34*</td>
<td>37*</td>
<td>17*</td>
<td>12*</td>
</tr>
<tr>
<td>at Basic</td>
<td>38*</td>
<td>39</td>
<td>15*</td>
<td>7*</td>
</tr>
<tr>
<td>at Proficient</td>
<td>41</td>
<td>41</td>
<td>13*</td>
<td>5</td>
</tr>
<tr>
<td>at Advanced</td>
<td>45</td>
<td>41</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

*Significantly different from Advanced at the .05 level*
Table 1c shows that at grades 4, 8 and 12 students who score at lower NAEP achievement levels are more likely to have a greater number of monthly days absent.

- The proportion of below-Basic students having perfect attendance the prior month was 14 percentage points lower than for Advanced students at grade 4, 17 percentage points at grade 8, and 21 percentage points at grade 12.
- Conversely, the proportion of below-Basic students having 3 or more days absent (or 5 weeks a year) was 12 percentage points greater than for Advanced students at grade 4, 18 percentage points greater at grade 8, and 15 percentage points greater at grade 12.

Given the strong association between student achievement and absenteeism, it is sensible for schools to focus on improving the attendance of lower-achieving students with high absenteeism rates as part of their efforts to boost academic achievement. Certainly, if children are absent from school, they may well find it more difficult to learn what is taught in their classes.

The NAEP data show, however, that there was little or no change in the percentage of students absent 3 or more days between 1994 and 2011. Despite some improvement at grade 12, about the same proportion—around one-fifth—has had such high absenteeism rates for almost two decades. One possible cause of the failure to make substantial improvement may be that few states track absenteeism in their public reporting of school data. A recent study found only six states that do so (Balfanz & Byrnes, 2012). The old adage “what gets measured gets done” seems to hold here. Tracking excessive absenteeism could be an important step in leading schools to focus on the problem.

Table 1d shows that the overall pattern of public and private school students’ absenteeism is similar, although private school students are slightly more likely to have perfect attendance the prior month at both grades 4 and 8.
The proportion of private school students with no days absent the prior month is 4-5 percentage points greater at grades 4 and 8.

Racial/ethnic differences by number of days absent break out into three groups (Table 1e).

- American Indian/Alaskan Native students have the highest percentages falling at the upper end of the monthly absenteeism range (combined 3 or more days per month) at each grade. To illustrate, at grade 8, 31 percent of American Indian students were absent 3 or more days the prior month, 13 percentage points higher than the comparable white student percentage.
- Asians at each grade have the lowest rates of absenteeism. For example, at grade 8, 62 percent were not absent any days the prior month compared to only 45 percent of whites; and only 11 percent of Asians were absent 3 or more days the prior month compared with 18 percent for Whites.
- Whites, Blacks and Hispanic students fall in the mid-range on days absent the prior month, with the White absenteeism rate of 3 or more days somewhat lower than for Blacks and Hispanics.

### Table 1e. Percentages of students by race/ethnicity for days absent from school the last month at grades 4, 8 and 12: Reading 2011

<table>
<thead>
<tr>
<th>Race/ethnicity</th>
<th>None</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5 or more days</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gr 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>50</td>
<td>31</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Black</td>
<td>49</td>
<td>28*</td>
<td>13*</td>
<td>9*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>50</td>
<td>29*</td>
<td>13*</td>
<td>8*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>64*</td>
<td>23*</td>
<td>8*</td>
<td>5*</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>39*</td>
<td>31</td>
<td>17*</td>
<td>12*</td>
</tr>
<tr>
<td><strong>Gr 8</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>45</td>
<td>37</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Black</td>
<td>45</td>
<td>32*</td>
<td>15*</td>
<td>7*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>44</td>
<td>35*</td>
<td>14*</td>
<td>7*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>62*</td>
<td>27*</td>
<td>7*</td>
<td>4*</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>34*</td>
<td>35</td>
<td>19*</td>
<td>12*</td>
</tr>
<tr>
<td><strong>Gr 12 1/</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>36</td>
<td>40</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Black</td>
<td>39*</td>
<td>38*</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Hispanic</td>
<td>38</td>
<td>39</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>50*</td>
<td>33*</td>
<td>11*</td>
<td>5*</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>30</td>
<td>36</td>
<td>23*</td>
<td>11</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
1/ Grade 12 is 2009.
*Significantly different from White at the .05 level.
Table 1f. Percentages of school-lunch eligible and ineligible students by days absent from school the last month at grades 4 and 8: Reading 2011

<table>
<thead>
<tr>
<th>School-lunch eligibility</th>
<th>None</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5 or more days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>47</td>
<td>30</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Ineligible</td>
<td>54*</td>
<td>30</td>
<td>10*</td>
<td>6*</td>
</tr>
<tr>
<td>Grade 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>42</td>
<td>35</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Ineligible</td>
<td>48*</td>
<td>36*</td>
<td>11*</td>
<td>5*</td>
</tr>
<tr>
<td>Grade 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>36</td>
<td>39</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Ineligible</td>
<td>38*</td>
<td>40</td>
<td>15*</td>
<td>7*</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significantly different from Eligible at the .05 level.

School lunch eligible students have somewhat higher rates of school absenteeism, especially at grades 4 and 8 (Table 1f).

- At grades 4 and 8, school-lunch eligible students are 7 percentage points higher in the proportion of students absent 3 or more days a month compared with school-lunch ineligible students.
- At grade 12, school-lunch eligible students are only 3 percentage points greater in the proportion absent 3 or more days a month compared with school-lunch ineligible students.

2. Time Spent in Mathematics and Reading Instruction

The research literature generally finds that greater time spent on instruction improves learning, especially for at-risk students, but instructional time must be used effectively (Silva, 2007). This report focuses on the quantity of instructional time, but potentially other reports could examine NAEP background data on instructional content and practice that represent factors affecting the quality of instructional time.

The unique NAEP historical data cover changes in instructional time during enactment of two major U.S. national educational policies:

- Systemic reform passed under President Bill Clinton in the 1994 Improve Americas Schools Act required all states receiving federal Title I funds to develop challenging content standards in mathematics and language arts and report adequate yearly progress by school.
- The 2001 No Child Left Behind Act ushered in a new period of grade-by-grade assessments and required schools to achieve proficiency of all students in mathematics and reading by 2014.
To explore instructional time in mathematics and reading and how it has changed overtime, this section examines the following questions:

- What is the current distribution of instructional time students spend in mathematics and reading, two subjects covered under federal annual testing requirements?
- Has instructional time in these subjects increased since the 1990’s and, if so, how do the increases compare after passage of the 1994 Improving America’s Schools Act with changes since enactment of the 2002 No Child Left Behind Act?
- How does instructional time in the mathematics and reading compare for different types of students, including by NAEP achievement levels, public or private school enrollment, race/ethnicity and school lunch eligibility?

The following section 3 will explore whether instructional time spent on music and the visual arts, which are subjects not covered under federal testing and accountability provisions, has declined since the mid 1990’s?

**Mathematics Instructional Time**

Because mathematics is a subject learned primarily through formal instruction, unlike reading which has a significant independent home learning component, the time spent learning mathematics at school is particularly important for students’ opportunity-to-learn mathematics content.

However, analyzing the NAEP data to describe how time spent on mathematics has changed since the mid-1990s is complicated by shifts in the wording of the questions teachers are asked about instructional time in mathematics across years. These wording shifts limit the ability to make comparisons of instructional time in mathematics over different time periods. For example, the wording of the grade 4 question about mathematics instructional time has changed three times since 1996:

- 2005-2011 grade 4 question: “How many hours of mathematics instruction do your students receive in a typical week? (teacher-reported): Responses: Less than 3 hours, 3-4.9 hours, 5-6.9 hours, 7 hours or more”
- 2003 grade 4 question: “About how much time in total do you spend with this class on mathematics instruction in a typical week? (teacher-reported) Responses: Less than 7 hours, 7-9.9 hours, 10-12.9 hours, 13 hours or more
- 1996-2000 grade 4 question: How much time do you spend each week on mathematics instruction with this class? (teacher-reported) Responses: 2.5 hours or less, More than 2.5 hours but less than 4, 4 hours or more
Because the 1996-2000 intervals don’t match those for 2005-2011, the change in instructional time on mathematics for each period will be computed separately. This report also elected to discard the 2003 question because its starting interval is a very high 7 hours a week and is not comparable with the intervals in the other questions.

The grade 8 wording of the question about instructional time spent on mathematics is similar to grade 4 and is treated the same way. There is no NAEP teacher questionnaire at grade 12 so that information on instructional time is unavailable.

| Table 2a. Percentages of students by weekly hours of mathematics instruction for grades 4 and 8, 2011 |
|-------------------------------------------------|-------------------------------------------------|
| Grade 4                                          | Grade 8                                          |
| Less than 5 hrs                                  | Less than 5 hrs                                 |
| 5.0-6.9 hrs                                      | 5.0-6.9 hrs                                     |
| 7 or more hrs                                    | 7 or more hrs                                   |
| 12                                               | 63                                               |
| 59                                               | 28                                               |
| 29                                               | 9                                                |

Source: NAEP Data Explorer

Teachers of mathematics report that students spend significantly more hours per week on mathematics at grade 4 than at grade 8 (Table 2a):

- A majority of grade 4 students spend 5-6.9 hours in math a week, but at grade 8 more than half the students spend less than 5 hours in mathematics a week.
- Moreover, the proportion of students receiving 7 or more hours of math instruction is only one-third as great at grade 8 than at grade 4.

