# AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:30 pm</td>
<td>Executive Committee Overview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Welcome and Introductions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• New Committee Assignments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Agenda Overview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Plans for Charge to Board Committees</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*David Driscoll, Chair</td>
<td></td>
</tr>
<tr>
<td>4:35 pm</td>
<td>Committee Issues and Challenges</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Committee Chairs</td>
<td></td>
</tr>
<tr>
<td>4:50 pm</td>
<td>Update: Education Summit for Parent Leaders</td>
<td>*Attachment A</td>
</tr>
<tr>
<td></td>
<td>*Tonya Miles, Terry Mazany, Summit Planning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Committee Members</td>
<td></td>
</tr>
<tr>
<td>5:00 pm</td>
<td>Governing Board 25th Anniversary Planning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Mary Crovo, Deputy Executive Director</td>
<td></td>
</tr>
<tr>
<td>5:05 pm</td>
<td>NAEP Budget and Reauthorization Update</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Ray Fields, Assistant Director for Policy and Research</td>
<td></td>
</tr>
<tr>
<td>5:10 pm</td>
<td>Draft Schedule of Assessments</td>
<td>*Attachment B</td>
</tr>
<tr>
<td></td>
<td>*Cornelia Orr, Executive Director</td>
<td></td>
</tr>
<tr>
<td>5:25 pm</td>
<td>Executive Committee Suggestions for Future Meeting Topics</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*Cornelia Orr</td>
<td></td>
</tr>
<tr>
<td>5:30 pm</td>
<td>Adjourn</td>
<td></td>
</tr>
</tbody>
</table>
Education Summit for Parent Leaders

1. Introduction and Background

The National Assessment Governing Board (www.nagb.org) 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. Congressionally authorized and funded since 1969, NAEP is the only continuing source of national and state-comparable data on student achievement at grades 4, 8, and 12, as well as for 21 urban school districts participating in the Trial Urban District Assessment (TUDA) (see nationsreportcard.gov).

The Governing Board has approved an initiative to reach parent leaders with NAEP data and resources. A component of this initiative is a plan to conduct an Education Summit for Parent Leaders in Arlington, VA on January 13, 2014.

2. Goal

The goal of the Summit is to convey to parent leaders the urgency of improving student achievement in the United States for all children and the urgency of reducing achievement gaps between student subgroups.

The Summit will enable attendees to use NAEP data and resources to ask the right questions of their education leaders about the status of student achievement and gaps in achievement locally, and to promote productive conversations about what is being done to improve achievement and close achievement gaps.

3. Summit Description

The Summit will be conducted in Washington, D.C. and available across the nation via live-streaming internet video, and/or live TV and radio coverage if feasible. It will be designed to be compelling, informative and non-partisan. The Summit audience of approximately 150 would consist primarily of parent and community leaders.

The day-long event would be broken into two major components. The first component, in the morning, will focus on information sharing about NAEP, NAEP results, and the reasons why it is urgent to improve achievement and close achievement gaps.

The second component, in the afternoon, will consist of workshop sessions in which the attendees have hands-on experience using NAEP data and resources, applying what was learned in the morning session.

Morning Session

The morning session would include individual, panel, and media presentations. These will address the reasons for improving achievement and closing achievement gaps from a wide range of perspectives. The intent is for these perspectives, taken together, to convey the importance of the issue for individuals and the nation, raise awareness on the part of participants, and provide a compelling, unassailable argument for the urgent need to take action.
While a wide range of perspectives will be presented, no single specific course of action, ideology, or pedagogy will be advanced or promoted at the summit as “the solution.” Whether a parent of a school child, empty nesters, or childless, the presentations will answer questions of both individual and national interest in closing achievement gaps and improving achievement overall.

The multiple perspectives presented are intended to reach individuals on as many levels as possible, and to reach everyone on some level. Having impact both intellectually and emotionally is essential. Giving a human face through story-telling to convey the substantive concepts and the sense of urgency is critical.

Potential outcomes of the morning session would be acknowledgement by participants of the urgency for action and an expression of commitment to take action locally.

Diversity, balance, and bi-partisanship in a non-partisan manner are the watchwords for the planning and execution of the symposium.

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, will present the NAEP data as clear evidence of the critical need to improve student achievement and close achievement gaps between student subgroups.

The range of perspectives made by speakers or through media presentations could include the following (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
- National 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, political or ideological 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

For maximum impact, some of these perspectives may be combined, when appropriate. They should be presented in ways that engage, inform, and even entertain, always keeping the nature of the parent leader audience in mind.
Afternoon Session

The 3-4 hour hands-on workshops will familiarize participants about NAEP data and resources and their potential for prompting relevant questions to ask of education leaders. At a minimum, attendees will end the day knowledgeable about how to access NAEP data and resources. Optimally, attendees will end the day having framed a set of questions relevant to their locale and having identified specific NAEP resources to use as references in support of the questions.

The workshop will be conducted in a way that recognizes and honors, first and foremost, that state and local authorities are responsible for all aspects of education decision-making about standards, curricula, and assessments.

NAEP resources include:

- Student achievement results
- Background information on student demographics, instructional practices, and school policy
- Assessment frameworks and specifications (the blueprint for each assessment)
- Hundreds of NAEP released test questions in each subject assessed
- On-line tools for analyzing data, and producing tables and charts

Afternoon Hands-on Workshop

In the invitation/announcement, we will ask those who register to bring their laptop or tablet to the summit and to indicate in the RSVP if they will do so. From this, we will know the degree to which the lack of equipment will be an issue and we can plan accordingly. For example, we can organize the breakout groups so that all of those with/without equipment are together and fine-tune the content accordingly.

The breakout rooms will have wifi. Those who do not bring equipment will be able to follow the demonstrations by the group leader and will have specific take-home instructions among the materials provided.

Workshop Outline

1. Disclaimer to be made orally and in printed workshop materials (something along the lines of): Because NAEP is a public resource developed with public funds, and because all states and 21 large urban districts participate in NAEP, it is appropriate for parent leaders to become aware of NAEP data and resources to use as a mirror to help reflect on student achievement locally. However, the workshop will not in any way promote the content of NAEP as superior to local curricula or as a curriculum to be followed.

2. Assessment Literacy 101: will provide an overview of the testing landscape in the U.S.--NAEP, state tests, local tests, international assessments, college admissions tests and how these fit together.
3. Understanding and using NAEP data
   a. National, state, TUDA
   b. Trends and gaps
   c. Achievement levels
   d. Average scores
   e. Percentiles

4. Understanding and using the state mapping study

5. Understanding and using NAEP frameworks

6. Understanding and using NAEP items and data

7. Understanding and using background information and data

8. Accessing and using the on-line NAEP tools
Education Summit for Parent Leaders

The National Assessment Governing Board is planning a one-day education summit for parent leaders on January 13, 2014 in Arlington, VA. 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.

Individual and panel presentations in the morning plenary will 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 (listed alphabetically):

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

The afternoon hands-on workshops will familiarize participants about NAEP data and resources and their potential for prompting questions to ask of education leaders relevant to improving achievement and closing achievement gaps. At a minimum, attendees will end the day knowledgeable about how to access NAEP data and resources. Optimally, attendees will end the day having framed a set of questions relevant to their locale and having identified specific NAEP resources to use as references in support of the questions.

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, 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.
PLENARY SESSION (Will be adjusted to accommodate schedules of First Lady and Secretary)

9:00 AM Welcome and Overview
David Driscoll (and Eileen, Rebecca, Terry and Tonya—as parent leaders themselves—to share/trade off moderator role throughout the day)

9:10 AM Keynote—Why it is Urgent for Parent Leaders to Address Improving Achievement and Closing Achievement Gaps
Secretary Duncan/First Lady (Invited)

9:40 AM Presentation—Status of Achievement and Achievement Gaps in the U.S.
Kati Haycock, The Education Trust (confirmed)

10:00 AM Presentation—National Security Imperative/Parent Perspective for Improving Achievement and Closing Achievement Gaps
Otha Thornton, National PTA/Lt. Col., U.S. Army (ret.) (confirmed)

10:20 AM Presentation—Demographic Imperative for Improving Achievement and Closing Achievement Gaps
Steve Murdock, Rice University (confirmed)

10:45 AM Break

11:00 AM Presentation—Business/Economic Imperative for Improving Achievement and Closing Achievement Gaps
Lee Blitch (availability confirmed)

11:15 AM Panel: Civil Rights/Equity/Moral Imperative for Improving Achievement and Closing Achievement Gaps
Charles Payne, University of Chicago (confirmed)
Marc Morial, National Urban League (confirmed)
Janet Murguia (or Delia Pompa), National Council of La Raza (confirmed)

Noon Working Lunch—Partnering with Parent Leaders to Improve Achievement and Close Achievement Gaps
Panel of Local and State Education Leaders (to be determined—Andrés Alonso reaching out on our behalf)

BREAKOUT SESSIONS
1:00 PM Hands-on Workshops

PLENARY SESSION
4:15 PM What Parent Leaders Can Do/Asking the Right Questions
Panel of Parent Leaders (to be determined)

4:45 PM Next Steps and Follow-up
David Driscoll

5:00 PM Adjourn
NAEP Schedule of Assessments

By law, the Governing Board sets the NAEP schedule of assessments. Board policy is to maintain a schedule of assessments with a 10-year outlook. This provides notice to states and districts for planning for participation. The schedule also is the basis for NAEP operational planning and for developing scopes of work for NAEP contracts. Thus, the NAEP schedule is the primary driver of the NAEP budget. Under the Board By-laws, the Executive Committee is responsible for proposing changes to the NAEP schedule of assessments for Board consideration.

The Board has taken action at the last two meetings on the NAEP schedule. At the May 2013 meeting, the Board postponed the 2014 assessments at grades 4 and 12 in U.S. history, civics and geography. At the August 2013 meeting, the Board postponed the following:

- the 2015 High School Transcript Study
- the 2015 state-level assessments at grade 12 in reading, mathematics, and science
- the 2016 Long-Term Trend assessments in reading and mathematics at ages 9, 13 and 17

The current approved schedule ends at 2017 and additional assessment years need to be added for a 10-year outlook. The current schedule is on the second page following for reference.

Although no action is required at the December 2013 Board meeting, it is prudent for the Executive Committee to begin consideration of the subjects to be assessed in the five-year period following 2017.

As background for the discussion, a draft of a schedule of assessments through 2022 has been prepared by staff. The principles or assumptions underlying the discussion draft are:

- Continue to cover the broad range of subject areas
- Administer all assessments using technology beginning in 2017
- Conduct assessments in
  - reading and mathematics at grades 4 and 8 once every two years
  - reading and mathematics at grade 12 once every four years
  - science and writing once every four years in alternating biennia
  - U.S. history, civics, and geography once every four years
  - technology and engineering once every four years, expanding from grade 8 in 2014 by one additional grade each cycle through 2022
  - the arts, economics, foreign language, and world history at least twice in a ten-year period

The Executive Committee will discuss the pros and cons of these principles and assumptions.
## NAEP Schedule of Assessments – Discussion Draft

<table>
<thead>
<tr>
<th>Year</th>
<th>National</th>
<th>State</th>
</tr>
</thead>
<tbody>
<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 (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Civics (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geography (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>TECHNOLOGY AND ENGINEERING LITERACY (8)</strong> **</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>Reading</td>
<td>Reading (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td>Math (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Science**</td>
<td>Science (4, 8)</td>
</tr>
<tr>
<td>2016</td>
<td>Arts (8)</td>
<td></td>
</tr>
</tbody>
</table>

**Assessments involving partial or full 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.
<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>MATH (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 (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Civics (8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geography (8)</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)</td>
</tr>
<tr>
<td></td>
<td>Mathematics</td>
<td>Math (4, 8)</td>
</tr>
<tr>
<td></td>
<td>Science**</td>
<td>Science (4, 8)</td>
</tr>
<tr>
<td>2016</td>
<td>Arts (8)</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.
## AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 – 10:15 am</td>
<td>Welcome, Introductions, and Agenda Overview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comments from ADC Chair and Vice Chair</td>
<td>Shannon Garrison, Chair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cary Sneider, Vice Chair</td>
</tr>
<tr>
<td>10:15 – 10:30 am</td>
<td>Chairman’s Charge to the Committee for 2014</td>
<td>David Driscoll, Board Chair</td>
</tr>
<tr>
<td>10:30 – 11:00 am</td>
<td>Technology and Engineering Literacy (TEL) Assessment Update</td>
<td>Lonnie Smith, ETS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attachment A</td>
</tr>
<tr>
<td>11:00 – 11:15 am</td>
<td>Update on Reporting Grade 4 Computer-Based Writing Information</td>
<td>Ebony Walton Chester, NCES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attachment B</td>
</tr>
<tr>
<td>11:15 am – 12:05 pm</td>
<td>Transitioning to NAEP Technology-Based Assessments in Reading and Mathematics</td>
<td>William Ward, NCES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attachment C</td>
</tr>
<tr>
<td>12:10 – 12:45 pm</td>
<td><strong>Joint Meeting with Reporting and Dissemination Committee</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>NAEP Contextual Variables</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>ACTION:</strong> Contextual Information Framework for NAEP</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Larry Feinberg, NAGB Staff</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>NAEP Education Indicators</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Alan Ginsburg and Marshall Smith, Consultants</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Information Item</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NAEP Item Review Schedule</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attachment D</td>
</tr>
</tbody>
</table>
Summary of Technology and Engineering Literacy (TEL) Activities

Update Session

The NAEP Technology and Engineering Literacy (TEL) assessment, designed to gauge how well students can apply their understanding of technology principles to real-life situations, will be administered for the first time in 2014 to a nationally representative sample of 8th graders.

To better inform and prepare the public for this assessment and to increase overall awareness, a TEL information page (http://nces.ed.gov/nationsreportcard/tel/) is currently hosted on the Nation’s Report Card website. In addition to various resources (e.g., fact sheet for public schools), information on the TEL page includes two orientation videos, the student tutorial, and a sample scenario-based task. In the absence of any past cognitive items (since this is a first-time assessment) the sample task (Wells) is included to provide an exemplar of an actual TEL task. The Wells task was administered in the 2013 pilot.

Currently, there are no supporting materials (e.g., scoring guides) for the Wells task provided on the TEL web page. At the December 2013 meeting of the Assessment Development Committee (ADC), members will learn about forthcoming additions to the Wells web-based materials including a description of the specific TEL skills targeted by the task, the kinds of decisions students make, the types of information collected from student performance (e.g., the path a student takes to navigate through the task), and how students’ responses effect their task score.

Additionally, the plans for the full TEL administration in 2014 will be summarized, including dates and sampling methods.
Reporting Lessons Learned from the 2011 Grade 4 Writing Pilot:

Progress Report

Technology is changing the way we assess and report student achievement. In congruence with this technological shift, NAEP has updated how it assesses student writing by conducting a writing computer-based assessment (WCBA) at grades 8 and 12 and piloting this assessment at grade 4. While it was clear that students in the higher grades could aptly demonstrate their writing ability on the computer, it was unknown whether fourth-graders could type their responses fully in the allotted time. To answer this and several other related questions, NCES coordinated usability studies, interviews, and a WCBA pilot assessment to examine what type of platform students needed to successfully complete the assessment and how well did fourth-graders write in different assessment conditions.

To disseminate findings from these efforts and to share the process for developing the WCBA platform, NCES leveraged technology to create an interactive web tool that shares lessons learned from the grade 4 WCBA. The website covers four areas: performance, development, accommodations, and (assessment) questions. Each section will have brief summaries describing what NCES did and learned and will contain supporting images and graphics.

At the August 2013 Governing Board meeting, the Assessment Development Committee (ADC) received a preview of the website and provided feedback to NCES. Since that meeting, NCES completed a technical memorandum that summarizes the development, administration and outcomes of the pilot assessment. This document will be available to the public upon request through the website. Additionally, more sections of the website were populated. During the December 2013 ADC meeting, the Committee will view additional web content that pertains to the development of the WCBA platform. The presentation will cover information on what students found difficult to understand in the original WCBA platform, and how the assessment platform changed in response.

The website is scheduled for public release in January 2014 and could serve as a model for future dissemination of “lessons learned” from other assessment activities.
Transitioning to Technology-Based Assessments (TBA) for NAEP Reading and Mathematics

Transitioning reading and mathematics paper-and-pencil tests to technology-based assessments (TBA) is among the major challenges and opportunities facing the NAEP program. The intent is not to simply transfer the existing items to electronic delivery, but to introduce new types of technology-enabled items that can measure knowledge and skills that could not be tested—or could not be tested as well—on paper. NAEP being a trend assessment, there is also a desire to continue to measure progress within each subject area by maintaining NAEP trend lines, despite the change from paper-and-pencil to technology-based assessment modes.

At this December 2013 meeting of the Assessment Development Committee (ADC), NCES will present our proposed design for transitioning NAEP to technology-based assessments (including timelines and next steps). The design balances the introduction of innovative TBA content with the desire for trend maintenance, and features the use of a “TBA start-up” administration prior to the first operational TBA year. This TBA start-up would serve as an early bridge study to evaluate the feasibility of maintaining trends across the change in delivery mode.

Additionally, the individual TBA transitions for reading and mathematics, including the shifting emphases in item types as the transition is under way, will be discussed in greater detail. Topics will include how the paper-and-pencil items will be translated and/or transferred and administered for technology-based delivery and how pilot testing will be conducted in an effort to start to bring more TBA-dependent content into the assessment (while balancing innovation with a desire for trend maintenance).
### Assessment Development Committee
#### Item Review Schedule
**December 2013 – July 2014**
*(Updated 11/01/13)*

<table>
<thead>
<tr>
<th>Review Package to Board</th>
<th>Board Comments to NCES</th>
<th>Survey/ Cognitive</th>
<th>Review Task</th>
<th>Approx Number Items</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/21/13</td>
<td>12/13/13</td>
<td>Survey</td>
<td>2015 Operational Reading (4, 8)</td>
<td>78</td>
<td>Review in early Dec.</td>
</tr>
<tr>
<td>2/13/14</td>
<td>3/6/14</td>
<td>Cognitive</td>
<td>2015 Operational Reading (4, 8)</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>2/13/14</td>
<td>3/6/14</td>
<td>Cognitive</td>
<td>2015 Operational Math (4, 8)</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>5/1/14</td>
<td>5/22/14</td>
<td>Cognitive</td>
<td>Science ICT Beta Reviews (4, 8, 12)</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>7/17/14</td>
<td>8/7/14</td>
<td>Cognitive</td>
<td>Science ICT Clearance Review (4, 8, 12)</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>TBD</td>
<td>Survey</td>
<td></td>
<td>Reading question pool* (4, 8)</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>TBD</td>
<td>Survey</td>
<td></td>
<td>Civics question pool* (4, 8, 12)</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>TBD</td>
<td>Survey</td>
<td></td>
<td>Economics question pool* (12)</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>TBD</td>
<td>Survey</td>
<td></td>
<td>Geography question pool* (4, 8, 12)</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td>TBD</td>
<td>Survey</td>
<td></td>
<td>US History question pool* (4, 8, 12)</td>
<td>105</td>
<td></td>
</tr>
</tbody>
</table>

*NOTE: Alpha builds will be presented to the ADC during their in-person and virtual meetings. These will not be submitted before the review. The ADC will receive outlines and beta builds prior to the ICT review meetings. (Alpha and beta builds are the first- and second-draft versions of the rendered task, respectively.)*

*A survey question pool represents all the questions that have been administered in operational assessments.*
## AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 – 10:45 am</td>
<td>NAEP Testing and Reporting on Students with Disabilities and English Language Learners</td>
<td></td>
</tr>
</tbody>
</table>
|               | [Joint meeting with Reporting and Dissemination]                      | Larry Feinberg, NAGB Staff  
Grady Wilburn, NCES  
[Lou Fabrizio, COSDAM Chair]  
[Joint meeting with Reporting and Dissemination]  
See Attachment A under Reporting & Dissemination |
| 10:55 – 11:05 am | Introductions and Welcome to Lucille Davy  
Lou Fabrizio, COSDAM Chair |
| 11:05 – 11:45 am | Discussion on Achievement Level Setting (ALS) on the 2014 NAEP Technology and Engineering Literacy (TEL) Assessment (Closed Session)  
- 2013 TEL Field Trial Scaling Analyses  
- Implications for Planning the TEL ALS  
Sharyn Rosenberg, NAGB Staff  
Andreas Oranje, ETS |
| 11:45 – 12:20 pm | NAEP 12th Grade Academic Preparedness Research  
- Reporting Grade 12 Results Using Preparedness Research Findings  
- National and State Partnerships  
Ray Fields, NAGB Staff  
Sharyn Rosenberg, NAGB Staff |
| 12:20 – 12:35 pm | Board Chairman’s Charge to the Committee for 2014  
David Driscoll, NAGB Chair |
| 12:35 – 12:45 pm | Other Issues  
- Instructional Sensitivity and NAEP  
W. James Popham, COSDAM Member |

### Information Items:

- Update on Evaluation of NAEP Achievement Levels Procurement  
- Reading for Understanding: A Theory-Based, Developmental Approach  
- NAEP 12th Grade Academic Preparedness Research: Phase 2 Research Updates  
  - Course Content Analysis Research  
  - Research with Frameworks
Welcome to Lucille E. Davy, New Governing Board Member

The Committee on Standards, Design and Methodology welcomes new Governing Board member Lucille Davy. Lucille Davy became a member of the Governing Board on October 1, 2013 in the category of General Public Representative. She is President and CEO of Transformative Education Solutions, LLC.

Abbreviated Professional Biography for Lucille E. Davy

Lucille E. Davy is an education policy consultant through her roles as President and CEO of Transformative Education Solutions, LLC, and senior advisor for the James B. Hunt, Jr. Institute for Educational Leadership and Policy. Ms. Davy started her career as a lawyer and adjunct professor of mathematics, a subject for which she received a bachelor’s degree and K-12 teacher certification for New Jersey. She later served as a volunteer and leader for a variety of parent groups and organizations in Westfield Public Schools in New Jersey. Her experience led to service as special counsel for education policy for the New Jersey Governor’s office and as an education policy advisor for several entities, including the Committee for Working Families. From 2005-2010, Ms. Davy served as New Jersey’s Commissioner of Education, overseeing more than 2,400 schools in 600 districts that served 1.4 million children. Since 2010, she has, via her consulting firm and as a Hunt Institute advisor, focused on high school students’ college and career readiness after graduation and effective implementation of the Common Core State Standards in English language arts and mathematics.
Setting Achievement Levels on the NAEP 2014 Technology and Engineering Literacy (TEL) Assessment

Background

At the March 1, 2013 meeting, the Committee began discussion on setting achievement levels for the 2014 NAEP TEL assessment. For the May 17, 2013 meeting, an issues paper was developed to support procurement and project planning for developing recommended achievement levels for TEL. In the Committee’s May 2013 discussion, the Committee expressed a need for more information before proceeding with procurement plans, particularly regarding TEL scaling issues that could hinder a strong TEL Achievement Level Setting (ALS) effort. Initial results from the analysis of TEL field trial data were presented during the August 2, 2013 meeting, but extensive scaling analyses had not yet been conducted. Additional results are now available and will be presented in closed session at the December 2013 meeting. An overview of the presentation can be found on page 5.

Timeline

The following timeline provides a preliminary list of key dates and activities related to TEL assessment development and achievement level setting.

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity</th>
<th>Responsibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008 - 2010</td>
<td>TEL Framework development</td>
<td>ADC, Board, WestEd (contractor)</td>
</tr>
<tr>
<td>2010 - 2012</td>
<td>Assessment development for 2013 pilot test</td>
<td>NCES, NAEP contractors</td>
</tr>
<tr>
<td>2010 - 2012</td>
<td>Item review for 2013 pilot test</td>
<td>NCES, NAEP contractors, TEL Standing Committee, ADC</td>
</tr>
<tr>
<td>Early 2013</td>
<td>Pilot test – national sample, grade 8</td>
<td>NCES, NAEP contractors</td>
</tr>
<tr>
<td>May 2013</td>
<td>TEL ALS issues paper</td>
<td>COSDAM, consultant</td>
</tr>
<tr>
<td>Early 2014</td>
<td>ALS procurement and contract award</td>
<td>Board staff, COSDAM</td>
</tr>
<tr>
<td>Early 2014</td>
<td>Operational administration – national sample, grade 8</td>
<td>NCES, NAEP contractors</td>
</tr>
<tr>
<td>2015</td>
<td>Board action on TEL achievement levels</td>
<td>COSDAM, ALS contractor, Board</td>
</tr>
<tr>
<td>2015</td>
<td>Reporting TEL results</td>
<td>Board, NCES, contractors</td>
</tr>
</tbody>
</table>
TEL Assessment Design

The 2014 Technology and Engineering Literacy (TEL) assessment is based on the Board-adopted Framework and Specifications (see www.nagb.org, Publications).

The TEL assessment is composed of three major areas:

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

Another key dimension of the TEL assessment is the three practices, each of which is applicable to the three major areas noted above:

- Understanding Technological Principles
- Developing Solutions and Achieving Goals
- Communicating and Collaborating

The TEL assessment was developed using an evidence-centered design (ECD) approach. From the beginning, all TEL tasks and items were designed using an evidential chain of reasoning that links what is to be measured, the evidence used to make inferences, and the tasks used to collect the desired evidence. In addition to student responses to complex tasks and discrete items, the computer-based TEL assessment allows NAEP to capture a wide array of data on student performance. For example, NAEP will collect information on how students interact with the TEL simulations and experiments. Such data may include the number of experimental trials run and the number and types of variables controlled. These observable data on “strategies and processes” are intended to be used for reporting purposes but are not expected to contribute to the scoring of student performance.

TEL Reporting

Based on the ECD approach, TEL reporting includes plans to expand beyond the traditional NAEP scores. It is expected that data from complex performance tasks and discrete items will be reported in several ways:

- A composite or univariate scale score on which the achievement levels will be set
- Subscores for the content areas (Design and Systems; Information Communication Technology; Technology and Society)
Achievement Level Setting on the NAEP TEL Assessment

- Reporting on the practices (Understanding Technological Principles; Developing Solutions and Achieving Goals; Communicating and Collaborating)
- Information on students’ processes and strategies, related to the ECD model, captured as observable data from their work on the TEL scenario-based tasks.

Potential Discussion Questions for COSDAM

- Given the field trial results, is there sufficient evidence to warrant achievement level setting on the overall construct?
- What additional information would help to inform the standard setting process?
At the August Board meeting, an update was provided to COSDAM on the Technology and Engineering Literacy (TEL) field trial analyses with the goal of preparing for standard setting and setting a tentative timeline based on when pertinent (empirical) results would be available to support the standard setting. At that point, analyses had just begun and percent correct and item-block-biserial correlations were shared, indicating a reasonable item pool covering a range of proficiencies. In addition to sharing and discussing results, a discussion ensued about the need for some dimensionality analyses to determine at what level (e.g., overall, by domain) meaningful standards can and ought to be set on this new construct of Technology and Engineering Literacy. Considerable interest was generated for the correlations between subscales, to inform the question of whether it is appropriate to set standards on the overall assessment.

The analyses that can be performed on data from the Technology and Engineering Literacy field trial are more extensive than typical of field trials and more like those on operational assessments. This capability derives from the change in format from paper-based to technology-based assessments—the printing-cost limitation on the block spiral design was removed and a complete spiral design became possible. Of course, important limitations remain. The field trial will provide data for item selection for the operational assessment, so some items will not carry forward and the remaining items will be reconfigured into different blocks that will revise the current position and context effects. Some blocks were found to take too little time. Reconfiguration to create longer assessment units for the operational assessment also has position and context effect implications.

At this point, extensive analyses have been completed with the field trial data and a firmer timeline for the 2014 analysis is available. In this session ETS will:

- Share results, including the correlations between subscales and student performance across scales.
- Provide more detail about what further analyses are planned based on the field trial data and the goals of these analyses.
- Provide a timeline for the 2014 operational analysis and reporting of Technology and Engineering Literacy and an indication of when results and data products will be available.
Phase 1 Research

The first phase of the Governing Board’s research on academic preparedness is now complete; results from more than 30 studies are available at: http://www.nagb.org/what-we-do/preparedness-research.html. During the August 2013 meeting, the Board voted on a motion to use the phase 1 research on academic preparedness for college in the reporting of the 2013 grade 12 national results for reading and mathematics. The approved motion and supporting validity argument also appear on the aforementioned website.

During the December 2013 meeting, COSDAM will hear a brief update on plans for reporting the 2013 grade 12 results for reading and mathematics in terms of academic preparedness for college (scheduled for release in April 2014).

Phase 2 Research

The second phase of the Governing Board’s research on academic preparedness currently consists of the following studies that are planned or underway:

<table>
<thead>
<tr>
<th>Study name</th>
<th>Sample</th>
<th>December 2013 Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical linking of NAEP and ACT</td>
<td>National; FL, IL, MA, MI, TN</td>
<td>See pp. 7-10 for overview and draft research questions</td>
</tr>
<tr>
<td>Longitudinal statistical relationships: Grade 12 NAEP</td>
<td>FL, IL, MA, MI, TN</td>
<td></td>
</tr>
<tr>
<td>Statistical linking of NAEP and EXPLORE</td>
<td>KY, NC, TN</td>
<td></td>
</tr>
<tr>
<td>Longitudinal statistical relationships: Grade 8 NAEP</td>
<td>KY, NC, TN</td>
<td></td>
</tr>
<tr>
<td>Content alignment of NAEP and COMPASS</td>
<td></td>
<td>See pp. 11-12 for overview</td>
</tr>
<tr>
<td>Content alignment of NAEP and EXPLORE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College Course Content Analysis</td>
<td></td>
<td>See pp. 27-36 for informational update</td>
</tr>
<tr>
<td>Evaluating Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training (Research with Frameworks)</td>
<td></td>
<td>See pp. 37-38 for informational update</td>
</tr>
</tbody>
</table>

During the December 2013 meeting, COSDAM will receive an update on the status of the national and state partnerships and will discuss draft research questions for the statistical relationship studies.