These data on instructional time raise the possibility of increasing the time spent on mathematics instruction at grade 8 as a means to strengthen middle school mathematics. As the National Mathematics Panel (2008) has noted, math in the middle grades is particularly important in building a foundation in fractions and other concepts to prepare students for algebra in high school.

| Table 2b. Changes in the percentages of students by weekly hours of mathematics instruction for grades 4 and 8, 2005-2011 |
|-------------------------------------------------|-------------------------------------------------|
| Grade 4                                          | Grade 8                                          |
| Less than 5 hrs                                  | Less than 5 hrs                                 |
| 5.0-6.9 hrs                                      | 5.0-6.9 hrs                                     |
| 7 or more hrs                                    | 7 or more hrs                                   |
| -9*                                              | -6*                                              |
| -6*                                              | 14*                                              |
| -6*                                              | -6*                                              |
| 4*                                               | 2*                                               |

Source: NAEP Data Explorer

*Significant 2005-2011 difference at the .05 level
Table 2c. Changes in the percentages of students by weekly hours of mathematics instruction, grades 4 and 8: 1996-2000

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 4 hrs</td>
<td>4 or more hrs</td>
<td>Less than 4 hrs</td>
</tr>
<tr>
<td>-6*</td>
<td>6*</td>
<td>-18*</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant 1996-2000 difference at the .05 level

As described above, the NAEP assessments over the last 15 years have used different respondent time intervals when asking teachers about mathematics instructional time. This necessitates displaying separately the historical responses to the questions for 2005-2011 and 1996-2000. Also as noted above, the 2003 assessment question was not compatible with either period and was discarded.

Although the time intervals cannot be directly compared, the changes indicate that time spent on mathematics has increased within both periods, although differentially by grade (Tables 2b and 2c):

- A significant increase in weekly mathematics time occurred during a 4-year NAEP assessment period 1996-2000 following the 1994 Improving America’s Schools Act. The increases were especially large at grade 8 where 18 percent of the students moved from less than 4 hours a week in mathematics to 4 or more hours a week.
- The increase in weekly mathematics time continued for grade 4 between 2005-2011 with a 14-percentage point increase in the proportion of students at 7 hours of instruction or more per week. The 2005-2011 increase in mathematics time at grade 8 was smaller than at grade 4 and primarily occurred from a shift of 6 percent out of the less than 5 hours a week of instruction.

It is noteworthy that the NAEP trends indicate that increases in weekly time spent on mathematics was already occurring during the 1990’s following enactment of the Improving Americas Schools Act. Increases in weekly time spent on mathematics continued during the period covering No Child Left Behind.
Table 2d. Percentages of students by weekly hours of mathematics instruction at NAEP mathematics achievement levels for grades 4 and 8, 2011

<table>
<thead>
<tr>
<th>NAEP achievement levels</th>
<th>Less than 5 hours</th>
<th>5-6.9 hours</th>
<th>7 or more hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>below Basic</td>
<td>12</td>
<td>56*</td>
<td>31*</td>
</tr>
<tr>
<td>at Basic</td>
<td>12</td>
<td>58*</td>
<td>30*</td>
</tr>
<tr>
<td>at Proficient</td>
<td>12</td>
<td>60*</td>
<td>28</td>
</tr>
<tr>
<td>at Advanced</td>
<td>12</td>
<td>63</td>
<td>25</td>
</tr>
<tr>
<td>Grade 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>below Basic</td>
<td>55*</td>
<td>33*</td>
<td>13*</td>
</tr>
<tr>
<td>at Basic</td>
<td>63*</td>
<td>28*</td>
<td>8*</td>
</tr>
<tr>
<td>at Proficient</td>
<td>68*</td>
<td>25*</td>
<td>7</td>
</tr>
<tr>
<td>at Advanced</td>
<td>73</td>
<td>21</td>
<td>5</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Advanced at the .05 level

Table 2d examines how weekly time spent in mathematics instruction compares for students at the four different NAEP achievement levels. These data portray opportunity-to-learn mathematics in terms of instructional time and should not be interpreted as causally relating instructional time to learning. Indeed, instructional time may reflect achievement rather than the reverse, if lower performing students are given additional mathematics instructional time.

Overall, the distribution of instructional time in mathematics is compensatory with respect to students’ mathematics achievement on NAEP, but a significant proportion of the lowest achievers still receive less than an hour of daily instruction in mathematics, particularly at grade 8.

- Students at the Basic achievement level or below are 6 to 8 percentage points more likely to receive 7 or more hours of math instruction per week at grades 4 and 8 than students at Advanced.
- Nevertheless, at grade 4, 12 percent of below-Basic students receive less than an average of an hour of math instruction a day, and at grade 8 the percentage receiving less than an hour of math a day increases to more than half of the students below Basic.

While the overall distribution of weekly instructional time is compensatory with respect to student achievement, these data suggest school systems should consider whether the many lowest achieving students who now receive less than an hour a day of mathematics instruction would benefit from greater exposure to mathematics teaching.
Table 2e. Public and private school percentages of students by weekly hours of mathematics instruction at grades 4 and 8: 2011

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th></th>
<th></th>
<th>Grade 8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 5 hrs</td>
<td>5.0-6.9 hrs</td>
<td>7 or more hrs</td>
<td>Less than 5 hrs</td>
<td>5.0-6.9 hrs</td>
<td>7 or more hrs</td>
</tr>
<tr>
<td>Public</td>
<td>10*</td>
<td>59</td>
<td>31*</td>
<td>62*</td>
<td>29*</td>
<td>9*</td>
</tr>
<tr>
<td>Private</td>
<td>35</td>
<td>58</td>
<td>7</td>
<td>77</td>
<td>21</td>
<td>2</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Private at the .05 level

Compared with students in private schools, students in public schools receive significantly greater amounts of weekly mathematics instructional time at grade 4, but the difference is much less at grade 8 (Table 2e).

- At grade 4, the percentage of public students spending 7 or more hours in mathematics a week is 24 percentage points more than for private school students.
- By grade 8, the public school percentage of students spending 7 or more hours weekly in mathematics instruction falls to 9 percent and the differential over private schools is down to 7 percentage points.

TABLE 2f. Percentages of students by weekly hours of mathematics instruction and race/ethnicity at grades 4 and 8: 2011

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th></th>
<th></th>
<th>Grade 8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 5 hrs</td>
<td>5-6.9 hrs</td>
<td>7 hrs or more</td>
<td>Less than 5 hrs</td>
<td>5-6.9 hrs</td>
<td>7 hrs or more</td>
</tr>
<tr>
<td>White</td>
<td>13</td>
<td>62</td>
<td>25</td>
<td>70</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>Black</td>
<td>10*</td>
<td>53*</td>
<td>37*</td>
<td>48*</td>
<td>37*</td>
<td>15*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>11*</td>
<td>54*</td>
<td>36*</td>
<td>55*</td>
<td>33*</td>
<td>13*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>10*</td>
<td>62</td>
<td>28</td>
<td>68</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td>Amer Indian/Alaska Native</td>
<td>12</td>
<td>52*</td>
<td>36*</td>
<td>52*</td>
<td>35*</td>
<td>13*</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from White at the .05 level

The NAEP background data indicate that students’ weekly instructional time on mathematics with respect to race/ethnicity is greater for the lower achieving Black, Hispanic and American Indian students than for Whites or Asians (Table 2f).

- At grade 4, slightly more than 35 percent of Black, Hispanic and American Indian/Alaskan Native students spend 7 hours or more per week in mathematics.
The corresponding percentages for White and Asian students are about 10 percentage points lower.

- While grade 8 students weekly time in mathematics displays the same general compensatory pattern as for grade 4, the Black, Hispanic and Native American advantage in mathematics instructional time is particularly large at the lowest not the highest amount of weekly instructional time. Among Whites, 70 percent spend less than 5 hours (1 hour per day) in math compared with only 48 percent of Blacks, 55 percent of Hispanics and 52 percent of American Indian/Alaskan Natives.

<table>
<thead>
<tr>
<th>Table 2g. Percentages of students by weekly hours of mathematics instruction and school-lunch eligibility at grades 4 and 8: 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4</td>
</tr>
<tr>
<td>Less than 5 hrs</td>
</tr>
<tr>
<td>Less than 5 hrs</td>
</tr>
<tr>
<td>Ineligible</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Ineligible at the .05 level

School-lunch eligible students are a significantly lower performing group on the NAEP mathematics assessment at grades 4 and 8 than school-lunch ineligible students (Table 2g). Similar to the findings for race/ethnicity, the amount of time spent in mathematics is compensatory with respect to school-lunch eligibility.