Brief overviews and status updates on the College Course Content Analysis and Research with Frameworks are provided as information items in Attachment H.

Overarching Research Questions for Statistical Relationship Studies:
1. What scores on the 2013 grade 8 and 12 NAEP Reading and Mathematics assessments predict academic preparedness for college?
2. Is it feasible to use NAEP to make state-level inferences about academic preparedness for college?
In 2013, the Governing Board is planning to partner with ACT, Inc. to conduct a statistical linking study at the national level between NAEP and the ACT in Reading and Mathematics. Through a procedure that protects student confidentiality, the ACT 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 was performed with the SAT in 2009. There will not be a statistical linking study performed for NAEP and the SAT in 2013.

In addition, the state-level studies, begun in 2009 with Florida, will be expanded in 2013. Again using a procedure that protects student confidentiality, ACT scores of NAEP 12th grade test takers in the state samples in partner states will be linked to NAEP scores. We are in the planning stages with five states to be partners in these studies at grade 12: Florida, Illinois, Massachusetts, Michigan, and Tennessee. In three of these states (IL, MI, TN), the ACT is administered to all students state-wide, regardless of students’ intentions for postsecondary activities.

Draft Research Questions for National and State Statistical Linking Studies with the ACT:

1. What are the correlations between the grade 12 NAEP and ACT student score distributions in Reading and Math?
2. What scores on the grade 12 NAEP Reading and Math scales correspond to the ACT college readiness benchmarks? (concordance and/or projection)
3. What are the average grade 12 NAEP Reading and Math scores and interquartile ranges (IQR) for students below, at, and at or above the ACT college readiness benchmarks?
4. Do the results differ by race/ethnicity or gender?
In addition to the linking of ACT scores to NAEP 12th grade test scores in partner states, the postsecondary activities of NAEP 12th grade test takers will be followed for up to six years using the state longitudinal databases in Florida, Illinois, Massachusetts, Michigan, and Tennessee. These studies will examine the relationship between 12th grade NAEP scores and scores on placement tests, placement into remedial versus credit-bearing courses, GPA, and persistence. Data sharing agreements are in development for each state partner.

**Draft Research Questions for Longitudinal Statistical Relationships, Grade 12 NAEP:**

1. What is the relationship between grade 12 NAEP Reading and Math scores and grade 8 state test scores?
2. What are the average grade 12 NAEP Reading and Math scores and interquartile ranges (IQR) for students with placement in remedial and non-remedial courses?
3. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students with a first-year GPA of B- or above?
4. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who remain in college after each year?
5. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who graduate from college within 6 years?
State Statistical Linking Studies with EXPLORE

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, are planned with partners in three states: Kentucky, North Carolina, and Tennessee. In all three of these states, EXPLORE is administered to all students state-wide during grade 8.

Draft Research Questions for State Statistical Linking Studies with EXPLORE:

1. What are the correlations between the grade 8 NAEP and EXPLORE scores in Reading and Math?
2. What scores on the grade 8 NAEP Reading and Math scales correspond to the EXPLORE college readiness benchmarks (concordance and/or projection)?
3. What are the average grade 8 NAEP Reading and Math scores and the interquartile ranges (IQR) for students below, at, and at or above the EXPLORE college readiness benchmarks?
Longitudinal Statistical Relationships: Grade 8 NAEP

In 2013, the Governing Board will also expand the state-level studies by partnering with a few states at grade 8. Again using a procedure that protects student confidentiality, secondary and postsecondary data for NAEP 8th grade test takers in the state samples in partner states will be linked to NAEP scores. These studies will examine the relationship between 8th grade NAEP scores and scores on state tests, future ACT scores, placement into remedial versus credit-bearing courses, and first-year college GPA.

Three states will be partners in these studies at grade 8: Kentucky, North Carolina, and Tennessee. Data sharing agreements are in development for each state partner.

Draft Research Questions for Longitudinal Statistical Relationships, Grade 8 NAEP:

1. What is the relationship between NAEP Reading and Math scores at grade 8 and state test scores at grade 4?
2. What are the average NAEP Reading and Math scores and the interquartile ranges (IQR) at grade 8 for students below the ACT benchmarks at grade 11/12? At or above the ACT benchmarks?
3. What are the average NAEP Reading and Math scores and the interquartile ranges (IQR) at grade 8 for students who are placed in remedial and non-remedial courses in college?
4. What are the average NAEP Reading and Math scores (and the IQR) at grade 8 for students who obtain a first-year college GPA of B- or above?
5. What is the relationship between grade 8 NAEP Reading and Math scores and grade 12 NAEP Reading and Math scores? (contingent on feasibility of sampling the same students in TN, NC, and KY)
Content Alignment Study of Grade 12 NAEP Reading and Mathematics and COMPASS

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.

We plan to conduct an alignment study of grade 12 NAEP Reading and Mathematics and ACT COMPASS. At this point in time, details of our agreement with ACT are still being worked out. Detailed plans for conducting this study will be presented at a future meeting.
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.

We plan to conduct an alignment study of grade 8 NAEP Reading and Mathematics and ACT EXPLORE. Results from this content alignment study will be particularly important for interpreting the findings from the NAEP-EXPLORE statistical linking studies. At this point in time, details of our agreement with ACT are still being worked out. Detailed plans for conducting this study will be presented at a future meeting.
For additional background information, the following list presents a brief description of the assessments referenced in the phase 2 academic preparedness research 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.

- **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 EXPLORE** – ACT EXPLORE assesses academic progress of eighth and ninth grade students. It is a component of the ACT College and Career Readiness System and includes assessments of English, Mathematics, Reading, and Science. ACT EXPLORE 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%) by the time students graduate high school with particular score ranges on the EXPLORE assessment.

- **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 SAT has calculated preparedness benchmarks are defined as the SAT scores corresponding to a 65% probability of earning a first-year college grade-point average of 2.67 (B-) or better.
THE EVALUATIVE MISUSES OF COMPARATIVELY FOCUSED TESTS

W. James Popham

University of California, Los Angeles

For almost a full century, the mission of U.S. educational measurement has been to elicit test-takers' scores so that those scores can be compared with one another. This is a good and useful thing to do. It is particularly good and useful thing to do in situations where the numbers of applicants exceeds the numbers of openings. To make a flock of important educational decisions, we need to identify those students who are our strongest or weakest performers. I am an enthusiastic supporter of tests that yield comparative score-interpretations.

The legitimacy of such test-based comparisons was firmly established way back in World War One, almost 100 years ago, when a group-administered intelligence test, the Army Alpha, was administered to about 1,750,000 U.S. Army recruits in an effort to identify men who would be the most suitable candidates for officer training programs. This use of the Alpha to provide comparative score-interpretations was regarded as a smashing success and, although the test was clearly a measure of a test-taker's aptitude, the Alpha’s focus on comparative score-interpretations was soon emulated by the makers of educational achievement tests. Indeed, a number of the test-construction and test-refinement tactics used for today’s U.S. achievement tests can be traced back to the comparative assessment procedures associated with the Army Alpha.

But tests capable of providing comparative score-interpretations are not necessarily tests that should be used to evaluate schools or teachers. Such evaluative applications of educational assessment, although similar in some ways to comparative applications of educational assessment, are fundamentally different. Increasingly, however, America’s educators are being evaluated on the basis of their students’ performances on tests that were created to yield comparative score-interpretations rather than to measure instructional quality. This is a terrible mistake.

1 A written accompaniment to oral remarks, A Trip to Intolerability, presented at the first International Instructional Sensitivity Conference hosted by the Achievement and Assessment Institute of the University of Kansas, Lawrence, Kansas, November 13-15, 2013.
This mistake is being made because of a pervasive but erroneous belief by Americans that students’ test-measured achievement levels, namely, the knowledge and skills students display when responding to achievement tests, can be attributed to what those students have learned in school. In some instances, this is a warranted belief. Certain skills and bodies of knowledge measured by today’s achievement tests have definitely been learned by students because of instruction those student received in school.

Yet, what if the tests we traditionally employ to measure students’ achievement, because of those tests’ preoccupation with providing comparative score-interpretations, also measure many things other than what students were taught in school? What if our traditional achievement tests, in an effort to provide the necessary variance in total-test scores that are so vital for comparative score-interpretations, also measure test-takers’ status with respect to such variance-inducing factors as students’ socioeconomic status and inherited academic aptitudes? Clearly, such a confounding of causality would make such traditional achievement tests less appropriate for evaluating how well students have been taught. To what extent is a student’s performance on a traditional achievement test attributable to what was taught in school rather than what was brought to school? Realistically, for many of today’s achievement tests, we just can’t tell.

I contend that the traditional way we build and burnish our educational achievement tests may lead to those tests’ being inappropriate for use in the evaluating of schools and teachers. The italicized may is intended to emphasize my conviction that, to date, the suitability of today’s traditional achievement tests for evaluative use has not been rigorously scrutinized. But it should be.

Clearly, if one wishes to evaluate the performance of a school’s instructional staff, or the performance of a particular teacher, then it would be better to have evidence on hand from students’ performances on almost any sort of achievement test rather than relying on no achievement evidence at all. Thus, I’d certainly rather use students’ scores from the tests we now employ for such evaluative purposes than have access to no data whatsoever regarding students’ achievement. But the choice before us is not whether we should try to carry out evaluations using flawed tests instead of using no tests at all. Instead, our challenge is to carry out today’s increasingly high-stakes evaluations using the most appropriate tests we can employ. I am certain we can do a better job of evaluating our schools and teachers than we do by using today’s achievement tests.

The Cornerstone of Our Assessment Castle

If you were to ask today’s educators—irrespective of how much they actually knew about educational testing—what is the single, most important concept in educational measurement, the most frequent response to your query would surely be “validity.” That
response, happily, turns out to be the correct answer. Educational measurement is predicated on the conviction that by getting students to make *overt* responses to stimuli such as a test’s items, educators can arrive at valid inferences about students’ *covert* knowledge and skills. Determination of the covert based on the overt, indeed, lies at the heart of all educational assessment.

It is not a *test*, however, that is valid or invalid. Instead, it is the score-based *inference*—an inference based on students’ test scores—that is valid or invalid. Validity thus represents the accuracy of test-based inferences (or, if you prefer, test-based interpretations). Increasingly these days, assessment validity is regarded not only as the accuracy of a test-based inference, but also as the appropriateness of the use to which a test’s score-based inferences are put (Kane, 2013). Optimally, therefore, not only would a test-based inference be accurate, but then that accurate inference would be employed to accomplish a suitable consequence such as subsequently making sound educational decisions about students.

The validity of a score-based inference, therefore, gets our test-usage ball rolling. If we can’t establish that test-takers’ performances lead to an accurate inference about what test-takers’ scores signify, then the likelihood of then making a sensible inference-based decision is definitely diminished. And this is where we currently are with respect to the tests we use to evaluate U.S. schools and teachers. Although educators have been urged (or, in some instances, been statutorily required) to evaluate schools and teachers using students’ performances on educational tests, *we have no meaningful evidence at hand indicating that these tests can accurately distinguish between well taught and badly taught students.* This state of affairs is truly astonishing.

**Instructional Sensitivity**

Yes, our nation increasingly relies on students’ scores on tests, typically using standardized achievement tests, to arrive at inferences about the quality of instruction provided to those students. Yet, the evidence to support the accuracy of such score-based inferences about instructional quality is essentially nonexistent. Today’s educators are being asked to sidestep the most important tenet of educational measurement, namely, the obligation to supply validity evidence regarding the interpretations and significant uses of an educational test’s results. Putting it differently, no evidence currently exists about these evaluative tests’ *instructional sensitivity*.

What is this “instructional sensitivity,” and how is it determined? Actually, the concept is quite a straightforward one, and it simply refers to how well a test can accurately distinguish between test-takers who have been taught well and test-takers who have been taught badly. Although a certain amount of definitional disagreement about
Instructional sensitivity can be found in the measurement community, the following definition reflects what most writers on this topic understand when they refer to a test’s instructional sensitivity:

*Instructional sensitivity is the degree to which students’ performances on a test accurately reflect the quality of instruction specifically provided to promote students’ mastery of what is being assessed* (Popham, 2006).

As you can see, this definition revolves around the “quality of instruction” insofar as it specifically contributes to “students’ mastery” of whatever the test is measuring. A test, then, can vary in the degree to which it is instructionally sensitive. We need not, therefore, distinguish between a test that is totally sensitive to instruction or totally insensitive to instruction. Instructional sensitivity is a continuous rather than a dichotomous variable. Our quest, therefore, should be to determine a minimum threshold of instructional-sensitivity acceptability for any test being used to evaluate the caliber of instruction. The more significant the stakes are that are associated with a test’s use, the higher should be our acceptability-threshold.

The instructional sensitivity of education tests is not a brand new concept. More than 30 years ago, when the high-stakes accountability movement began to capture the attention of American educators, Haladyna and Roid (1981) described the role of instructional sensitivity when judging the merits of accountability tests.

Much earlier, when the initial proponents of criterion-referenced measurement were attempting to sort out how to create and improve tests leading to criterion-referenced inferences, Cox (1971) and other measurement specialists tried to devise ways to maximize a test item’s sensitivity to instruction. But those early deliberations among advocates of criterion-referencing were focused almost exclusively on measurement challenges, that is, how to build tests capable of yielding more valid criterion-referenced inferences. As the years tumbled by, however, the evaluative use of students’ test performances has become more significant. During the next several years, for instance, it is almost certain that many American teachers will lose their jobs primarily because of their students’ poor performances on tests. The high-stakes decisions riding on students’ test scores have, without argument, become higher and higher and higher.

Nonetheless, despite the increased importance now attached to evaluative test-based consequences, the attention given to the instructional sensitivity of the tests being used to arrive at those consequences still ranges from trifling to nonexistent. Perhaps, one might think, today’s inattention to tests’ instructional sensitivity simply stems from our not knowing how to go about determining the degree of a test’s sensitivity to instructional quality. Yet, we already have on hand a demonstrably successful strategy
drawn from our experiences in reducing the assessment bias found in our important educational tests. Let’s look at the chief elements of that strategy.

**Serious Problems Demand Serious Responses**

Rarely today is a significant educational test created for which considerable attention has not been devoted to the reduction of assessment bias. That is, we currently regard the diminishment of assessment bias as a canon of good test-building. But it was not always thus.

Go back to the 1960s and 1970s, and you will find that if any attention whatsoever was given to the reduction of tests’ assessment bias, it was apt to be perfunctory. Usually, it was completely absent. This skimpy attention to assessment bias was quite understandable. That’s because in those days we rarely analyzed test results in such a way as to reveal differences in performances among test-taker groups associated with their gender, race, or ethnicity. However, the rules of the educational testing game changed dramatically in the late seventies when a substantial number of states—dismayed by what they perceived to be the poor quality of their state’s public schools—began to link high-school graduation to a students’ passing “minimum competency tests” demonstrating that students possessed at least rudimentary skills in reading, mathematics, and sometimes writing.

Because those minimum competency tests were administered to all students in a state’s public schools, and those students’ scores were typically made public, we soon began to see astonishing disparities between the performances of racial groups as well as students drawn from different socioeconomic strata. Indeed, it was the difference in the racial pass rates on Florida’s diploma-denial tests that triggered a class-action lawsuit in the precedent-setting *Debra P. v. Turlington* case (Popham and Lindheim, 1981). In that case which, even now, remains the operative case law in such litigation, it was affirmed by a federal appellate court that a violation of the U.S. Constitution occurs when students are denied a property right (such as a high-school diploma) if they are tested using a test whose content had not been taught. In the Florida case, the precipitating circumstance was that far more African-American students were failing the state’s basic skills test than were white students. The *Debra P.* litigation, and similar disparities in racial pass rates elsewhere, presented a serious problem to America’s educational-measurement specialists. They quickly grasped the significance of the situation—and they set out to fix it.

**A Two-Pronged Bias-Reduction Strategy**

Having recognized the legitimacy of complaints that the nation’s tests were biased against certain subgroups, members of the measurement community soon devised a
two-tactic strategy to minimize such bias. The first of these two tactics was a during-development *judgmental review* of each test item in an effort to identify and eliminate any items thought to offend or unfairly penalize test-takers because of test-takers’ personal characteristics such as their gender or ethnicity. Second, an *empirical analysis* of students’ actual test performances was undertaken, usually during field-testing of new items, so that items potentially contributing to a test’s assessment bias could be spotted. The typical analytic approach that evolved, after several years of exploratory analyses, was to employ “differential item functioning” (DIF) techniques in which items were isolated that were being answered differently by different subgroups of test-takers. Items identified by DIF as possibly biased were then modified or jettisoned before being used in an operational test. As a consequence of employing this two-tactic strategy, over many years, we have witnessed a substantial reduction in the number of items on high-stakes tests that are biased against particular groups of test-takers.

The actual procedures for these two approaches to the reduction of assessment bias are now well known among measurement specialists. While their use may not have completely *eliminated* assessment bias from the nation’s high-stakes assessments, the marked impact of these procedures on the reduction of assessment bias is undisputed.

**Benign Borrowing**

The methodological strategy we could employ in reducing the instructional insensitivity of today’s evaluatively oriented achievement tests might be nothing more than a straight-out lift from what has been used in the reduction of assessment bias, that is, to employ a blend of judgmental and empirical procedures.

Although we currently do not have a definite, well-honed set of procedures for dealing with the instructional sensitivity of our tests, the essential elements of an attack on this problem could be directly derivative from previous work in minimizing assessment bias. For example, the charge to be issued when asking a group of seasoned educators to scrutinize a set of test items for instructional *insensitivity* could be quite similar to the language employed when we ask a committee of bias reviewers to look for biased elements in test items. To illustrate, a review committee composed of experienced teachers (who are thoroughly familiar with the content and age-levels of the students to be tested) could be oriented to their item-review responsibilities by learning about the most likely ways an item might be instructionally insensitive. After such an orientation, reviewers could then be given the following charge and asked to render a per-item judgment regarding each item intended for inclusion in a high-stakes evaluative test:

> **Attention reviewers:** Please note the specific curricular aim which, according to the test’s developers, this item is assessing. Only then, answer the following
question: *If a teacher has provided reasonably effective instruction to promote students’ mastery of the specific curricular aim being assessed, it is likely that the bulk of the teacher’s students will answer this item correctly?* (Choose one: YES, NO, NOT SURE). (Popham, 2014, 397)

Items for which one or more reviewers have supplied a specified proportion of negative and/or not-sure responses would then be scrutinized to discern if the items embody elements apt to render them instructionally insensitive. Such items would, as is true when acting on the judgments of bias-review committees, be revised or removed.

Similarly, procedural elements for carrying out empirical DIF-like studies for instructional sensitivity must surely be generated and refined. The overriding thrust of such DIF analyses is to identify two groups of teachers who, for item-analysis purposes, are decisively different in their demonstrated effectiveness in bringing about improvements in students’ assessed achievement levels. Having identified two extreme groups of teachers on the basis of, for instance, their students’ performances for several previous years’ worth of annual assessments, we can then see if those teachers’ current students’ responses to a new set of items are consonant with what would be predicted. For example, if students taught by lower-effectiveness teachers actually perform better on particular items than students taught by higher-effectiveness teachers, then those items should certainly be subjected to serious scrutiny to discern what seems to be rendering them instructionally insensitive. Although Joseph Ryan and I (Popham and Ryan, 2012) have proposed one use of DIF procedures using student-growth-percentiles to carry out item-sensitivity analyses, much more exploratory work on this problem should be undertaken.

As with the reduction of assessment bias in high-stakes educational tests, the implementation of the previously described two-tactic strategy for dealing with instructional sensitivity will not transform instructionally insensitive tests, overnight, into assessment that reek of instructional sensitivity. But our colleagues who coped with assessment bias have given us a set of MapQuest.com directions for making our evaluative tests *more* instructionally sensitive. And progress in that direction, of course, will increase not only the validity of test-based inferences about instructional quality, but also the subsequent decisions we make about the teachers or schools being evaluated.

**A Discontented Winter**

“Now is the winter of our discontent . . .” are the initial seven words of Shakespeare’s *Richard the Third*. Well, it is currently winter and I am definitely discontented. I find it altogether intolerable to be a member of a measurement clan that allows hugely important educational decisions to be made on the basis of students’ scores on tests
not demonstrated to be suitable for their evaluative applications. How can we let such misuses continue? How can we, in good conscience, permit our nation’s educational leaders and policymakers to rely on test results that may be completely unsuitable for the purposes to which they are being put? How can we allow teachers to be fired because of students’ scores on the wrong tests? How can we? And yet we do.

The only way to begin changing an indefensible practice is to set out seriously to alter that practice. It is time, indeed past-time, for those of us who recognize the seriousness of this situation to don our alteration armor and head into battle.

References


Update on Evaluation of NAEP Achievement Levels Procurement

Objective
To receive a brief informational update from NCES on the current status of the procurement being planned to evaluate NAEP achievement levels. Ongoing updates will be provided at each COSDAM meeting.

Background
The NAEP legislation states:

The achievement levels shall be used on a trial basis until the Commissioner for Education Statistics determines, as a result of an evaluation under subsection (f), that such levels are reasonable, valid, and informative to the public.

In providing further detail, the aforementioned subsection (f) outlines:

(1) REVIEW-

A. IN GENERAL- The Secretary shall provide for continuing review of any assessment authorized under this section, and student achievement levels, by one or more professional assessment evaluation organizations.

B. ISSUES ADDRESSED- Such continuing review shall address--

(i) whether any authorized assessment is properly administered, produces high quality data that are valid and reliable, is consistent with relevant widely accepted professional assessment standards, and produces data on student achievement that are not otherwise available to the State (other than data comparing participating States to each other and the Nation);

(ii) whether student achievement levels are reasonable, valid, reliable, and informative to the public;

(iii) whether any authorized assessment is being administered as a random sample and is reporting the trends in academic achievement in a valid and reliable manner in the subject areas being assessed;

(iv) whether any of the test questions are biased, as described in section 302(e)(4); and

(v) whether the appropriate authorized assessments are measuring, consistent with this section, reading ability and mathematical knowledge.

(2) REPORT- The Secretary shall report to the Committee on Education and the Workforce of the House of Representatives and the Committee on Health,
Education, Labor, and Pensions of the Senate, the President, and the Nation on the findings and recommendations of such reviews.

(3) USE OF FINDINGS AND RECOMMENDATIONS- The Commissioner for Education Statistics and the National Assessment Governing Board shall consider the findings and recommendations of such reviews in designing the competition to select the organization, or organizations, through which the Commissioner for Education Statistics carries out the National Assessment.

Responsively, a procurement has been planned to administer an evaluation of NAEP achievement levels. The last update COSDAM reviewed on this topic was in August 2013.

In this brief written update, NCES provides the Committee with a summary of the status of this procurement.
Evaluation of NAEP Achievement Levels

The National Center for Education Evaluation and Regional Assistance (NCEERRA), part of the Institute for Education Sciences (IES), will administer the Evaluation of the NAEP Achievement Levels. NCEERRA and the Department of Education’s Contracts and Acquisitions Management (CAM) office will begin this procurement during fiscal year 2014. Tentatively, NCEERRA will deliver the Request for Comments (RFC) package to CAM in December 2013 and the scheduled award date is June 2014. This will be a full and open competition.
Reading for Understanding: A Theory-Based, Developmental Approach

Objective: To provide a brief overview of an IES grant on assessment innovations.

Background

During the August 2013 COSDAM meeting, Committee members were invited to provide comments on “Other issues or questions.” John Easton noted that there is an IES grant on assessment innovations for parsing out prior knowledge. Committee members expressed interest in hearing more about this project. In this brief overview, NCES provides the Committee with a description of the Reading for Understanding grant.
Reading for Understanding: A Theory-Based, Developmental Approach

IES Grant Award Number: R305F100005
Principal Investigators: John Sabatini, Tenaha O’Reilly (ETS)

Project Goals: Develop a series of age-appropriate, developmentally-sensitive, and theoretically-based summative reading comprehension assessments.

Population: The assessments are intended for students in grades Pre-Kindergarten through twelfth grade.

Administration: All assessments are computer delivered and take approximately 45-60 minutes to administer.

The Assessments: Our system incorporates two types of assessments that are designed to measure the components of reading as well as higher-level comprehension.

Component assessments: The first type of assessment is designed to measure the components of reading including decoding, phonological awareness, word recognition, morphology, syntax, vocabulary, listening comprehension and spelling. The components assessment is designed to help contextualize and interpret performance on the Global Integrated Scenario-based Assessment (GISA).

GISA: The second type of assessment, the GISA, is designed to measure a set of integrated skills associated with higher-level comprehension. Students are presented with a realistic purpose for reading that requires them to integrate, synthesize, and evaluate a collection of diverse reading materials (e.g., blog, website). Tasks and activities are sequenced to model complex thinking while simultaneously collecting evidence of partial understanding for developing students. Scaffolding techniques coupled with simulated peer interactions are designed to promote the social nature of reading and the structured nature of learning. Particular emphasis was put on measuring and accounting for variables known to affect reading comprehension, but seldom measured in a summative reading assessment. These variables include background knowledge, student motivation, self-regulation/metacognition, disciplinary reading, learning, and reading strategies.

Technical information: To date, both types of assessments have been piloted in 23 states in a mix of urban, suburban and rural areas. We have tested over 50,000 students on our GISA assessments and over 115,000 students on our component assessments. Preliminary analyses reveal that our GISA assessments are demonstrating good reliability (typically $\alpha=.80$ or higher). Our component assessments tend to have even higher reliability (typically $\alpha=.90$ or higher). We have demonstrated validity evidence (e.g., eye tracking data) and answered a number of key research questions. For example, analyses are uncovering key relationships between components and comprehension, dimensionality of the measures, the role of background knowledge and motivation in testing, the ability of students to learn new information during a reading test, the added value of constructed response items and the relationships between local and global comprehension processes.
INTRODUCTION AND BACKGROUND

The College Course Content Analysis (CCCA) study is one of a series of studies contributing to the National Assessment of Educational Progress (NAEP) Program of 12th Grade Preparedness Research conducted by the National Assessment Governing Board (NAGB). The purpose of the CCCA study is to identify a comprehensive list of the reading and mathematics knowledge, skills, and abilities (KSAs) that are pre-requisite to entry-level college mathematics courses and courses that require college level reading based on information from a representative sample of U.S. colleges. The Educational Policy Improvement Center (EPIC) is the contractor working for the Board to conduct this study.

Another goal of the CCCA study is to extend the work of the two previous preparedness studies—the Judgmental Standards Setting (JSS) study, implemented in 2011 and the Job Training Program Curriculum (JTPC) study, implemented in 2012. The CCCA study is designed so the results can be compared to the JSS and JTPC studies, reporting on how this new information confirms or extends interpretations of those earlier studies. The design of the CCCA study is based on the JTPC study but with modifications based on the lessons learned.

The CCCA study will answer four core research questions.

1. What are the 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?

The final report is due May 2014, and until then COSDAM will receive detailed reports at each Board meeting.
METHODOLOGY
The Design Document for the CCCA study is complete. It provides guidance for the study by describing:

- Criteria for collecting courses and artifacts;
- A sampling plan to comprise a representative sample of institutions;
- Review and rating processes, including a training plan and process for ensuring reviewer effectiveness and consistency; and
- The process for ensuring reliability across reviewers providing artifact analysis.

This study comprises three primary phases:
1. Identification and collection of course artifacts,
2. Review of course artifacts by Review Teams, and
3. Analysis and reporting.

OVERVIEW OF ACTIVITIES BY PHASE
Phase 1: Identification and collection of course artifacts

In the CCCA study, a course artifact is defined as a syllabus, a non-textbook based assignment or assessment, and textbook excerpt. In mathematics, there are some instances where the only specifically identified assignments were listed in the syllabus and were from the textbook. In those cases, a textbook based assignment or assessment was allowed. The CCCA sample of artifacts is derived from extant artifacts and combined with newly gathered course artifacts. Extant artifacts contributing to the CCCA sample were extracted from EPIC’s repository of artifacts compiled during previous research on entry-level curricula at postsecondary educational institutions. Project staff solicited new course artifacts as needed to create a complete and nationally representative sample.

EPIC identified a set of inclusion criteria that courses must meet to be included in the CCCA study as well as a set of institutional characteristics of which the final CCCA Artifact Bank must be representative. The final CCCA Artifact Bank comprises a set of courses and artifacts that are to be used as the basis for the content reviews to be conducted by mathematics and reading content review teams in the second phase of the study.

Phase 1 preparatory work also included the convening of NAEP advisory panels, for reading and mathematics respectively, to obtain content-based guidance and recommendations. In these meetings, preliminary coding schemas, training materials and decision rules were reviewed. NAEP advisors also reviewed all of the course packets to be used in validation data analyses, training sessions, and determining sufficient reviewer competence (qualifying). Guidance from these NAEP advisory panels was integrated into the implementation of the study.
Phase 2: Review of course artifacts by Review Teams

In Phase 2, content reviewers are recruited and training materials are developed in preparation for the review of course artifacts and the content reviews are conducted. Content reviewers are first trained to review the course packets from a “holistic” perspective and identify prerequisite mathematics and reading KSAs. In the second independent review training, the NAEP frameworks for grade 12 reading and mathematics are used as a basis for coding the packets. All additional KSAs beyond the NAEP frameworks are documented and included in all successive reviews, comparisons and data analyses. The overarching goal of the CCCA study is to identify all prerequisite KSAs, not just those KSAs associated with the NAEP frameworks.