- Grade 4 school lunch eligible students have an 11-percentage point greater enrollment in instruction of 7 hours or more per week than school-lunch ineligibles.
- At grade 8 the school-lunch eligible advantage is only 6 percentage points, but school lunch eligible students are also 16 percentage points less likely to spend 5 or less hours per week in mathematics.

Reading Instructional Time

The NAEP background questions on students’ time spent in classroom reading instruction cover two major instructional components. One component is instruction in formal reading related activities, such as phonics and reading comprehension. A second component is grammar and writing. Language arts is the more common and inclusive terminology for covering reading, grammar and writing.

Unfortunately, the NAEP historical questions in this area have at different times used both the terms reading and language arts, which causes difficulties in making
meaningful comparisons over time. To illustrate the difficulties, three questions use different wording and ask different information about grade 4 instructional time in the reading area since the mid-1990s.

- **2005-2011 question:** About how much time in total do you spend with this class on language arts instruction in a typical week? Language arts refers to reading, writing, literature, and related topics. (teacher-reported)
  Values: Less than 3 hours, 3-4.9 hours, 5-6.9 hours, 7-9.9 hours, 10 hours or more.

- **2002-2003 question:** About how much time in total do you spend with your class on language arts instruction in a typical week? (teacher-reported)
  Values: Less than 7 hours, 7-9.9 hours, 10-12.9 hours, 13 hours or more

- **1998-2000 question:** About how much time do you spend with this class for reading instruction on a typical day? (teacher-reported)
  Values: Less than 30 minutes, 30-44 minutes, 45-59 minutes, 60-90 minutes, More than 90 minutes

The 2005-2011 grade 4 question covers the broader instructional category of language arts rather than reading and the intervals range from under 3 to 10 or more hours. By comparison, the 2002-2003 question still focuses on language arts, but the least time interval is less than 7 hours a week or an hour and 20 minutes a day. The 1998-2000 question asks about reading not language arts and shifts the intervals to minutes.

At grade 8, 2002-2011 question covers the time students spend in language arts. Unfortunately, again, the prior year questions are not comparable as the 1998 and 1994 questions ask only about reading, not language arts.

Given these differences in questions, this report for grade 4 is limited to the 2005-2011 data and for grade 8 to the 2002-2011 data on time spent in reading-language arts.

<table>
<thead>
<tr>
<th>Table 2h. Percentages of students by weekly hours of reading-language arts instruction at grades 4 and 8: 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade 4</strong></td>
</tr>
<tr>
<td>Less than 5 hrs</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
In 2011, like mathematics, students are exposed to a greater number of weekly hours in reading instruction at grade 4 than at grade 8 (Table 2h).

- Nearly half the grade 4 students spend 10 hours or more a week on reading and another 30 percent between 7 and 10 hours a week.
- By comparison, nearly half the grade 8 students spend less than 5 hours a week on reading and about one-fifth of the students spend over 7 hours.

This greater time spent on reading instruction at grade 4 may reflect the instructional time required to transition students from decoding skills to reading comprehension as well as supporting the building of foundational skills in grammar and writing.

<table>
<thead>
<tr>
<th>Table 2i. Change in the percentages of students by weekly hours of reading-language arts instruction at grades 4 and 8, 2002-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4: 2005-2011</td>
</tr>
<tr>
<td>Less than 5 hrs</td>
</tr>
<tr>
<td>2*</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference in NAEP change score at the .05 level

Because of the shifts in the wording of the NAEP questions noted above, the time period for measuring the change in the number of weekly hours devoted to reading is limited to 2005-2011 at grade 4 and 2002-2011 at grade 8 (Table 2i). Over these time periods, weekly time spent on reading held constant at grade 4, but increased somewhat at grade 8.

- At grade 4 over the period of 2005 to 2011, the distribution of weekly hours of reading instruction displayed no substantial changes.
- At grade 8 from 2002 to 2011, the proportion of students receiving less than 5 hours a week of reading instruction decreased by 10 percentage points. Much of the increase was in the next highest category (5.0-6.9 hours per week), and even with these changes, nearly half the grade 8 students still were receiving less than 5 hours per week of reading-language arts.
Table 2j. Percentages of students by weekly hours of reading-language arts instruction at NAEP reading achievement levels for grades 4 and 8, 2011

<table>
<thead>
<tr>
<th>NAEP achievement levels</th>
<th>Less than 5 hrs</th>
<th>5.0-6.9 hrs</th>
<th>7-9.9 hrs</th>
<th>10 hrs or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>below Basic</td>
<td>12*</td>
<td>12</td>
<td>27*</td>
<td>49*</td>
</tr>
<tr>
<td>at Basic</td>
<td>10</td>
<td>13</td>
<td>31*</td>
<td>47</td>
</tr>
<tr>
<td>at Proficient</td>
<td>9</td>
<td>14</td>
<td>32</td>
<td>46</td>
</tr>
<tr>
<td>at Advanced</td>
<td>9</td>
<td>14</td>
<td>34</td>
<td>44</td>
</tr>
</tbody>
</table>

Grade 8

<table>
<thead>
<tr>
<th>NAEP achievement levels</th>
<th>Less than 5 hrs</th>
<th>5.0-6.9 hrs</th>
<th>7-9.9 hrs</th>
<th>10 hrs or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>below Basic</td>
<td>39*</td>
<td>34*</td>
<td>20*</td>
<td>8*</td>
</tr>
<tr>
<td>at Basic</td>
<td>47*</td>
<td>32*</td>
<td>15</td>
<td>6*</td>
</tr>
<tr>
<td>at Proficient</td>
<td>51</td>
<td>30</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>at Advanced</td>
<td>55</td>
<td>27</td>
<td>14</td>
<td>4</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Advanced at the .05 level

Table 2j indicates that the amount of weekly reading time is compensatory in the sense that students who are at the lower NAEP achievement levels receive more reading instruction, especially at grade 8. But sizeable percentages of students who are reading at Basic or below-Basic fall into the fewest hours of weekly reading time, especially at grade 8.

- At grade 4, more than three-fourths of the students across all the achievement levels receive 7 or more hours of reading instruction a week. Time spent on reading exhibits a small compensatory pattern, as the proportion of the below-Basic reading achievement group to spending 10 hours a week in reading is 5 percentage points greater than the Advanced.

- At grade 8, reading patterns are more sharply differentiated by NAEP achievement group. The percentage of students below-Basic in reading receiving 5 or less hours of reading instruction per week is 16 percentage points less than the Advanced percentage. Also, a greater proportion of below-Basic students receive reading instruction in each of the three higher amounts of weekly time spent on reading than for Advanced students.

- However, there remain significant percentages of below-basic students receiving less than 5 hours a week of reading instruction – at grade 8 about 39 percent of the students fall into this lowest amount of weekly time spent on reading.
Table 2k. Percentages of students by weekly hours of reading-language arts instruction and public and private school attendance at grades 4 and 8: 2011

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th></th>
<th>Grade 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 5 hrs</td>
<td>5.0-6.9 hrs</td>
<td>7 - 9.9 hrs</td>
<td>10 hrs or more</td>
</tr>
<tr>
<td>Public</td>
<td>9*</td>
<td>12*</td>
<td>30*</td>
<td>49*</td>
</tr>
<tr>
<td>Private</td>
<td>15</td>
<td>27</td>
<td>35</td>
<td>22</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Private at the .05 level

Public school students at grade 4 experience higher weekly hours of time in reading than those in private school, but differences in weekly hours of reading instruction diminish at grade 8 (table 2k).

- Grade 4 public school teachers report about half the students receive 10 or more hours of reading instruction a week compared with only about one-fifth the private school students.
- At grade 8, public and private school students are equally likely to receive higher amounts of weekly reading instruction of 7 or more hours.

Table 2l. Percentages of students by weekly hours of reading-language arts instruction and race/ethnicity at grades 4 and 8: 2011

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th></th>
<th>Grade 8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 5 hrs</td>
<td>5.0-6.9 hrs</td>
<td>7 - 9.9 hrs</td>
<td>10 hrs or more</td>
</tr>
<tr>
<td>White</td>
<td>9</td>
<td>14</td>
<td>33</td>
<td>43</td>
</tr>
<tr>
<td>Black</td>
<td>11*</td>
<td>13</td>
<td>29*</td>
<td>47*</td>
</tr>
<tr>
<td>Hispanic</td>
<td>12*</td>
<td>9*</td>
<td>24*</td>
<td>56*</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>9</td>
<td>13*</td>
<td>31</td>
<td>48</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>12</td>
<td>14</td>
<td>30</td>
<td>44</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from White at the .05 level

Weekly time spent in reading instruction tends to be greater for Hispanics at grade 4 and Hispanics and Blacks at grade 8, although the pattern of the extra time differs.
between grades 4 and 8 (Table 2l).

- At grade 4, 12 percentage points more Hispanic students receive 10 hours a week in instruction on reading than Whites.
- At grade 8, Black and Hispanic students participate at higher rates in the 7-9.9 and 10 or more hours per week reading categories, but American Indians receive little time advantage over Whites. Also, about 15 percentage points fewer Black and Hispanic students spend less than 5 hours of weekly instruction in reading than Whites. American Indian students, although also lower scoring on NAEP achievement, are almost as likely as Whites to receive less than 5 hours of reading instruction.