Subsets of the course artifact packets were set aside to serve as training packets and qualifying packets. These packets are annotated by the NAEP advisory panel members for use as exemplars of expert coding. After the holistic reviews, content reviewers are trained with respect to the NAEP frameworks, and as part of the training process, the reviewers code the training packets with respect to the NAEP frameworks in small groups. Then, the reviewers code qualifying packets independently. EPIC project staff then compares and scores this coding with respect to the exemplars provided by the NAEP advisory panel. If a reviewer scores below a certain threshold, retraining is provided. Reviewers who receive a second low score are not invited to participate in the study. Qualified reviewers proceed to the next stage: coding 28 course artifact packets independently. Group review meetings are then held to discuss discrepancies identified in independent reviews.

NAEP experts attend the group review meetings as on-site assistance, answering questions about the NAEP framework as they arise. Validity checks are also embedded in the group review process. Validity packets are annotated by the NAEP experts at the advisory panel meetings to be used as reference coding. Those packets are reviewed by all content reviewers without the knowledge that the packets were for validity purposes. This provides the opportunity for evaluating the reliability of the review team coding. The percent agreement between the four review teams’ group consensus coding on the validation packets and the NAEP reference coding as reliability evidence will be calculated within each course title and across course titles.

In summary, the CCCA Study’s Phase 2 combines independent individual judgments with panel processes. The primary goal of the second, or group, review is to adjudicate differences where possible in coding of the packets completed during the independent review and to produce group-level coding of the additional prerequisite KSAs that were not found in the NAEP frameworks. The final result of this two-part review process is a comprehensive list of prerequisite KSAs, answering the Board’s first research question: what are the prerequisite KSAs in reading and mathematics to qualify for entry-level, credit-bearing courses that satisfy general education requirements?
The final step in Phase 2 is for the NAEP experts to review the results of the KSA prerequisite data collected from the content reviewers, which will be summarized in content maps. The NAEP experts’ primary task is to compare these data with the 12th grade NAEP 2009 and 2013 items, achievement level descriptions, and minimal academic preparedness descriptions (from the JSS studies) in both mathematics and reading.

**Phase 3: Analysis and reporting**

Phase 3 includes processing and analyzing the judgments collected during the review of course artifacts by review teams, and preparing the data to be reported in ways that are directly responsive to research questions in accordance with the analysis plan specified within the Design Document. Standard statistical methods and metrics necessary will provide evidence of validity and reliability, and both conceptual (information processing/document analysis) and technical (quantitative) analyses will be conducted. The CCCA study is structured to provide a fully crossed, three factor design to ensure that results can be reviewed in statistical generalizability analyses, which will allow us to evaluate the reliability of the study design.

Final results will include narrative summaries of the prerequisite knowledge, skills, and abilities in mathematics and reading. Summary analyses will also address all aspects of the CCCA study design (see Illustration 1). As project elements are completed, the appropriate sections of Illustration 1 are shaded in dark gray. Project elements that have begun and are in progress are shaded in a lighter gray. Those project elements that have just begun have no shading in the diagram.
Update on College Course Content Analysis and Research with Frameworks Projects

Illustration 1: Project Design
Attachment H
Update on College Course Content Analysis and Research with Frameworks Projects

Illustration 2 displays a schedule of the CCCA study. As meetings or events are completed, they are noted and shaded in dark gray.

Illustration 2: CCCA Study Gantt chart

<table>
<thead>
<tr>
<th>MEETING OR EVENT</th>
<th>Start Date</th>
<th>End Date</th>
<th>Duration</th>
<th>Quarter 2</th>
<th>Quarter 3</th>
<th>Quarter 4</th>
<th>Quarter 1</th>
<th>FINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APR</td>
<td>MAY</td>
<td>JUN</td>
<td>JUL</td>
<td>AUG</td>
</tr>
<tr>
<td>NAEP Technical Panel</td>
<td>21-Jun</td>
<td>23-Jun</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meeting</td>
<td>5-Jun</td>
<td>9-Jun</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAEP Technical Panel</td>
<td>8-Jul</td>
<td>12-Jul</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holistic Training</td>
<td>8-Jul</td>
<td>12-Jul</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Content</td>
<td>8-Jul</td>
<td>12-Jul</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Content NAEP</td>
<td>22-Jul</td>
<td>24-Jul</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>22-Jul</td>
<td>24-Jul</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent Content</td>
<td>9-Jul</td>
<td>16-Aug</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviews</td>
<td>19-Aug</td>
<td>6-Sep</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Review Meeting</td>
<td>26-Sep</td>
<td>29-Sep</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meeting 2</td>
<td>3-Oct</td>
<td>9-Oct</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPIC Data Analysis</td>
<td>29-Sep</td>
<td>15-Nov</td>
<td>COMPLETE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period 2</td>
<td>3-Jan</td>
<td>5-Jan</td>
<td>3 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAEP Math Expert</td>
<td>15-Nov</td>
<td>17-Nov</td>
<td>3 days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review Meeting</td>
<td>1-Feb</td>
<td>28-Mar</td>
<td>7.5 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Report Writing</td>
<td>1-Apr</td>
<td>30-Apr</td>
<td>4+ weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent Content</td>
<td>30-Apr</td>
<td>30-Apr</td>
<td>FINAL DELIVERY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PROGRESS UPDATE
Phase 1, Identification and collection of course artifacts, is complete. For Phase 2, Review of course artifacts by Review Teams, the work of the content reviews, independent and group reviews, is complete. The outcome of the content review is a list of prerequisite KSAs, including additional KSAs and those associated with the NAEP frameworks. This list of prerequisite KSAs, in the form of content maps, is the basis of the upcoming NAEP content expert review. Phase 3, Analysis and Reporting, is ongoing. Data are being compiled from the group reviews, the NAEP expert reviews and generalizability analyses.

Independent Content Review (Phase 2)

Content reviews included training, the independent review and group review by mathematics and reading content reviewers. Content reviewers, separated into four mathematics and four reading groups, conducted reviews of 20 course packets and 8 validity packets. The initial holistic review was conducted to elicit additional KSA for each course packet without the influence of the NAEP framework, and familiarized the content reviewers with the review process and course
The independent review generated the following data that is the basis of the group content review:

- Applicability and importance coding on all NAEP framework objective level statements
- KSA exclusions on relevant NAEP framework objective level statements (i.e., phrases in the NAEP framework objectives that were not applicable)
- List of additional KSAs evident within the course packets

Thirty-two mathematics and reading content experts participated in training for independent review. Training consisted of the following:

- Attend Holistic Review training webinar
- Complete Holistic Review of 28 course packets
- Attend Independent (NAEP) Review training webinar
- Complete Training Packet #1
- Review feedback and attend re-training (as necessary)
- Complete Training Packet #2 (optional)
- Complete Qualifying Packet #1
- Review Scoring and attend re-training (as necessary)
- Complete Qualifying Packet #2 (as necessary)

Content Reviewers were required to reach a level of coding consistency with NAEP advisory panel coding on a set of qualifying course packets in order to proceed to independent (NAEP framework) review. One content reviewer did not reach the level required and was released from further work on the CCCA project.

Twenty-nine content reviewers completed the independent (NAEP framework) review. Twenty-four of content reviewers were asked to attend the group review meetings. Consistency of content reviewer validity packet coding with the NAEP advisory panel coding was the main factor to determine whether a reviewer was selected to attend the group review meetings.

**Group Review (Phase 2)**

Two group review sessions were scheduled upon completion of the independent (NAEP framework) review in the Portland, Oregon metropolitan area. Two mathematics and two reading groups attended each session.
A technical working group, familiar with the CCCA project, considered the multiple content area configurations of the group review meetings and potential threats to the procedural validity of the study. It was determined that no threats to validity existed as all decision rules were finalized prior to the group review sessions. Also, content-area calibration in both mathematics and reading primarily occurred during training and through the review process.

EPIC staff, trained in facilitation and data collection, was engaged to facilitate the group review process. Two required training sessions were convened for five facilitators and five scribes. The first training session focused on the project overview and independent review materials and data. The second session instructed facilitators and scribes on roles, data collection process, and survey instrument instruction.

The objective of the group review meeting sessions was to determine if consensus could be reached on discrepant coding across reviewers in how well the KSAs described in the NAEP framework align with the KSAs evident in course materials. All three members of each review group coded course packets for evidence of mathematics or reading KSAs during independent review by applying the decision rules and using their expert judgment based on evidence in the course packet. At the group review, the same groups discussed and reached consensus on discrepant applicability, importance coding and related KSA exclusions.

*Decision points* are the number of decisions a group of reviewers was asked to make over the course of the independent review and to come to consensus on in the group review. The number of points is calculated by multiplying the number of packets (28) times the sum of the decisions applicability/importance coding (130 for mathematics or 37 for reading) and KSA exclusions times the number of reviewers.
Illustration 4: Distribution of Coding Decisions

<table>
<thead>
<tr>
<th>Group</th>
<th>NAEP KSA Coding Decision Points</th>
<th>Points of Discussion/ KSA Coding Discrepancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics Group 1</td>
<td>21,840</td>
<td>1,467</td>
</tr>
<tr>
<td>Mathematics Group 2</td>
<td>21,840</td>
<td>771</td>
</tr>
<tr>
<td>Mathematics Group 3</td>
<td>21,840</td>
<td>959</td>
</tr>
<tr>
<td>Mathematics Group 4</td>
<td>21,840</td>
<td>971</td>
</tr>
<tr>
<td>Reading Group 1</td>
<td>6,216</td>
<td>965</td>
</tr>
<tr>
<td>Reading Group 2</td>
<td>6,216</td>
<td>827</td>
</tr>
<tr>
<td>Reading Group 3</td>
<td>6,216</td>
<td>883</td>
</tr>
<tr>
<td>Reading Group 4</td>
<td>6,216</td>
<td>938</td>
</tr>
</tbody>
</table>

In the group reviews, coding for applicability and importance of KSAs were reviewed. Reviewers also examined their areas of agreement and as well as phrases in the NAEP framework objectives that were not applicable, i.e., partial matches to NAEP framework objectives (termed “KSA exclusions”). All of the coding is the basis for, and summarized in, content maps to be used for at the NAEP expert review meetings, which are being held in November 2013 and January 2014.

**Analysis Conducted to Date (Phase 3)**

EPIC conducted a fully-crossed generalizability study on the independent review data in order to determine the inter-rater reliability (i.e., consistency of the reviewers’ ratings) on the NAEP standards and framework objectives (KSA) for both reading and mathematics. Generalizability analyses allow analysts to disentangle the contributions made to measurement error by different facets. EPIC analyzed three facets for their contributors to the variance in coding: individual reviewer, NAEP standard, and packet for reading and mathematics. As the G and Phi coefficients approach 1.0, consistency increases; coefficients between .70 and 1.0 are in the acceptable range. Because there are more NAEP standards and framework objectives in mathematics than in reading, EPIC anticipated that the G and Phi coefficients for mathematics would be smaller than for reading, however preliminary results indicate that raters consistently rated the packets at both the standard and objective level for both mathematics and reading.

Content maps have been prepared from the group review data for both mathematics and reading to show the coding provided for each KSA across course packet and course title. Content maps will be generated in spreadsheet form and will be incorporated into a narrative document during the NAEP review material preparation.
Preparation for analysis and the final reporting have begun with the majority of the effort in data management. Staff are working with sample data and testing to ensure that accurate data collection protocols and routines of effective quality control, data cleaning procedures and data storage/security protocols are in place and use.

The final report is also underway. The table of contents has been established and preliminary table shells have been drafted.

**STATUS SUMMARY**

The first phase of the study is complete. The course artifacts have been identified, all artifacts have been collected, review packets have been created from those artifacts, and the course packets were reviewed by content reviewers independently and then again in group review meetings.

The second phase of the study is nearing completion. Based on guidance from NAEP advisory panels, in both reading and mathematics, feedback was integrated into the content review training and coding schemes and the overall approach to training. Content reviewers were trained in two sessions and required to obtain an acceptable score on training and qualifying packets prior to beginning the process of remote independent content reviews. Next, trained facilitators managed a process to determine and record group level coding of the course packets at onsite group review meetings in September. Project staff will facilitate the comparison work of the NAEP experts at onsite meetings in November 2013 and January 2014. Process evaluations were conducted after training, after independent review and after the group review meetings. Evaluations were largely positive.

The third phase of the study has begun. The data from the independent reviews was compiled for presentation at the onsite group review meeting using online recording tools. A generalizability analysis was conducted on the independent coding data to quantify the variance that certain factors contribute to the dataset.

The data from the group reviews is being compiled for presentation at meetings of the NAEP mathematics and reading content experts. These data are also being used in a generalizability analysis on the group coding data. Preparation of the final report is ongoing.
Evaluating Reading and Mathematics Frameworks and Item Pools as Measures of Academic Preparedness for College and Job Training

Project Status Update
Contract ED-NAG-13-C-0001

The National Assessment Governing Board contracted with the Human Resources Research Organization (HumRRO) in June 2013 to conduct three tasks related to research on 12th grade preparedness:

1. **Evaluation of the Alignment of Grade 8 and Grade 12 NAEP to an Established Measure of Job Preparedness:** This study will extend prior analysis of the relation of NAEP to measures such as WorkKeys by including the NAEP grade 8 assessments and by expanding the method for assessing content alignment. The study method will follow the Governing Board content alignment design document for preparedness research studies, with some modifications. The two-pronged approach includes alignment of:
   - (a) the training preparedness measure to the NAEP frameworks; and
   - (b) NAEP items to the framework from which the training preparedness measure was developed.

2. **O*NET Linkage Study:** This study is a content validity investigation. Major duties (MDs) for the five target occupations will be identified. The occupations are automotive master technician, computer support specialist, HVAC technician, licensed practical nurse, and pharmacy technician. Expert raters will link NAEP content to MDs; NAEP content to O*NET knowledge, skills, and abilities (KSAs); and O*NET KSAs to MDs. This study will identify any disconnects between the level of constructs measured by NAEP and the level of those constructs required for entry into job training programs.

3. **Technical Advisory Panel (TAP) Symposium:** As part of the current contract, HumRRO assembled a technical advisory panel (TAP) of five experts in educational measurement and five experts in industrial-organizational (I-O) psychology to review extant research and to generate ideas for commissioned papers on preparedness. Each panelist is being asked to propose a paper that he/she could develop. Governing Board staff and members will review the proposals and commission up to 10 papers. Panelists will have several months to develop the papers, after which the TAP will reconvene in a late 2014 symposium. Authors will present their papers and the entire panel will discuss implications for preparedness research. HumRRO will produce a proceedings document summarizing the commissioned papers and discussion. (A list of TAP members is included on the next page.)

In addition, HumRRO will produce a comprehensive report at the conclusion of the contract in December 2014.
Work completed to date:

**O*NET Linkage:** HumRRO has developed lists of major duties (MDs) for each occupation based on O*NET task lists and course objectives from training curricula. Content experts in each occupation have reviewed and vetted the MDs. We also have obtained NAEP items for the O*NET Linkage Study. We are currently assembling materials for conducting the linking exercise.

**TAP Symposium:** The initial Brainstorming Meeting of the TAP was convened on October 25, 2013 in Crystal City, VA. Panelists will submit paper proposals in late 2013.

**Technical Advisory Panel (TAP) Members**

**John Campbell**  
Professor of Psychology  
University of Minnesota  
(Member, NAGB Technical Panel on 12th Grade Preparedness Research, 2007-2008)

**Michael Campion**  
Herman C. Krannert  
Professor of Management  
Purdue University

**Gregory Cizek**  
Professor of Educational Measurement and Evaluation  
University of North Carolina at Chapel Hill

**Brian Gong**  
Executive Director of Center for Assessment National Center for the Improvement of Educational Assessment, Inc.

**Ronald Hambleton**  
Distinguished University Professor, Educational Policy, Research, & Administration  
Executive Director, Center for Educational Assessment  
University of Massachusetts at Amherst

**Suzanne Lane**  
Professor, Research Methodology  
University of Pittsburgh School of Education

**Kenneth Pearlman**  
Independent Consultant in Industrial-Organizational Psychology  
Sarasota, FL

**Barbara Plake**  
University Distinguished Professor, Emeritus  
University of Nebraska-Lincoln

**Ann Marie Ryan**  
Professor of Psychology  
Michigan State University

**Nancy Tippins**  
Senior Vice President  
CEB Valtera
# National Assessment Governing Board
## Reporting and Dissemination Committee

### December 6, 2013
10:00 a.m.-12:45 p.m.

## AGENDA

<table>
<thead>
<tr>
<th>Time</th>
<th>Item</th>
<th>Presenter(s)</th>
<th>Attachment</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00 – 10:45 am</td>
<td>NAEP Testing and Reporting on Students with Disabilities and English Language Learners</td>
<td>Larry Feinberg, NAGB Staff; Grady Wilburn, NCES</td>
<td>Attachment A</td>
</tr>
<tr>
<td></td>
<td><strong>[Joint meeting with COSDAM]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:45 – 11:15 am</td>
<td>Embargo Policy for National Assessment Reports</td>
<td>NAGB Staff</td>
<td>Attachment B</td>
</tr>
<tr>
<td>11:15 – 11:40 am</td>
<td>Board and Committee Input into NAEP Reports</td>
<td>Stephaan Harris, NAGB Staff</td>
<td>Attachment C</td>
</tr>
<tr>
<td>11:40 – 11:50 am</td>
<td>Review of NAEP 2013 Reading and Mathematics Release</td>
<td>Stephaan Harris, NAGB Staff; Amy Buckley, Reingold</td>
<td>Attachment D</td>
</tr>
<tr>
<td>11:50 – 11:55 am</td>
<td><strong>ACTION:</strong> Release Plan for 2013 TUDA Report Cards</td>
<td>Stephaan Harris, NAGB Staff</td>
<td>Attachment E</td>
</tr>
<tr>
<td>11:55 am – 12:10 pm</td>
<td>Board Chairman’s Charge to the Committee for 2014</td>
<td>David Driscoll, NAGB Chair</td>
<td></td>
</tr>
<tr>
<td>12:10 pm – 12:45 pm</td>
<td>NAEP Contextual Variables</td>
<td>a. <strong>ACTION:</strong> Contextual Information Framework</td>
<td>Attachment F</td>
</tr>
<tr>
<td></td>
<td>b. Using NAEP for Education Indicators</td>
<td>Alan Ginsburg and Marshall Smith, Consultants</td>
<td>Attachment G</td>
</tr>
<tr>
<td></td>
<td><strong>[Joint Meeting with ADC]</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Information Items</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Projected Schedule for Future NAEP Reports</td>
<td></td>
<td>Attachment H</td>
</tr>
<tr>
<td></td>
<td>• NAGB Parent Summit</td>
<td></td>
<td>Attachment I</td>
</tr>
</tbody>
</table>


NAEP Testing and Reporting on Students with Disabilities and English Language Learners

The inclusion of students with disabilities (SD) and English language learners (ELL) is key to the success of the Nation’s Report Card. It is important that we measure all of the students in NAEP jurisdictions to give the most accurate picture of student knowledge and skills in those areas. The National Assessment Governing Board, believing that the validity of NAEP results has been threatened by high exclusion rates and substantial variations from state-to-state, has adopted a policy to promote increased inclusion and testing of these two important groups.

Since the policy was enacted in March 2010, we have seen inclusion rates improve across all subjects and grades. For example, ten years ago in 4th grade reading 19 states would not have met the 95 percent inclusion goal that the Board policy set. In 2013, that number dropped to one state. In fact, in 2013 only one state—Maryland in both 4th and 8th grade reading—did not meet the 95 percent inclusion goal across all grades/subjects. The exclusion rates in 2013 are the lowest since state NAEP began more than 20 years ago.

The December presentation to the joint committees of COSDAM and Reporting and Dissemination will focus on (1) improvements resulting in the lowest exclusion rates in NAEP’s history, and (2) NCES plans for implementing the follow-up actions endorsed at the Governing Board meeting in May 2013. These include the following:

• giving greater prominence to full population estimates, which present adjusted scores that seek to account for differences in students excluded,

• providing additional information online on the proportion of students excluded because they use an accommodation NAEP does not allow, and

• developing a special report after the main data release that includes a full discussion of exclusion, participation and refusal issues, and the total participation rate for each jurisdiction in NAEP.
The following materials are attached:

A.1 Governing Board Policy on NAEP Testing and Reporting on Students with Disabilities and English Language Learners – Adopted March 6, 2010

A.2 Action on reporting SD and ELL data as part of 2013 NAEP results—Joint meeting of Committee on Standards, Design and Methodology and Reporting and Dissemination—May 17, 2013

A.3 State-by-state data on exclusions, including proportion of students excluded because they use an accommodation on state tests that is not permitted by NAEP [Available in Excel]

A.4 NAEP Accommodations—Permitted and Used (2013)

A.5 Read-aloud accommodation

A.6 Calculator accommodation

A.7 Full-population estimates

A.8 NAEP 2013 data on students assessed, accommodated, excluded, and absent, including total participation rate [Available in PDF and Excel]

A.9 Former English-language learners
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.
POLICY PRINCIPLES

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.
National Assessment Governing Board
Committee on Standards, Design and Methodology

May 17, 2013
EXCERPT

JOINT MEETING WITH REPORTING AND DISSEMINATION COMMITTEE

NAEP Testing and Reporting on Students with Disabilities and English Language Learners

The Committees considered two sets of issues: (1) implementation of the Board policy that deals with testing of English language learners (ELL), and (2) reporting options for exclusions, participation rates, and the adjusted scores (termed full-population estimates), which try to take into account the differences in exclusion rates between the states and districts participating in NAEP.

Grady Wilburn, of NCES, gave an update to the committees on these two topics. He noted that under the policy adopted in 2010, the only ELL students that schools may exclude from NAEP are those who have been in U.S. schools for less than one academic year. Even students in this category should not be excluded if NAEP offers a translation in their home language. Students who speak Spanish now account for about 80 percent of ELLs nationwide. NAEP offers Spanish translations of all its tests (in bilingual booklets) except for reading and writing, which under the frameworks adopted by the Board are reading and writing in English.

Mr. Wilburn said the decision tree incorporating the Board policy on which ELL students to test and how to test them was implemented smoothly in the 2013 NAEP. He said exclusion rates went down.

On the reporting issues Mr. Wilburn noted that a joint meeting of the two committees in March had received a full report on implementation of the policy on SD students. Under this policy the only students that may be excluded from NAEP by school personnel are those with the most significant cognitive disabilities who take alternate state assessments with alternate achievement standards, expected to be about 1 percent of enrollment. For practical reasons NCES decided that schools could also continue to exclude students with an individualized education program (IEP) or 504 plan that provides for accommodations on state tests that NAEP does not allow. The non-allowable accommodations in nearly all cases have been read-aloud on the NAEP reading assessment or calculator use on all sections of NAEP math.

In 2013 for the first time NCES permitted students with an IEP requiring calculator use to take calculator-active blocks on NAEP even if they would have been assigned non-calculator blocks as part of the normal NAEP sampling. Mr. Wilburn said a study in 2011 indicated that this would have little impact on results, and, in any case, the number of students involved is small.
By law, student participation in NAEP is voluntary. Parents can refuse to have their children participate for any reason. Under the Board policy, “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.”

NCES has said doing this would break trends, depress reported scores, and contravene sound psychometric procedures. Under long-standing practice, excluded students are omitted from any calculations of NAEP results, and have no effect on state or district averages. Adjustments are made for refusals or absent students (a much larger group) by reweighting the scores of those with similar characteristics, which tends to lower state and district averages.

There is another analytic procedure, called full-population estimates (FPE), which NCES has used for about a decade to adjust state and district results by imputing scores for excluded SD and ELL students based on the performance of similar SD and ELL students who are tested. Data showing year-to-year changes in the full-population estimates are published on the NAEP website for participating states and districts, but these are given little prominence and do not include the adjusted scores themselves. The FPE scores were provided to the Board at this meeting. They showed most state averages to be about 3 to 6 points lower than reported. In only a few cases were year-to-year changes significantly different.

George Bohrnstedt, of AIR, chair of the NAEP Validity Studies Panel, said his group is concluding a study which shows that FPEs provide less biased results than the current NAEP analysis method, which overstates true scores considerably more.

Another proposal considered for reporting is to publish a total participation rate, based on all students in a sample divided into those tested and not tested for any reason whether excluded, absent, or refused. At present the reported participation rates are calculated after excluded students are subtracted from the number in the sample.

After considerable discussion, the Committees endorsed the following in regard to the reporting of 2013 NAEP results:

1. Continue previous analysis procedures for exclusions and refusals.

2. Give greater prominence and easier accessibility to full-population estimates as part of the information available online at the time of data release.

3. Provide additional information online on the proportion of students excluded because they use an accommodation on state tests that is not allowed on NAEP.

4. Issue a special report after the main data release with a full discussion of exclusion, participation, and refusal issues that includes data on the total participation rate for each jurisdiction in NAEP.
Attachment A3 - Excel file: State-by-state data on exclusions, including proportion of students excluded because they use an accommodation on state tests that is not permitted by NAEP
<table>
<thead>
<tr>
<th>Type of accommodation</th>
<th>SD and/or ELL</th>
<th>SD</th>
<th>ELL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual dictionary</td>
<td>0.8</td>
<td>0.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Braille presentation</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Braille response</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Breaks</td>
<td>4.1</td>
<td>3.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Calculator</td>
<td>1.1</td>
<td>1.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Cue to stay on task</td>
<td>1.2</td>
<td>1.1</td>
<td>0.2</td>
</tr>
<tr>
<td>Directions read aloud in English</td>
<td>3.3</td>
<td>2.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Directions read aloud in Spanish</td>
<td>0.1</td>
<td>#</td>
<td>0.1</td>
</tr>
<tr>
<td>Extended time</td>
<td>11.4</td>
<td>8.1</td>
<td>4.4</td>
</tr>
<tr>
<td>Large-print booklet</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Magnification device</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>One-on-one</td>
<td>0.6</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>Read aloud (all)</td>
<td>5.7</td>
<td>5.0</td>
<td>1.5</td>
</tr>
<tr>
<td>Read aloud (occasional)</td>
<td>1.3</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Read aloud in Spanish</td>
<td>0.1</td>
<td>#</td>
<td>0.1</td>
</tr>
<tr>
<td>School staff administers</td>
<td>0.5</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Scribe</td>
<td>0.5</td>
<td>0.5</td>
<td>#</td>
</tr>
<tr>
<td>Sign language presentation</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Sign language response</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Small group</td>
<td>9.8</td>
<td>8.0</td>
<td>2.9</td>
</tr>
<tr>
<td>Spanish-English booklet</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Special equipment</td>
<td>0.4</td>
<td>0.4</td>
<td>#</td>
</tr>
<tr>
<td>Other</td>
<td>0.3</td>
<td>0.3</td>
<td>#</td>
</tr>
</tbody>
</table>

# Rounds to zero.

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories. SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973.

Percentage of eighth-grade public and nonpublic school students identified as students with disabilities (SD) and/or English language learners (ELL) assessed in NAEP mathematics with accommodations, by SD/ELL category and type of accommodation: 2013

<table>
<thead>
<tr>
<th>Type of accommodation</th>
<th>SD and/or ELL</th>
<th>SD</th>
<th>ELL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual dictionary</td>
<td>0.7</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td>Braille presentation</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Braille response</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Breaks</td>
<td>2.5</td>
<td>2.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Calculator</td>
<td>2.3</td>
<td>2.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Cue to stay on task</td>
<td>0.7</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Directions read aloud in English</td>
<td>2.9</td>
<td>2.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Directions read aloud in Spanish</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Extended time</td>
<td>9.8</td>
<td>8.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Large-print booklet</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Magnification device</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>One-on-one</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Read aloud (all)</td>
<td>3.6</td>
<td>3.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Read aloud (occasional)</td>
<td>1.3</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Read aloud in Spanish</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>School staff administers</td>
<td>0.3</td>
<td>0.3</td>
<td>#</td>
</tr>
<tr>
<td>Scribe</td>
<td>0.2</td>
<td>0.2</td>
<td>#</td>
</tr>
<tr>
<td>Sign language presentation</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Sign language response</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td>Small group</td>
<td>8.4</td>
<td>7.6</td>
<td>1.5</td>
</tr>
<tr>
<td>Spanish-English booklet</td>
<td>0.2</td>
<td>#</td>
<td>0.2</td>
</tr>
<tr>
<td>Special equipment</td>
<td>0.3</td>
<td>0.3</td>
<td>#</td>
</tr>
<tr>
<td>Other</td>
<td>0.3</td>
<td>0.3</td>
<td>#</td>
</tr>
</tbody>
</table>

# Rounds to zero.

NOTE: Students identified as both SD and ELL were counted only once under the combined SD and/or ELL category, but were counted separately under the SD and ELL categories. SD includes students identified as having either an Individualized Education Program or protection under Section 504 of the Rehabilitation Act of 1973.