<table>
<thead>
<tr>
<th>School lunch eligibility</th>
<th>Grade 4</th>
<th>Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 5 hrs</td>
<td>5.0-6.9 hrs</td>
</tr>
<tr>
<td>Eligible</td>
<td>11*</td>
<td>11*</td>
</tr>
<tr>
<td>Ineligible</td>
<td>8</td>
<td>13</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Ineligible at the .05 level

The federal Title I program targets low-income children to receive more and better quality instruction in reading and other core subjects. In terms of greater time spent on reading, Table 2m suggests that reading instructional time is more compensatory at grade 8 than at grade 4.

- At grade 4, low-income children eligible for school lunch are not likely to spend more time in reading instruction than other students. About one-fifth of school lunch eligibles and ineligibles receive less than 7 hours a week. School lunch eligible students do have a 3 percent greater frequency of spending 10 or more hours in reading, but an offsetting 5 percentage point fewer are spending 7-9.9 hours.
- At grade 8, however, 11 percentage points fewer school-lunch eligible students spend less than 5 hours per week in reading. Also, 9 percentage points more school-lunch eligibles spend 7 or more hours in reading than their school-lunch ineligible peers.
3. Frequency of Visual Arts and Music Instruction

Education experts and parents interested in music and art have expressed concerns that the emphasis on mathematics and reading to meet federal test-based accountability at grades 3-8 has diminished the amount of instructional time devoted to music and the visual arts. This section examines trends in grade 8 teacher reports on the weekly frequency of music and visual arts instruction to examine the evidence supporting diminished exposure between 1994 and 2008.

The literature on trends in time spent on music and art sends conflicting signals. Some surveys (Centre on Education Policy, 2007) of school systems ask about whether time spent on music and art instruction has changed since enactment of No Child Left Behind in 2002. These surveys are typically based on retrospective data and methodologically are weaker than comparing actual times reported for different years.

The likely more accurate methodology employed by NCES (Parsad & Spiegleman, 2012) was to compare survey responses from public school staff in 1999-2000 and 2009-2010 rather than rely on retrospective recollections of instructional time. The NCES study found that the percent of public schools offering some instruction in music and the visual arts had not significantly changed in the decade spanning 2000-2010, with most schools offering both subjects.

The NAEP background variables can be used to validate the Fast Response findings. Furthermore, it provides additional information about the frequency of instruction in music and the visual arts each week. That is, the Fast Response survey only asked schools whether instruction in music and the visual arts was offered and not how frequently, but the NAEP questionnaire asked about frequency each week. For example, the music question asked in 1994 and 2008

- How often does a typical eighth-grade student in your school receive instruction in each of the following subjects? (school-reported)
  Values: Every Day, 3-4 Times A Week, 1-2 Times A Week, Less than once a week, Subject not taught

### Visual Arts

<table>
<thead>
<tr>
<th>Year</th>
<th>Subject not taught</th>
<th>Less than once a week</th>
<th>1-2 times a week</th>
<th>3-4 times a week</th>
<th>Every Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>14</td>
<td>10</td>
<td>30</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>1994</td>
<td>13</td>
<td>13</td>
<td>33</td>
<td>14</td>
<td>28</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
Despite concerns over test-based accountability in reading and math diminishing the frequency of students’ exposure to the visual arts, the NAEP data indicate that the frequency of visual arts instruction per week was about the same in 2008 as in 1994 across the full range of frequencies (Table 3a).

- Using the NCES the fast-response survey criteria about whether visual arts instruction was taught at all, slightly under 15 percent of the students failed to receive any instruction in the visual arts in 1994 and 2008. This is consistent with the NCES fast-response finding that in 2010 about 17 percent of the elementary schools failed to offer visual arts instruction.

- Given that most students receive some visual arts instruction, the NAEP data go beyond the information in the Fast Response survey in that they address how often visual arts instruction is offered when it is available. Overall, the distribution of visual arts instruction across the weekly frequency levels for 2008 is quite similar to the distribution 14 years earlier. There is some evidence of a small increase in visual arts instructional time, as 47 percent of the students received at least 3 days of visual arts instruction in 2008 compared with a lesser 42 percent in 1994.

Table 3b. Percentages of public and private school students by frequency of instruction in visual arts at grade 8: 2008

<table>
<thead>
<tr>
<th>Subject not taught</th>
<th>Less than once a week</th>
<th>1-2 times a week</th>
<th>3-4 times a week</th>
<th>Every Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>13</td>
<td>10</td>
<td>26*</td>
<td>18</td>
</tr>
<tr>
<td>Private</td>
<td>17</td>
<td>10</td>
<td>70</td>
<td>--</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Private at the .05 level

Public school students are more likely than private to have frequent exposure to visual arts (Table 3b).

- Most grade 8 students in public or private school receive some visual arts instruction each week.
- About half of public school grade 8 students are exposed to visual arts 3 or more times a week, while private school students exposure is likely to be no more than only one or two times a week.
Table 3c. Percentages by racial/ethnic group and school lunch eligibility for frequency of instruction in visual arts at grade 8: 2008

<table>
<thead>
<tr>
<th>Year</th>
<th>Subject not taught</th>
<th>Less than once a week</th>
<th>1-2 times a week</th>
<th>3-4 times a week</th>
<th>Every Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/Ethnicity1/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• White</td>
<td>11</td>
<td>11</td>
<td>34</td>
<td>17</td>
<td>28</td>
</tr>
<tr>
<td>• Black</td>
<td>18</td>
<td>10</td>
<td>24</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td>• Hispanic</td>
<td>17</td>
<td>5</td>
<td>23</td>
<td>15</td>
<td>41</td>
</tr>
<tr>
<td>• Asian Pacific</td>
<td>5</td>
<td>11</td>
<td>29</td>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>School Lunch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Eligible</td>
<td>18</td>
<td>9</td>
<td>26</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>• Ineligible</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>19</td>
<td>31</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
1/Data for American Indian/Alaskan Natives are not available because of insufficient sample size.
*Significant difference from White within race/ethnicity or Ineligible within school lunch at the .05 level

With respect to students having access to any visual arts instruction, Whites and especially Asians have greater access than Blacks or Hispanics to some visual arts instruction, but among those receiving instruction, Hispanics are likely to have a greater frequency of exposure (Table 3c).

- Only 5 percent of Asian students have no exposure to the visual arts, but 18 percent of black students and 17 percent of Hispanics.
- However, Hispanics have a 13 percent higher proportion receiving daily visual arts instruction than Whites.

Notably, the proportion of school-lunch eligibles receiving no visual arts instruction is 8 percentage points higher than for students from higher-income families, who are ineligible for school lunch.
Music

Table 3d. Percentages of students by frequency of instruction in music at grade 8: 2008 and 1994

<table>
<thead>
<tr>
<th>Year</th>
<th>Subject not taught</th>
<th>Less than once a week</th>
<th>1-2 Times a Week</th>
<th>3-4 Times a Week</th>
<th>Every Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>8</td>
<td>8</td>
<td>27</td>
<td>20</td>
<td>37</td>
</tr>
<tr>
<td>1994</td>
<td>12</td>
<td>9</td>
<td>30</td>
<td>15</td>
<td>34</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer

The comparison of music instruction between 1994 and 2008, as with the visual arts, shows no reduction in the frequency of music instruction offered (Table 3d). The 8 percent of the students receiving no music instruction is similar to the NCES Fast Response Survey finding of 6 percent of the schools not offering any music in 2010. Indeed, the frequency of music offerings may have increased.

- There was an increase of 8 percentage points in the proportion of students receiving music 3 or more times a week.

Table 3e. Percentages of public and private school students by frequency of instruction in music at grade 8: 2008

<table>
<thead>
<tr>
<th></th>
<th>Subject not taught</th>
<th>Less than once a week</th>
<th>1-2 Times a Week</th>
<th>3-4 Times a Week</th>
<th>Every Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>8</td>
<td>7</td>
<td>24*</td>
<td>22</td>
<td>39</td>
</tr>
<tr>
<td>Private</td>
<td>10</td>
<td>15</td>
<td>71</td>
<td>#</td>
<td>3</td>
</tr>
</tbody>
</table>

# Rounds to zero.
Source: NAEP Data Explorer
*Significant difference from Private at the .05 level

In 2008, students in public schools were more likely than in private schools to receive instruction in music multiple times a week (Table 3e).

- Most public or private school students are exposed to some music instruction.
- Sixty-one percent of public school students but only three percent of private school students receive music instruction 3 or more days a week.
Table 3f. Percentages of students by racial/ethnic group and school-lunch eligibility for frequency of music instruction at grade 8: 2008

<table>
<thead>
<tr>
<th>Subject not taught</th>
<th>Less than once a week</th>
<th>1-2 Times a Week</th>
<th>3-4 Times a Week</th>
<th>Every Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicity 1/</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>6</td>
<td>8</td>
<td>29</td>
<td>22</td>
</tr>
<tr>
<td>Black</td>
<td>10</td>
<td>8</td>
<td>26</td>
<td>18</td>
</tr>
<tr>
<td>Hispanic</td>
<td>14</td>
<td>6</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>7</td>
<td>8</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>School lunch Eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>10</td>
<td>6</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>Ineligible</td>
<td>8</td>
<td>8</td>
<td>26</td>
<td>21</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
1/Data for American Indian/Alaskan Natives are not available because of small sample.
*Significant difference from White within race/ethnicity or Ineligible within school lunch at the .05 level

The frequency of exposure to music is largely unrelated to race/ethnicity or school lunch eligibility (Table 3f).

- Hispanic students are the only group that is markedly different than White students in their frequency of music instruction each week, although the direction of the difference shifts. Compared with White students, an 8 percentage point greater proportion of Hispanic students receive no music each week and also an 8 percentage point greater proportion of Hispanic students receive daily music instruction.

4. Expected Homework Time

Homework serves different purposes. Homework reinforces the content of classroom instruction. In addition, homework may be tailored to address specific student’s academic needs. Students who are having difficulty in a subject can use homework to review content and provide greater practice to facilitate understanding basic concepts. Students who are more academically able may benefit from homework that stretches knowledge.

Because of differences in students’ abilities, assessing the quantity of homework in terms of students’ time spent on homework and its relationship with achievement may yield counter intuitive results. That is, lower achieving students may require a greater amount of time to complete the same quantity of homework than a higher achieving student, producing a negative correlation between homework time and student
achievement. In fact, research that does a good job controlling for initial differences in student achievement finds that appropriate homework can exert a positive relationship on achievement, especially beyond the early elementary grades (Cooper et al., 2006).

The NAEP question about time spent on homework, asks the teacher rather than the student the amount of homework time expected. It is available only for mathematics homework. We expect that asking the teacher expected homework time should yield a measure of homework time that a student of average ability would take to complete the homework.

As an example, at grade 4, NAEP asks a consistent question between 1996 and 2011 that covers a broad range of daily homework time intervals:

Approximately how much mathematics homework do you assign to students in this class each day? (teacher-reported).
Responses: None, 15 minutes, 30 minutes, 45 minutes, 1 hour, More than 1 hour

At grade 8, the questions about time spent on homework are less satisfactory. The question for both 2009 and 2011 is:

Approximately how much mathematics homework do you assign to students in your mathematics class each day? (teacher-reported).
Responses: None, Less than 1 hour, About 1 hour, About 2-3 hours, More than 3 hours

Unlike the prior question that breaks out time into 15-minute segments up to an hour, the responses grade 8 question skips from none to less than 1 hour and it then extends to unlikely amounts just for mathematics of 2-3 or more hours daily. Thus, the grade 8 homework time analyses will be limited to identifying only very broad differences in homework time of none, less than an hour, or one hour or more.

<table>
<thead>
<tr>
<th>Table 4a. Percentages of students by mathematics homework time teacher assigns per day at grades 4 and 8, 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4</td>
</tr>
<tr>
<td>None, 15 min, 30 min, 45 or more min</td>
</tr>
<tr>
<td>4, 48, 43, 5</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
As expected, the amount of daily homework time in mathematics increases from grade 4 to grade 8 (Table 4a).

- Grade 4 mathematics homework typically centers around 15 to 30 minutes per day.
- The grade 8 question intervals, as noted, are not discriminating below an hour, but 17 percent of the grade 8 students have at least an hour of expected mathematics homework. This compares with only 5 percent of the grade 4 students having 45 minutes or more of daily mathematics homework.

<table>
<thead>
<tr>
<th>Table 4b. Change in percentages of students by mathematics homework time teacher assigns per day at grades 4 and 8, 1996-2011</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade 4</strong></td>
</tr>
<tr>
<td>None</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

Given the pressures to improve student outcomes generated by systematic reform and No Child Left Behind, increases in teacher expected homework time might be anticipated between 1996 and 2011. Interestingly, expected homework time does not appear to have increased at grade 4 although increases did occur at grade 8 (Table 4b).

- At grade 4, there was no change between 1996 and 2011 in the percentage of students at the upper end of 45 minutes or more expected daily homework and there was only a modest increase of 5 percentage points of those expected to spend between 15 minutes and 30 minutes of mathematics homework daily.
- At grade 8, there was an increase of 13 percentage points in the proportion of students with expected homework time over an hour a day between 1996 and 2011.

<table>
<thead>
<tr>
<th>Table 4c. Percentages of students at NAEP achievement levels by mathematics homework time teacher assigns per day at grades 4 and 8, 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NAEP Achievement Levels</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>below Basic</td>
</tr>
<tr>
<td>at Basic</td>
</tr>
<tr>
<td>at Proficient</td>
</tr>
<tr>
<td>at Advanced</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
*Significant difference from Advanced at the .05 level
Table 4c indicates that the expected amount of homework teachers assign does not differ much across student achievement levels. This suggests that teachers have the same expectations in terms of the amount of homework time it would take to complete typical assignments regardless of the performance levels of students in their class. It is unclear how this translates into actual assigned homework. For example, a teacher with high-performing students may expect them to do a mathematics problem in less time than a teacher with low-performing students, and hence assign the higher performers a greater quantity or more difficult problems to complete within the same time.

| Table 4d. Percentages of public and private students by mathematics homework time teacher assigns per day at grades 4 and 8, 2011 |
|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|
| Grade 4                          | Grade 8                          |
| None  | 15 min  | 30 min  | 45 or more min  | None  | Less than 1 hour  | 1 hour or more  |
| Public | 3       | 48      | 43      | 5       | 2       | 82      | 16      |
| Private | 5       | 41      | 48      | 6       | 1       | 74      | 25      |

Source: NAEP Data Explorer

At both grade 4 and 8 private schools expect students to do somewhat more homework in mathematics each day than public schools (Table 4d).

- At grade 4, about 6 percentage points more private than public school students spend at least 30 minutes on homework.
- At grade 8, 9 percentage points more private school students are expected to take at least an hour to complete mathematics homework.

Table 4e shows that Black and Hispanic students are assigned more time on mathematics homework each day than Whites but not Asian students. Although American Indian students, like Blacks and Hispanics score below the NAEP average on mathematics, their expected homework time is similar to Whites and not as high as Blacks, Hispanics or Asians.

- At grade 4, 42 percent of White students are assigned at least 30 minutes of homework a day, which is considerably less than the 56 percent of Blacks, 60 percent of Hispanics and 54 percent of the Asian students. However, the 41 percent of American Indian students assigned 30 minutes or more daily on mathematics homework is no greater than on average the percentage for higher-achieving Whites.

- At grade 8, compared with White students, 6 percentage points more Black students are assigned an hour or more on homework, 8 percentage points more
Hispanics and 10 percentage points more Asians. American Indians again are more similar to Whites in the percentage of students expected to spending at least an hour daily on mathematics homework.

Low-income students (eligible for school lunch) are expected to devote greater time on mathematics homework at grade 4 but expectations are more equal at grade 8.

<table>
<thead>
<tr>
<th>Race/ethnicity</th>
<th>Grade 4</th>
<th></th>
<th></th>
<th>Grade 8</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None</td>
<td>15 min</td>
<td>30 min</td>
<td>45 or</td>
<td>more min</td>
<td>None</td>
</tr>
<tr>
<td>White</td>
<td>4</td>
<td>54</td>
<td>39</td>
<td>3</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Black</td>
<td>3*</td>
<td>41*</td>
<td>48</td>
<td>8*</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2*</td>
<td>37*</td>
<td>51</td>
<td>9*</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>2*</td>
<td>44*</td>
<td>47</td>
<td>7*</td>
<td></td>
<td>1*</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>8*</td>
<td>50</td>
<td>37</td>
<td>4</td>
<td></td>
<td>6*</td>
</tr>
<tr>
<td>School Lunch Eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>4</td>
<td>43*</td>
<td>46*</td>
<td>7*</td>
<td></td>
<td>3*</td>
</tr>
<tr>
<td>Ineligible</td>
<td>3</td>
<td>53*</td>
<td>41</td>
<td>4</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Source: NAEP Data Explorer
1/Data for American Indian/Alaskan Natives are not available because of small sample size.
*Significant difference from White within race/ethnicity or ineligible within school lunch at the .05 level

V. Recommendations for Further Analyses and Improvements in the Data

This report has examined at the national level three important aspects of student time and learning: absenteeism; instructional time in mathematics, reading, music and visual art; and the time expected to complete homework. The following recommendations are proposed for the Governing Board to consider extending these analyses in three areas:

1. Two additional analyses on time for learning are proposed for NAGB consideration.
   a. Track instructional time for additional academic subjects to be presented in separate reports. Science would be a high priority for an additional report because of wide public interest in this field and the federal requirement for annual state science assessments.
b. **Disaggregate and display selected key indicators of time use for individual states and participating urban districts.** A publication of key time indicators for states and urban districts in NAEP would be of great interest to enable these jurisdictions to compare their use of time for learning against other similar jurisdictions. See recommendation 4b of the Expert Panel Report to NAGB on Strengthening NAEP Background Questions (Smith et al., 2012).