NAEP Accommodations Increase Inclusiveness

The responses of students with disabilities (SD) and English language learners (ELL) on NAEP assessments represent those of hundreds of other similar students. Without them, information about how to best meet the educational needs of these students would be lost. NAEP incorporates inclusive policies and practices into every aspect of the assessment, including selection of students, participation in the assessment administration, and valid and effective accommodations. This is essential to ensuring an assessment that yields meaningful NAEP results for all students. By representing their peers across the nation on NAEP, students with disabilities and English language learners help to ensure that NAEP results can be used to inform efforts to improve educational programs.

Just like any other student, SD and ELL students are selected to participate in NAEP. Within each selected school and grade to be assessed, students are chosen at random to participate in NAEP. Regardless of race/ethnicity, socioeconomic status, disability, status as an English language learner, or any other factors, every student has the same chance of being selected, because NAEP is administered to a sample of students who represent the student population of the nation as a whole, and for state level tests, of each individual state.

The accommodations allowed on NAEP and those allowed in states are often similar, but there may be some differences. Sometimes these differences result from the way that the subject being measured is defined in the NAEP frameworks. For example, NAEP does not allow read-aloud of any part of the NAEP reading test except the instructions, because decoding words is part of what the NAEP framework is measuring.

The many accommodations available in NAEP can be categorized for ease of understanding:

- Some are regarded as **Standard NAEP Practice**, available in almost all NAEP assessments for SD and ELL students.
- Other **accommodations for SD students** require special preparation, such as Braille or signing.
- Other **accommodations for ELL students**.
- Some accommodations are actually built-in features of the computer-based assessments that are available to all students and so are referred to as **Universal Design Elements**.

### Standard NAEP Practice, for SD and/or ELL

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended time</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Small group, or one-on-one</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>One-on-one</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Directions only read aloud in English</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Test items read aloud in English – occasional or most/all</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Breaks during test</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Writes directly in the booklet</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

### Other Accommodations for SD students

<table>
<thead>
<tr>
<th>Other Accommodations for SD Students</th>
<th>Math</th>
<th>Reading</th>
<th>Science</th>
<th>Writing (CBA)</th>
<th>Civics, Economics, Geography, U.S. History</th>
<th>Music and Visual Arts</th>
<th>TEL (pilot 2013)</th>
<th>Writing (before CBA; paper and pencil)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Other Accommodations for ELL students

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilingual dictionary with definitions in any language</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Directions only read aloud in Spanish</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (in Music, No in VisArt)</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Spanish/English version of the test (not g12)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No FN6</td>
<td>Yes FN3</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Test items read aloud in Spanish (not g12 Math)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Universal Design Elements and descriptions

<table>
<thead>
<tr>
<th>Universal Design Elements available for all students in Computer-Based Assessments</th>
<th>Used for CBA Writing 2011 and TEL pilot 2013</th>
</tr>
</thead>
</table>
| Adjusting font size | Text size options are provided for short standalone items (e.g., multiple choice and short constructed response items), but not for the scenario-based tasks which will be the standard 14-point font. For the short standalone items, students can change between three sizes ranging from approximately 14-point font to approximately 48-point font. Forty-eight point font is approximately ¾ - inch tall.  
NOTE: This only enlarges the short standalone items, NOT the scenario-based tasks, tool icons, menus, etc. |
| Small group | Unnecessary due to the mode of test administration. All students are interacting on a one-on-one basis with the computer and will have earbuds to reduce distractions. Read aloud and other accommodations will be provided through the computer |
| One-on-one | Unnecessary due to the mode of test administration. All students are interacting on a one-on-one basis with the computer and will have earbuds to reduce distractions. Read aloud and other accommodations will be provided |
'Read-Aloud' Assistance on Common Tests Proves Contentious

Accommodation for students with print-related disabilities has assessment consortia taking differing approaches

By Christina A. Samuels

Faced with the decision of whether to allow students with dyslexia and other print disabilities the option of having text passages on the common-core tests read aloud to them, the two federally financed consortia responsible for creating the general assessments took a Solomonic approach.

Rather than prohibit the so-called "read-aloud accommodation" entirely or allow reading aloud with no restriction, the Partnership for Assessment of Readiness for College and Careers decided to permit text passages to be read to students, with a notation on score reports saying no claims can be made regarding the student's foundational reading skills. The Smarter Balanced Assessment Consortium opted against the read-aloud accommodation for students in grades 3-5, saying it would invalidate the language constructs being measured; students taking the test in higher grades may use that accommodation.

Like many decisions that attempt to strike a balance between two opposing philosophies—those who wish to leave the use of read-aloud accommodations to a school-based team and those who want to prohibit its use entirely—the decisions of the testing consortia appear to have left no one entirely happy.

And now critics on either side say the consortia's decisions on reading aloud could be setting districts up for violations of special education law, or could ultimately leave a swath of students unable to read because the read-aloud accommodation was used as a crutch.

Fierce Defenders

Richard Allington, a professor of education at the University of Tennessee and one of the country's most recognized experts on early literacy, calls the accommodation "cheating."

"What special education does best is create illiterates," Mr. Allington said. "I know why they don't want their kids tested on reading activity. It's because they've done a terrible job of providing those kids with high-quality reading instruction."

But special educators believe just as strongly that for some children, a read-aloud accommodation is the tool they need to demonstrate what they know.
Lindsay Jones, the director of public policy and advocacy for the National Center for Learning Disabilities, said she was "stunned" that Smarter Balanced chose to ban read-aloud accommodations for elementary students, even for visually impaired children who may be in the early stages of learning Braille. (An official with the consortium explained that the testing group was told by its legal experts that it could not open up an accommodation to blind children without opening it up to all students with disabilities.)

The learning-disabilities center has argued on behalf of restricting read-aloud accommodations only on test items that gauge print decoding and fluency. Questions that measure other literacy skills, such as picking out a main idea of a text passage, should be open to the accommodation.

"There's been a lot of conjecture and anecdotes about evidence that states overuse read-aloud. I don't know why we would let that dominate our conversation going forward," Ms. Jones said. "Why can't we create items that will test comprehension, that will test decoding? That is a failure, to me, of the promise of these exams."

The restrictions also could represent a conflict with the Individuals with Disabilities Education Act, which gives broad powers to school-based teams of teachers, administrators, parents, and others who have deeper knowledge of a child's capabilities to determine what accommodations may be necessary. Perry A. Zirkel, a professor of education and law at Lehigh University in Bethlehem, Pa., said that the approach taken by Smarter Balanced could be more legally sound than drawing attention to an accommodation. The Educational Testing Service, which administers the SATs, was sued by a student for its policy of flagging tests that were taken under an extended-time accommodation. The testing company agreed to end flagging in 2001.

"My prediction is that [PARCC] thinks it's getting out of the problem," Mr. Zirkel said. "If this gets challenged, they will likely follow the ETS approach."

**Policy Differences**

Since 1990, the National Center for Educational Outcomes, based at the University of Minnesota -Twin Cities, has been tracking accommodations policies for students with disabilities. It found that while the read-aloud accommodation on test reading passages may be controversial, it is not rare. Several states currently allow text passages to be read aloud to students with certain disabilities, though in some cases, that accommodation results in the test being invalidated for accountability purposes.

Kentucky is an example of a state with an expansive read-aloud policy, "on the premise that the intent of reading is to measure comprehension." Hawaii's guidelines for the read-aloud accommodation are much more strict: It allows the use of text-to-speech on its online assessments, but only if the student is a nonreader who will never be able to read any words at any grade level throughout the student's education.

**Patchwork of Policies**

States currently vary on whether they allow text passages on state tests to be read aloud to students. Some prohibit this accommodation, while others allow it in certain circumstances, such as for students who are blind or visually impaired. Some states allow it with implications for scoring; for example, the test may be invalidated.
lifetime. The student must also receive all printed material for every subject in an audio format, at all times.

Somewhat more widely used were read-aloud accommodations for test instructions and for math tests, the center found. PARCC and Smarter Balanced have both decided to leave that particular use of read-aloud accommodations up to the school-based special education team working with a given student.

In allowing the use of read-aloud accommodations, even with some caveats or limitations, the common-core-testing consortia are striking a different path from the National Assessment of Educational Progress, which examines a nationally representative sample of students at grades 4, 8, and 12. NAEP allows a wide range of accommodations, but does not allow passages to be read aloud to students. That restriction has led to large student exclusion rates in some states.

The National Center for Educational Outcomes has suggested that the common-core assessments offer an opportunity for states to develop a coherent, and in its view necessary, multistate policy.

Such a policy should take into account that some students may be able to draw inferences from a text, or answer questions based on its main idea, even if they struggle with translating letters into sounds, also known as decoding. Requiring those students to decode to answer a question based on text comprehension is an "artificial barrier," according to a report researchers at the center wrote as guidance for the testing consortia.

"Ensuring that common standards have addressed accessibility concerns does not mean lowering the standards. It does mean, for example, providing a way for students who cannot hear to demonstrate their 'listening' skills; for students who cannot see to demonstrate their 'viewing' skills; and for students who cannot decode to demonstrate their comprehension skills in reading," the report says. But it’s not so easy to separate the tasks of reading comprehension and decoding, said literacy expert Timothy Shanahan, the chairman of the department of curriculum and instruction at the University of Illinois at Chicago.

"Part of the task of reading and learning to read is learning to get the words off the page while you think about them. Not having to get the words off the page gives a measurable advantage in most studies," he said.

In a blog post, Mr. Shanahan gave PARCC's read-aloud accommodation decision the "Lindsay Lohan Award for Bad Judgment," a cheeky reference to the troubled starlet. Smarter Balanced made a better move by limiting the read-aloud accommodation in the early grades, he said, but his preference would be not to use it at all.

http://www.edweek.org/ew/articles/2013/10/30/10cc-readaloud.h33.html?tkn=RZXFI5Ao0...
"Part of my concern is I have absolutely no doubt there are going to be school districts and states where people are going to be shopping to get the highest scores they can get," he said. And in states where read-aloud accommodations are used liberally, "those kids are going to do better if they're read to than if they have to read it themselves."

Diane Cordry Golden, the policy coordinator for the Missouri Council of Administrators of Special Education and the project coordinator for the Association of Assistive Technology Act Programs, has a foot in the special education and the assistive-technology worlds. She criticizes views such as Mr. Shanahan's, saying they are akin to "telling a child who can walk with great difficulty that using a wheelchair is not as good. You need to walk, period," she said. "We can't give you technology to work around it; we've got to fix this thing that we see as not normal in you."

It's a legitimate concern that a read-aloud accommodation could be overused, Ms. Golden added, but "this summative test is not an appropriate mechanism to use to address those issues. If you have kids who are really failing readers because of poor instruction, that needs to be driven by something else other than calling out kids with disabilities."

**Moving Forward**

The consortia are moving ahead with field-testing on a united front, though some state education leaders, such as in Colorado, a PARCC state, registered concerns about allowing the accommodation when the final vote was taken.

"Read-aloud is one of our major concerns. It's pretty fundamental to us," said Robert Hammond, Colorado's commissioner of education, at that meeting, held in June.

But both consortia say it's important to have some policy in place in time for the field-testing. If the accommodation is overused or badly deployed, the field tests will reveal those problems.

Mitchell D. Chester, the commissioner of education in Massachusetts, a state that does use a read-aloud accommodation on reading tests in limited circumstances, voted for the policy for the purpose of "getting it out into the bloodstream."

"I don't think there's a right or wrong answer to this," said Mr. Chester, the chairman of the PARCC governing board. "I think it's more a philosophical question."
PATCHWORK OF POLICIES

States currently vary on whether they allow text passages on state tests to be read aloud to students. Some prohibit this accommodation, while others allow it in certain circumstances, such as for students who are blind or visually impaired. Some states allow it with implications for scoring; for example, the test may be invalidated. Other variables are whether states allow human readers, software-based text-to-speech, prerecorded audio, or some combination of these.

SOURCES: National Center for Educational Outcomes; Education Week
Evaluation of Effects of Implementing
A NAEP Mathematics Calculator Booklet
Accommodation
Final Report

Prepared for: U.S. Department of Education
National Center for Education Statistics
Seventh and D Streets, SW
Washington, DC 20202

Prepared under: MOBIS Contract No. GS-10F-0087J
E-D08-DQ-0081
Task Order 0036

Authors: Monica A. Gribben
Tirso E. Diaz

Date: May 28, 2013
Evaluation of Effects of Implementing
a NAEP Mathematics Calculator Booklet Accommodation

Table of Contents

Background ................................................................................................................................. 1
  NAEP Inclusion Policy ........................................................................................................... 2
  NAEP Inclusion Strategies .................................................................................................. 3
  NAEP Calculator Use ........................................................................................................... 3
  Calculator Booklet Special Study ....................................................................................... 4
  NAEP Calculator Accommodation ..................................................................................... 5

Research Questions .................................................................................................................. 6

Method .................................................................................................................................... 6
  Gather and Review Data and Prior Results ........................................................................... 6
  Student Subgroups ............................................................................................................... 6
  Conduct Analyses ............................................................................................................... 9

Results ................................................................................................................................... 9
  Effect of Calculator Accommodation on Student Performance ........................................... 9
  Effect of Calculator Accommodation on State Reporting ............................................... 12
  Profiles of Students .......................................................................................................... 20
  Comparison of Omit Rates and Items Not Reached ......................................................... 28
  Comparison of Calculator Item Types .............................................................................. 30

Discussion ............................................................................................................................. 31
  Effects on Student Performance ....................................................................................... 31
  Effects on State Reporting ............................................................................................... 32
  Comparison of Students Requiring Calculator Accommodations ...................................... 33
  Conclusion ......................................................................................................................... 33

References ............................................................................................................................ 35

List of Tables

Table 1. Number of Students by Disability Status and Calculator Block Assignment .............. 7
Table 2. Mean Scale Scores for Student Subgroups on 2011 NAEP Mathematics (Grade 8) ...... 10
Table 3. Overall Percent Correct for Student Subgroups on 2011 NAEP Mathematics (Grade 8) ................................................................................................................................. 10
Table 4. Percent Correct by Block for Student Subgroups on 2011 NAEP Mathematics (Grade 8) ................................................................................................................................. 11
Table 5a. Mean scale score comparisons for Oklahoma grade 8 students on NAEP mathematics items (2009 and 2011) ........................................................................................................ 13
Table of Contents (Continued)