2. **Establish a series of NAEP Portraits of American Education.** The series could be initiated by assembling all of the time and learning analyses along with the exploratory study of NAEP data on charter schools, which has also been prepared for the Governing Board. The series could then be extended to cover other sets of background variables such as those for teachers, curriculum, technology, private schools and students’ out-of-school learning.

3. Our experiences in preparing this report have prompted several methodological suggestions to improve future analyses of NAEP background information:

- **Consider adding questions to the school questionnaire on the length of the school day and school year so that a more comprehensive picture of student time for learning in school is available.** In particular, there is considerable interest in learning more about the length of the school day. For example, the National Center on Time and Learning (2012) in partnership with the Ford Foundation has announced a five-state time collaborative to support expanded learning time in schools. Yet, currently basic data are not available on the distribution of the length of the school day across American schools.

- **Consider adding to the teacher or school survey a question on whether academic performance is used to assign students to more instructional time in a particular subject.** This question would address an important possible reason for instructional time differences.

- **Consider adding questions to the student questionnaire on time spent in out-of-school learning situations**—including formal classes or tutoring, visits to museums and historic sites, cultural programs, and online activities related to learning. While currently after-school learning situations consist mostly of formal tutoring and other organized activities, the rapid expansion of online learning will make it increasingly important to have a good picture of how students spend their time learning outside the regular school day. Indeed, educators are widely discussing the flipped classroom—primarily delivering instruction electronically and doing additional tutoring or homework activities in school. NAEP could provide important information on these out-of-school learning trends.

- **When considering the wording of the time-for-learning questions, examine consistency with the wording of similar questions on the major international...**
assessments of PIRLS, PISA and TIMSS. Also, consider consistency in wording over time, as proposed in recommendation 1d of the Expert Panel Report on Background Questions (Smith et al., 2012). It might also be desirable to have consistent wording of the questions in different subjects. For example, the questions about music and the visual arts have asked about the frequency of instruction each week, but not about the amount of instructional time, which is asked—with different time intervals—about instruction in mathematics and reading. Consistent wording would improve time-use comparisons across subjects. However, some of these goals may be conflicting, and careful judgments should be made about which type of consistency is most important.

- Further improve the powerful online tools for NAEP data analysis. (Recommendation 4e of the Expert Panel Report.) In particular, software should be extended to build in the capability for multivariate analysis.
References

Nichols, J. (2003). Prediction indicators for students failing the state of Indiana high school graduation exam. Preventing School Failure, 47, 112-120.
Silva, E. (2007). *On the clock: rethinking the way schools use time*. Education Sector Reports. Available online November 2012:  

http://www.nagb.org/content/nagb/assets/documents/publications/expert-panel-naep-bq-report.pdf

U.S. Congress, HR6, 103rd Congress 1994). *Improving America’s schools act*. Available December 2012 online:  
http://www.gpo.gov/fdsys/pkg/BILLS-103hr6enr/pdf/BILLS-103hr6enr.pdf

U.S. Congress, PL 107-110, 107th Congress. *No child left behind act of 2001*. Available online December 2012:  

<table>
<thead>
<tr>
<th>Sub-Group</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5+ days</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ days</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table A-1. Percentages of students for reading, grade 4, 8, and 12 by days absent from school in the last month students nationally and by achievement levels, public/private school, race/ethnicity and school-lunch eligibility year 2011, 2009 and 1994 (Source NAEP Data Explorer)

### Achievement levels - discrete

<table>
<thead>
<tr>
<th>Sub-Group</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5+ days</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ days</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Public/Private

<table>
<thead>
<tr>
<th>Sub-Group</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5+ days</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ days</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Race/Ethnicity

<table>
<thead>
<tr>
<th>Sub-Group</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5+ days</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ days</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### School Lunch

<table>
<thead>
<tr>
<th>Sub-Group</th>
<th>1-2 days</th>
<th>3-4 days</th>
<th>5+ days</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ days</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table A-2a. Percentages of students by weekly hours of mathematics instruction at grades 4, 8, and 12 by students nationally and by achievement levels, public/private school, race/ethnicity and school-lunch eligibility: 2011, 2009 and 1994 (Source NAEP Data Explorer)

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Sub-group</th>
<th>Less than 5 hours (collapsed)</th>
<th>5-6.9 hour</th>
<th>7 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard error</td>
<td>Percentage</td>
<td>Standard error</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>12* (0.4)</td>
<td>59* (0.6)</td>
<td>29* (0.5)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>21 (0.3)</td>
<td>65 (0.5)</td>
<td>15 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Grade 8</td>
<td>National</td>
<td>63* (0.7)</td>
<td>28* (0.6)</td>
<td>9* (0.4)</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>National</td>
<td>69 (0.6)</td>
<td>24 (0.6)</td>
<td>7 (0.3)</td>
<td></td>
</tr>
</tbody>
</table>

*Significant difference from 2005 at .05 level.

#### Achievement Levels

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Sub-group</th>
<th>Less than 5 hours (collapsed)</th>
<th>5-6.9 hour</th>
<th>7 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard error</td>
<td>Percentage</td>
<td>Standard error</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>below Basic</td>
<td>12 (0.7)</td>
<td>56* (0.9)</td>
<td>31* (0.7)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Basic</td>
<td>12 (0.5)</td>
<td>58* (0.7)</td>
<td>36* (0.6)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>National</td>
<td>at Proficient</td>
<td>12 (0.5)</td>
<td>60* (0.7)</td>
<td>28 (0.6)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Advanced</td>
<td>12 (0.8)</td>
<td>63 (1.2)</td>
<td>25 (1.2)</td>
</tr>
</tbody>
</table>

*Significant difference from 2005 at .05 level.

#### Public/Private

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Sub-group</th>
<th>Less than 5 hours (collapsed)</th>
<th>5-6.9 hour</th>
<th>7 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard error</td>
<td>Percentage</td>
<td>Standard error</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>Public</td>
<td>10* (0.4)</td>
<td>59 (0.6)</td>
<td>31* (0.5)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Private</td>
<td>35 (2.5)</td>
<td>56 (2.4)</td>
<td>7 (1.1)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>National</td>
<td>Public</td>
<td>62* (0.7)</td>
<td>29* (0.6)</td>
<td>9* (0.5)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Private</td>
<td>77 (1.9)</td>
<td>21 (1.8)</td>
<td>2 (0.6)</td>
</tr>
</tbody>
</table>

*Significant difference from Private at .05 level.

#### Race/Ethnicity

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Sub-group</th>
<th>Less than 5 hours (collapsed)</th>
<th>5-6.9 hour</th>
<th>7 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard error</td>
<td>Percentage</td>
<td>Standard error</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>White</td>
<td>13 (0.6)</td>
<td>62 (0.7)</td>
<td>25 (0.4)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Black</td>
<td>10* (0.6)</td>
<td>53* (0.9)</td>
<td>37* (0.9)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>National</td>
<td>Hispanic</td>
<td>11* (0.9)</td>
<td>54* (1.3)</td>
<td>36* (1.3)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Asian/Pacific Islander</td>
<td>10* (0.8)</td>
<td>62 (1.8)</td>
<td>28 (1.9)</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>American Indian/Alaska Native</td>
<td>12 (1.3)</td>
<td>52* (1.6)</td>
<td>36* (1.4)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>White</td>
<td>70 (0.6)</td>
<td>24 (0.6)</td>
<td>6 (0.3)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>National</td>
<td>Black</td>
<td>48* (1.2)</td>
<td>37* (1.6)</td>
<td>15* (0.9)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Hispanic</td>
<td>55* (1.4)</td>
<td>33* (1.3)</td>
<td>13* (1.1)</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>Asian/Pacific Islander</td>
<td>68 (2.0)</td>
<td>26 (1.8)</td>
<td>6 (0.6)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>American Indian/Alaska Native</td>
<td>52* (2.1)</td>
<td>35* (1.6)</td>
<td>13* (1.9)</td>
</tr>
</tbody>
</table>

*Significant difference from White at .05 level.

#### School-Lunch Eligibility

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Sub-group</th>
<th>Less than 5 hours (collapsed)</th>
<th>5-6.9 hour</th>
<th>7 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard error</td>
<td>Percentage</td>
<td>Standard error</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>Eligible</td>
<td>10* (0.5)</td>
<td>55* (0.8)</td>
<td>35* (0.7)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Not eligible</td>
<td>12 (0.5)</td>
<td>63 (0.8)</td>
<td>26 (0.6)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>National</td>
<td>Eligible</td>
<td>54* (0.9)</td>
<td>33* (0.8)</td>
<td>13* (0.7)</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Not eligible</td>
<td>70 (0.7)</td>
<td>24 (0.6)</td>
<td>6 (0.3)</td>
</tr>
</tbody>
</table>

*Significant difference from Not Eligible at .05 level.

### Table A-2b. Percentages of students by weekly hours of mathematics instruction at grades 4, 8, 2000 and 1996

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Less than 4 hr (collapsed)</th>
<th>4 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard error</td>
</tr>
<tr>
<td>Grade 4</td>
<td>National</td>
<td>28* (2.0)</td>
<td>72* (2.0)</td>
</tr>
<tr>
<td>2000</td>
<td>National</td>
<td>24 (2.5)</td>
<td>66 (2.5)</td>
</tr>
<tr>
<td>Grade 8</td>
<td>National</td>
<td>49* (0.7)</td>
<td>51* (0.7)</td>
</tr>
<tr>
<td>1996</td>
<td>National</td>
<td>67 (2.6)</td>
<td>33 (2.6)</td>
</tr>
</tbody>
</table>

*Significant difference from 1996 at .05 level.
Table A-3. Percentages of students by time per week on reading-language arts at grades 4 and 8 for students nationally and by achievement levels, public/private school, race/ethnicity and school-lunch eligibility: 2011, 2005 and 2002 (Source NAEP Data Explorer)

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Sub-group</th>
<th>Less than 5 hours (collapsed)</th>
<th>5-6.9 hours</th>
<th>7-9.9 hours</th>
<th>10 hours or more</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Standard error</td>
<td>Percentage</td>
<td>Standard error</td>
<td>Percentage</td>
</tr>
</tbody>
</table>

### All Students

#### Grade 4
- **2011 National**
  - 10* (0.4)
  - 13* (0.4)
  - 30 (0.6)
  - 47* (0.8)
- **2005 National**
  - 8 (0.3)
  - 16 (0.4)
  - 31 (0.4)
  - 44 (0.5)

#### Grade 8
- **2011 National**
  - 47* (0.7)
  - 32* (0.6)
  - 16 (0.5)
  - 6* (0.3)
- **2002 National**
  - 57 (1.1)
  - 26 (0.9)
  - 14 (0.8)
  - 3 (0.4)

*Significant difference from earliest year at the .05 level

### Achievement Levels

#### Grade 4
- **2011 National**
  - Below Basic: 12* (0.5)
  - Basic: 10 (0.5)
  - Proficient: 9 (0.4)
  - Advanced: 9 (0.7)
- **2011 National**
  - Below Basic: 39* (-0.8)
  - Basic: 47* (-0.7)
  - Proficient: 51 (-0.8)
  - Advanced: 55 (-1.7)

*Significant difference from Advanced at the .05 level

#### Grade 8
- **2011 National**
  - White: 9 -0.4, 14 (0.5), 33 (0.7)
  - Black: 11 (-0.6), 13 (0.8), 29* (1.0)
  - Hispanic: 12* (-0.8), 9* (0.6), 24* (0.8), 56* (1.2)
  - Asian/Pacific Islander: 9 (-0.9), 13 (1.0), 31 (1.7), 48 (2.5)
  - Indian/Alaska Native: 12* (-1.3), 14 (1.9), 30 (2.2)
- **2011 National**
  - White: 52 (0.7), 30 (0.7), 14 (0.4), 4 (0.3)
  - Black: 35* (1.1), 37* (1.2), 20* (0.9), 7* (0.4)
  - Hispanic: 39* (1.3), 33 (1.1), 19* (1.2), 9* (0.9)
  - Asian/Pacific Islander: 51 (2.2), 28 (1.6), 13 (1.6), 8* (1.1)
  - Indian/Alaska Native: 46 (2.4), 32 (2.0), 15 (1.7), 6 (0.8)

*Significant difference from White at the .05 level

### Public/Private

#### Grade 4
- **2011 National**
  - Public: 9* (0.4), 12* (0.4), 30* (0.6), 49* (0.8)
  - Private: 15 (1.7), 27 (2.0), 35 (2.2), 22 (2.1)
- **2011 National**
  - Public: 46* (0.7), 32* (0.7), 16 (0.5), 6* (0.3)
  - Private: 52 (2.2), 26 (2.2), 18 (1.8), 4 (0.8)

*Significant difference from Private at the .05 level

### Race/Ethnicity

#### Grade 4
- **2011 National**
  - White: 9 -0.4, 14 (0.5), 33 (0.7), 43 (0.8)
  - Black: 11 (-0.6), 13 (0.8), 29* (1.0), 47* (1.0)
  - Hispanic: 12* (-0.8), 9* (0.6), 24* (0.8), 56* (1.2)
  - Asian/Pacific Islander: 9 (-0.9), 13 (1.0), 31 (1.7), 48 (2.5)
  - Indian/Alaska Native: 12* (-1.3), 14 (1.9), 30 (2.2)
- **2011 National**
  - White: 52 (0.7), 30 (0.7), 14 (0.4), 4 (0.3)
  - Black: 35* (1.1), 37* (1.2), 20* (0.9), 7* (0.4)
  - Hispanic: 39* (1.3), 33 (1.1), 19* (1.2), 9* (0.9)
  - Asian/Pacific Islander: 51 (2.2), 28 (1.6), 13 (1.6), 8* (1.1)
  - Indian/Alaska Native: 46 (2.4), 32 (2.0), 15 (1.7), 6 (0.8)

*Significant difference from White at the .05 level

### School-lunch Eligible

#### Grade 4
- **2011 National**
  - Eligible: 11* (0.5), 11* (0.5), 28* (0.6), 50* (0.9)
  - Not eligible: 8 (0.4), 13 (0.5), 33 (0.6), 47 (1.0)
- **2011 National**
  - Eligible: 40* (0.8), 34* (0.8), 19* (0.7), 8* (0.5)
  - Not eligible: 51 (0.8), 31 (0.7), 13 (0.5), 5 (0.3)

*Significant difference from Not Eligible at the .05 level
### Table A-4. Percentages of grade 8 students by frequency of instruction in visual arts nationally, public/private students/j race/ethnicity and school-lunch eligibility: 2008 and 1994 (Source NAEP Data Explorer)

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Subgroup</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All Students</td>
<td></td>
<td></td>
<td>Public/Private</td>
<td></td>
<td>Race/Ethnicity</td>
<td>School-Lunch Eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Every Day</td>
<td>3-4 Times A Week</td>
<td>1-2 Times A Week</td>
<td>Less than once a week</td>
<td>Subject not taught</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>30</td>
<td>(3.1)</td>
<td>17</td>
<td>(2.4)</td>
<td>30</td>
<td>(3.5)</td>
<td>10</td>
<td>(2.5)</td>
<td>14</td>
<td>(2.4)</td>
<td></td>
</tr>
<tr>
<td>1994 National</td>
<td></td>
<td>28</td>
<td>(3.4)</td>
<td>14</td>
<td>(2.7)</td>
<td>13</td>
<td>(3.5)</td>
<td>13</td>
<td>(2.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public/Private</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>33</td>
<td>(3.5)</td>
<td>✶</td>
<td>(2.8)</td>
<td>28</td>
<td>(3.5)</td>
<td>10</td>
<td>(2.7)</td>
<td>13</td>
<td>(2.4)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>#</td>
<td>✶</td>
<td>3</td>
<td>(0.9)</td>
<td>70</td>
<td>(10.2)</td>
<td>10</td>
<td>(6.1)</td>
<td>37</td>
<td>(8.2)</td>
<td></td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>28</td>
<td>(3.6)</td>
<td>17</td>
<td>(3.1)</td>
<td>34</td>
<td>(4.4)</td>
<td>11</td>
<td>(3.4)</td>
<td>11</td>
<td>(2.6)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>31</td>
<td>(4.9)</td>
<td>18</td>
<td>(4.6)</td>
<td>24</td>
<td>(4.5)</td>
<td>10</td>
<td>(4.9)</td>
<td>18</td>
<td>(4.9)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>41</td>
<td>(6.6)</td>
<td>15</td>
<td>(3.9)</td>
<td>23</td>
<td>(4.7)</td>
<td>5</td>
<td>(2.0)</td>
<td>17</td>
<td>(4.5)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>33</td>
<td>(7.8)</td>
<td>22</td>
<td>(6.0)</td>
<td>29</td>
<td>(6.0)</td>
<td>11</td>
<td>(4.6)</td>
<td>5</td>
<td>(2.5)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>#</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td></td>
</tr>
<tr>
<td>School-Lunch Eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>32</td>
<td>(4.0)</td>
<td>15</td>
<td>(3.1)</td>
<td>26</td>
<td>(3.7)</td>
<td>9</td>
<td>(2.2)</td>
<td>18</td>
<td>(3.4)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>31</td>
<td>(3.6)</td>
<td>15</td>
<td>(3.1)</td>
<td>30</td>
<td>(4.1)</td>
<td>10</td>
<td>(2.2)</td>
<td>18</td>
<td>(2.4)</td>
<td></td>
</tr>
</tbody>
</table>

*Significant difference from Private at the .05 level

### Table A-5. Percentages of grade 8 students by frequency of instruction in music nationally, public/private students/j race/ethnicity and school-lunch eligibility: 2008 and 1994 (Source NAEP Data Explorer)

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Sub-group</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
<th>Percentage</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All Students</td>
<td></td>
<td></td>
<td>Public/Private</td>
<td></td>
<td>Race/Ethnicity</td>
<td>School-Lunch Eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Every Day</td>
<td>3-4 Times A Week</td>
<td>1-2 Times A Week</td>
<td>Less than once a week</td>
<td>Subject not taught</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Students</td>
<td></td>
<td>37</td>
<td>(3.3)</td>
<td>20</td>
<td>(2.7)</td>
<td>27</td>
<td>(3.1)</td>
<td>8</td>
<td>(2.0)</td>
<td>9</td>
<td>(2.0)</td>
<td></td>
</tr>
<tr>
<td>All Students</td>
<td></td>
<td>34</td>
<td>(3.6)</td>
<td>15</td>
<td>(3.0)</td>
<td>30</td>
<td>(4.0)</td>
<td>9</td>
<td>(2.8)</td>
<td>12</td>
<td>(2.4)</td>
<td></td>
</tr>
<tr>
<td>Public/Private</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>39</td>
<td>(3.7)</td>
<td>22</td>
<td>(2.9)</td>
<td>24*</td>
<td>(3.2)</td>
<td>7</td>
<td>(2.1)</td>
<td>8</td>
<td>(2.1)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>#</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td></td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>35</td>
<td>(3.8)</td>
<td>22</td>
<td>(3.2)</td>
<td>29</td>
<td>(4.0)</td>
<td>8</td>
<td>(2.5)</td>
<td>6</td>
<td>(2.5)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>39</td>
<td>(5.7)</td>
<td>18</td>
<td>(4.9)</td>
<td>26</td>
<td>(5.5)</td>
<td>8</td>
<td>(4.5)</td>
<td>10</td>
<td>(2.9)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>43</td>
<td>(5.4)</td>
<td>16</td>
<td>(4.1)</td>
<td>21</td>
<td>(4.1)</td>
<td>6</td>
<td>(2.3)</td>
<td>14</td>
<td>(4.3)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>34</td>
<td>(6.5)</td>
<td>27</td>
<td>(8.9)</td>
<td>25</td>
<td>(6.6)</td>
<td>8</td>
<td>(3.4)</td>
<td>7</td>
<td>(4.1)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td>✶</td>
<td></td>
</tr>
<tr>
<td>School-Lunch Eligibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>37</td>
<td>(4.1)</td>
<td>21</td>
<td>(3.3)</td>
<td>26</td>
<td>(3.4)</td>
<td>6</td>
<td>(2.1)</td>
<td>10</td>
<td>(2.1)</td>
<td></td>
</tr>
<tr>
<td>2008 National</td>
<td></td>
<td>38</td>
<td>(3.9)</td>
<td>21</td>
<td>(3.0)</td>
<td>26</td>
<td>(4.0)</td>
<td>8</td>
<td>(2.5)</td>
<td>9</td>
<td>(2.6)</td>
<td></td>
</tr>
</tbody>
</table>

*Significant difference from Private at the .05 level

*Significant difference from White at the .05 level

*Significant difference from not eligible at the .05 level

49
### Table A6a

Percentages of students at grade 8 by amount of math homework teacher assigns per day (collapsed) at grade 4 for students nationally and by achievement levels, public/private school, race/ethnicity and school-lunch eligibility: 2011 and 1996 (Source NAEP Data Explorer)

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Subgroup</th>
<th>None</th>
<th>Standard error</th>
<th>15 minutes</th>
<th>Standard error</th>
<th>30 minutes</th>
<th>Standard error</th>
<th>45 min or more (collapsed)</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Percentage</td>
<td></td>
<td></td>
<td>Percentage</td>
<td></td>
<td></td>
<td>Percentage</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All Students</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td></td>
<td>4 (0.2)</td>
<td></td>
<td>48 (0.6)</td>
<td></td>
<td>43 (0.6)</td>
<td></td>
<td>3 (0.3)</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>National</td>
<td></td>
<td>4 (0.8)</td>
<td></td>
<td>52 (2.4)</td>
<td></td>
<td>38 (2.3)</td>
<td></td>
<td>6 (0.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Achievement Levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>below Basic</td>
<td>5* (0.4)</td>
<td></td>
<td>42* (0.7)</td>
<td></td>
<td>45 (0.9)</td>
<td></td>
<td>8* (0.5)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Basic</td>
<td>4* (0.3)</td>
<td></td>
<td>47* (0.6)</td>
<td></td>
<td>44 (0.7)</td>
<td></td>
<td>6* (0.3)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Proficient</td>
<td>3* (0.2)</td>
<td></td>
<td>50 (0.7)</td>
<td></td>
<td>42 (0.7)</td>
<td></td>
<td>4 (0.3)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Advanced</td>
<td>2 (0.2)</td>
<td></td>
<td>52 (1.3)</td>
<td></td>
<td>43 (1.2)</td>
<td></td>
<td>4 (0.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from Advanced at .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Public/Private</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Public</td>
<td>3 (0.2)</td>
<td></td>
<td>48* (0.6)</td>
<td></td>
<td>43 (0.6)</td>
<td></td>
<td>5 (0.3)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Private</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from Private at .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>White</td>
<td>4 (0.1)</td>
<td></td>
<td>54 (0.7)</td>
<td></td>
<td>39 (0.6)</td>
<td></td>
<td>2 (0.2)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Black</td>
<td>3* (0.2)</td>
<td></td>
<td>41* (0.8)</td>
<td></td>
<td>48* (0.9)</td>
<td></td>
<td>8* (0.6)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Hispanic</td>
<td>2* (0.3)</td>
<td></td>
<td>37* (1.1)</td>
<td></td>
<td>51* (1.1)</td>
<td></td>
<td>9* (0.8)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Asian/Pacific Islander</td>
<td>2* (0.4)</td>
<td></td>
<td>44* (2.1)</td>
<td></td>
<td>47* (2.1)</td>
<td></td>
<td>7* (0.9)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>American Indian/Alaska Native</td>
<td>8* (1.0)</td>
<td></td>
<td>50 (1.8)</td>
<td></td>
<td>37 (1.8)</td>
<td></td>
<td>4 (0.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from White at .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School-lunch Eligibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Eligible</td>
<td>4 (0.3)</td>
<td></td>
<td>43* (0.7)</td>
<td></td>
<td>46* (0.7)</td>
<td></td>
<td>7* (0.4)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Not eligible</td>
<td>3 (0.2)</td>
<td></td>
<td>53 (0.7)</td>
<td></td>
<td>41 (0.7)</td>
<td></td>
<td>4 (0.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from not eligible at .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table A6b

Percentages of students at grade 8 by amount of math homework teacher assigns per day (collapsed) at grade 8 for students nationally and by achievement levels, public/private school, race/ethnicity and school-lunch eligibility: 2011 and 1996 (Source NAEP Data Explorer)

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Subgroup</th>
<th>None</th>
<th>Standard error</th>
<th>Less than 1 hour</th>
<th>Standard error</th>
<th>1 hr or more (collapsed)</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All Students</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td></td>
<td>2 (0.2)</td>
<td></td>
<td>81 (0.5)</td>
<td></td>
<td>17 (0.4)</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>National</td>
<td></td>
<td>2 (0.6)</td>
<td></td>
<td>94 (0.9)</td>
<td></td>
<td>4 (0.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Achievement Levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>below Basic</td>
<td>4* (0.4)</td>
<td></td>
<td>80 (0.8)</td>
<td></td>
<td>17* (0.7)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Basic</td>
<td>2* (0.2)</td>
<td></td>
<td>82 (0.5)</td>
<td></td>
<td>16* (0.5)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Proficient</td>
<td>1 (0.1)</td>
<td></td>
<td>81 (0.6)</td>
<td></td>
<td>18 (0.7)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>at Advanced</td>
<td>8* (1)</td>
<td></td>
<td>89 (1.1)</td>
<td></td>
<td>20 (1.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from Advanced at the .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Public/Private</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Public</td>
<td>2* (0.2)</td>
<td></td>
<td>82* (0.5)</td>
<td></td>
<td>16* (0.5)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Private</td>
<td>1 (0.4)</td>
<td></td>
<td>74 (2.0)</td>
<td></td>
<td>25 (1.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from Private at the .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>White</td>
<td>2 (0.2)</td>
<td></td>
<td>84 (0.5)</td>
<td></td>
<td>14 (0.4)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Black</td>
<td>2 (0.2)</td>
<td></td>
<td>78* (0.7)</td>
<td></td>
<td>20* (0.7)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Hispanic</td>
<td>3 (0.4)</td>
<td></td>
<td>75* (1.0)</td>
<td></td>
<td>22* (1.0)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Asian/Pacific Islander</td>
<td>1* (0.1)</td>
<td></td>
<td>75* (1.9)</td>
<td></td>
<td>24* (1.9)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>American Indian/Alaska Native</td>
<td>6* (1.3)</td>
<td></td>
<td>79* (1.7)</td>
<td></td>
<td>16 (1.6)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from White at the .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School-lunch Eligibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Eligible</td>
<td>3* (0.3)</td>
<td></td>
<td>79* (0.7)</td>
<td></td>
<td>18* (0.6)</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>Not eligible</td>
<td>1 (0.1)</td>
<td></td>
<td>83 (0.5)</td>
<td></td>
<td>15 (0.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Significant difference from Not Eligible at the .05 level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>