Table 5b. Mean scale score comparisons for Oklahoma grade 8 students on NAEP mathematics items (2009 and 2011) ................................................................. 13
Table 6a. Mean scale score comparisons for Maryland grade 8 students on NAEP mathematics items (2009 and 2011) ................................................................. 14
Table 6b. Mean scale score comparisons for Maryland grade 8 students on NAEP mathematics items (2009 and 2011) ................................................................. 14
Table 7a. Mean scale score comparisons for Texas grade 8 students on NAEP mathematics items (2009 and 2011) ................................................................. 16
Table 7b. Mean scale score comparisons for Texas grade 8 students on NAEP mathematics items (2009 and 2011) ................................................................. 16
Table 8a. Mean scale score comparisons for New Jersey grade 8 students on NAEP mathematics items (2009 and 2011) ................................................................. 17
Table 8b. Mean scale score comparisons for New Jersey grade 8 students on NAEP mathematics items (2009 and 2011) ................................................................. 17
Table 9. Demographic profiles of grade 8 students participating in NAEP mathematics items (2011) ............................................................................................................. 21
Table 10. Degree of disability of grade 8 students participating in NAEP mathematics (2011) ............................................................................................................. 22
Table 11. Type of disability of grade 8 students participating in NAEP mathematics (2011) ............................................................................................................. 26
Table 12. Types of accommodations used by grade 8 students participating in NAEP mathematics (2011) ......................................................................................... 27
Table 13. Number of accommodations used by grade 8 students participating in NAEP mathematics (2011) ......................................................................................... 28
Table 14. Percent of items omitted by grade 8 students participating in NAEP mathematics (2011) ......................................................................................... 28
Table 15. Percent of items not reached by grade 8 students participating in NAEP mathematics (2011) ......................................................................................... 29
Table 16. Percent correct by calculator activity of grade 8 students participating in NAEP mathematics (2011) ......................................................................................... 30

List of Figures

Figure 1. Student Groups in 2011 NAEP Mathematics Sample ......................................................... 8
Figure 2. Percent Correct on 2011 NAEP Mathematics (Grade 8) by Block for SD Requiring a Calculator Accommodation ................................................................. 11
Figure 3. Subgroup Mean Differences (2011-2009) for Oklahoma grade 8 students. ................. 18
Figure 4. Subgroup Mean Differences (2011-2009) for Maryland grade 8 students ................. 19
Figure 5. Subgroup Mean Differences (2011-2009) for Texas grade 8 students ................. 19
Table of Contents (Continued)

Figure 6. Subgroup Mean Differences (2011-2009) for New Jersey grade 8 students..............20
Figure 7. Demographic profiles by gender of grade 8 students participating in NAEP
    mathematics (2011)........................................................................................................22
Figure 8. Demographic profiles by race/ethnicity of grade 8 students participating in
    NAEP mathematics (2011)............................................................................................23
Figure 9. Demographic profiles by participants in the National School Lunch Program of
    grade 8 students participating in NAEP mathematics (2011).......................................23
Figure 10. Degree of disability of grade 8 students participating in NAEP mathematics (2011)....24
Figure 11. Grade level of performance of grade 8 students participating in NAEP
    mathematics (2011)........................................................................................................24
Figure 12. Type of disability of grade 8 students participating in NAEP mathematics (2011)........26
Figure 13. Types and number of accommodations used by grade 8 students participating
    in NAEP mathematics (2011).....................................................................................27
Figure 14. Percent of items omitted by grade 8 students participating in NAEP
    mathematics (2011).........................................................................................................29
Figure 15. Percent of items not reached by grade 8 students participating in NAEP
    mathematics (2011).......................................................................................................30
Figure 16. Percent correct by calculator activity of grade 8 students participating in
    NAEP mathematics (2011)............................................................................................31
Evaluation of Effects of Implementing
a NAEP Mathematics Calculator Booklet Accommodation

This study provides a detailed look at the effects of providing a calculator accommodation to students with disabilities (SD) who would otherwise not participate in the National Assessment of Educational Progress (NAEP). The purpose of this study is to determine how increased use of calculator booklets by SD would affect the quality of NAEP score reports. The analyses contained in this report provide information on the effects on reporting of offering calculator accommodation booklets, particularly at the demographic subgroup level. The analyses should be particularly relevant for 2013 reporting because the 2013 NAEP mathematics assessment included calculator accommodation booklets for the first time in NAEP history.

We used 2009 operational NAEP as a baseline and data from the 2011 NAEP inclusion study for the data analyses presented in this report. In 2011, approximately 900 grade 8 students participated in a NAEP special study providing a calculator accommodation (Educational Testing Service [ETS], 2012). Although calculators are used in NAEP, they are only allowed on designated calculator blocks which are assigned randomly to students. Students who would have been excluded from NAEP in 2011 because they used a calculator as an accommodation on their state mathematics assessment and NAEP did not allow a calculator accommodation were offered a calculator booklet to participate in the special study. The calculator booklets were operational booklets with two calculator blocks. In 2011, there were three grade 8 calculator blocks (e.g., A, B, C) as part of the operational assessment. These three blocks were used to create three pairs of blocks for three different calculator booklets (e.g., AB, AC, BC).

This report includes background history and information on the NAEP calculator accommodation issue. Description of the data analyses, results, and recommendations also are presented.

Background

The National Assessment of Educational Progress (NAEP) aims to represent all students. However, students with disabilities (SD) sometimes are excluded from participating because NAEP does not offer an accommodation students use on state assessments (e.g., read-aloud on reading, calculators on mathematics). Inclusion rates vary by jurisdiction, grade, subject area, and subgroup. In 2011, inclusion on NAEP was high with the overall inclusion rate for grade 8 in mathematics at 97 percent. The inclusion rate for SD was lower at 80 percent for grade 8 in mathematics (National Center for Education Statistics [NCES], 2011). The higher the participation rate of all students, the more accurate score reports will be. Bias introduced to NAEP statistics by exclusions, particularly differential exclusion rates, can affect conclusions related to the statistical significance of changes over time and differences among jurisdictions and subgroups. Put another way, excluding students from NAEP assessments because their accommodations are not allowed on NAEP is likely to increase overall scores because scores for SD are generally lower than scores for students without disabilities. For example, in 2011, the mean scale score for SD in grade 8 in mathematics was 250 compared to 288 for students without disabilities (NCES, 2011). Further, SD exclusion effects are becoming increasingly important as NAEP sample sizes increase and greater precision (i.e., smaller standard errors) is achieved by the program (Haertel, 2003). Despite sustained efforts by the U.S. Department of Education to increase the number of SD included in NAEP, high percentages of SD continue to...

1 Approximately 600 grade 4 students also participated in the inclusion study with calculator booklets.
be excluded in some jurisdictions, particularly Maryland, Oklahoma, and Texas where less than 60 percent of SD were included in grade 8 NAEP mathematics in 2011 (NCES, 2011).

**NAEP Inclusion Policy**

In March 2010, the National Assessment Governing Board (NAGB) revised the NAEP inclusion policy to improve the comparability of score reports of SD across states, districts, and other subgroups. One source of the variation in exclusion rates has been differences in rates of students defined as SD and differences in accommodation allowances across jurisdictions. To reduce variation in exclusion related to these causes, the policy statement (NAGB, 2010) included the following stipulations:

1. Only students with Individualized Education Programs (IEPs) or Section 504 plans can be considered for an accommodation.
2. If the student’s IEP or Section 504 plan specifies an accommodation permitted by NAEP, then the student takes NAEP with that accommodation.
3. If a student’s IEP or Section 504 plan does not address a need for an accommodation, school staff are encouraged to include the student without accommodations.
4. If the student requires an accommodation or modification that is not allowed by NAEP, the student is encouraged to take NAEP without it.
5. If a student takes an alternate assessment with modified achievement standards, that student should participate on NAEP with the NAEP allowable accommodations on his/her IEP.
6. If a student takes an alternate assessment with alternate achievement standards, that student is eligible for exclusion from NAEP. These SD are the only students eligible for exclusion.

The revised policy statement added guidance to the previous policy to encourage school staff to allow students with accommodations or modifications that are not permitted by NAEP to take NAEP without those accommodations (points 4 and 5). Further, the new policy statement explicitly noted that the only students eligible for exclusion are those who take an alternate assessment with alternate achievement standards (part of point 6). Other points of the revised policy were part of the prior policy statement on inclusion.

While most of the requirements are objective, participation by SD who require an accommodation not allowed by NAEP still depends on state and local policies, procedures, and interpretation. In states that allow a particular accommodation on state tests that is not allowed on NAEP, NAEP assessment staff have discussions with school personnel prior to the assessment. School personnel make the determination about participation on a student-by-student basis. Comparability of state results are threatened by differences in (a) state use of accommodations not allowed by NAEP and (b) participation decisions by state and local authorities regarding students who use accommodations not allowed by NAEP.

A new consideration for jurisdiction and school personnel when making their decisions is NAGB’s new inclusion goals: Ninety-five percent of all selected students will participate in NAEP (with or without accommodations), and 85 percent of SD and ELL students selected for assessment will participate in NAEP (with or without accommodations). States and jurisdictions that don’t meet the goals will be noted in results.
Inclusion rates are calculated as:

\[
\frac{\text{assessed} + \text{absent} + \text{refused}}{\text{assessed} + \text{absent} + \text{refused} + \text{excluded}}
\]

National, state, and district results with fewer than 85 percent inclusion of SD and/or English Language Learner (ELL) students participating in NAEP will be noted in upcoming reports releasing results (NAGB, 2010).

**NAEP Inclusion Strategies**

During the 2011 administration, the National Center for Education Statistics (NCES) implemented strategies to increase the inclusion rate. In one special study, assessment staff encouraged school personnel to include students who require access to a calculator when responding to mathematics items. Since this accommodation was not allowed on NAEP in 2011 but often allowed on state and other testing, it is a main contributor to differential participation across states and districts. Seventeen states at grade 8 that allow calculator accommodations on state assessments and in other testing situations took part in the study. Other states either did not offer calculator accommodations on their state assessments and/or did not permit calculator use on state assessments. Sampled students who had been excluded from NAEP for the lack of the calculator accommodation were provided with “calculator booklets” (including two calculator blocks). Their scores were not included in calculations for official score reports. The calculator booklet special study was conducted to gather data about how many more students would be included using the strategy and about the performance of this group of students who historically have not participated in NAEP mathematics assessments. NCES wanted to ensure that students requiring calculators on mathematics assessments could meaningfully participate in NAEP.

**NAEP Calculator Use**

In the regular distribution of items, or spiral, approximately one-third of NAEP’s grade 8 mathematics items allow students to use a calculator; the other two-thirds of items assess mathematics knowledge and skills without using a calculator (NAGB, 2007). Two blocks of cognitive items are included in each assessment booklet. Students can receive a booklet with (a) a block of items that allows a calculator and one that does not, (b) two blocks that do not allow a calculator, or (c) two blocks that allow a calculator. Calculator blocks and non-calculator blocks are considered of equal difficulty (Accommodations and Inclusion Coordinating Council [AICC], 2012). In 2011, there was a 52 percent chance that a student sampled in the eighth grade would receive a booklet with a calculator block (AICC, 2012).

Within the calculator blocks, items are categorized by the level of benefit a calculator is to the solution. Calculator active items do benefit from a calculator, calculator neutral items can be solved equally well with or without a calculator, and calculator inactive items are not benefited.

---

2 The 17 states participating in the special study at grade 8 included: Alaska, California, Indiana, Iowa, Louisiana, Maryland, Massachusetts, Montana, Missouri, New Hampshire, New Jersey, North Dakota, Oklahoma, South Carolina, Texas, Utah, and Vermont.

3 Some students who use calculator accommodations participate in NAEP without the accommodation. They are randomly assigned a booklet allowing calculator use on zero, one, or two blocks. They may use other accommodations as appropriate. These students were not part of the special study.
by a calculator (NAGB, 2007). Based on the mathematics framework (NAGB, 2010), most objectives permit use of a calculator; only some objectives related to computation limit the use of calculators. However, calculators are only permitted on items in calculator blocks.

**Calculator Booklet Special Study**

To align with the NAEP mathematics framework (NAGB, 2010), the students in the 2011 calculator booklet special study (892 in grade 8) were tested only on “calculator booklets” that included two calculator blocks (Educational Testing Service [ETS], 2012). The blocks/booklets in the calculator study were the same as those included in the operational assessment. The only difference is that students in the special study were assigned a booklet with two calculator blocks, similar to assigning an accommodation booklet (e.g., read-aloud, bilingual). Characteristics of SD who were assessed in the operational assessment and SD who were assessed in the calculator booklet special study were largely similar in the type of state assessment taken (e.g., with accommodations, modified, alternate) and degree of student disability (ETS, 2012).

NAEP’s Design, Analysis, and Reporting (DAR) contractor, ETS, conducted analyses using 2011 NAEP mathematics data that included students who took part in the calculator booklet special study. ETS found that an additional 0.4 percent of grade 8 students could be assessed using a calculator “accommodation” booklet when considering the entire national reporting sample (ETS, 2012). At the national level, exclusion rates would have decreased from 2.5 percent to 2.0 percent for grade 8. In several states use of the calculator booklet would have had a particularly large impact on exclusion rates. Three states at grade 8 would have made the goal of 95 percent inclusion only if the calculator accommodation was part of the operational assessment (Maryland, Oklahoma, and Texas). In Maryland, the exclusion rate would have dropped from 6.3 percent to 2.7 percent for grade 8 students. In Oklahoma, the exclusion rate would have declined from 9.8 percent to 3.3 percent for grade 8 students. In Texas, the exclusion rate would have decreased from 5.2 percent to 3.5 percent for grade 8 students. In other jurisdictions, the option of a calculator accommodation had little, if any, noticeable impact on exclusion rates (e.g., Missouri, New Hampshire, Vermont, and California). The option for a calculator accommodation had an even larger impact on exclusion rates on some Trial Urban Districts Assessment (TUDA) districts (e.g., Baltimore).

Predictably, the effects of a calculator accommodation option on inclusion are amplified when examining the SD subgroup (as compared to all students). The percentage of SD excluded would have decreased from 19.6 percent to 16.0 percent for grade 8 nationwide. At the state level, the percentage of SD excluded would have decreased from 60.1 percent to 19.3 percent for grade 8 in Oklahoma (NCES, embargoed data).

ETS-DAR examined the impact on performance by comparing differences in scale scores with and without students who participated in the calculator booklet special study. Across the national subgroups, scale scores were about 0.3 points greater at grade 8 when the students who participated in the special study were not included in the sample (ETS, 2012). For the SD subgroup, scale scores were about 0.6 points greater at grade 8 when those who participated in the calculator booklet special study were not included in the sample. For grade 8, scale scores without the special study students were about 2.1 points greater in Maryland, 3.5 points greater in Oklahoma, and 6.0 points greater in Baltimore (ETS, 2012). These data provide information on the relative magnitude of the differences in subgroup scale scores.4

---

4 Statistical significance of the scale score differences was not reported (ETS, 2012).
**NAEP Calculator Accommodation**

NCES, ETS, members of the November 2011 NAEP Quality Assurance Technical Panel (QATP), and others had concerns that aspects of the NAEP Mathematics Framework may not be equivalent in the NAEP calculator block item pool as compared to the entire mathematics operational item pool. There was concern for the validity of inferences based on scores where a large proportion of students in certain jurisdictions or subgroups do not complete parts of the framework. For example, items that evaluate computational skills are not included in the calculator booklets. In May 2012, ETS-DAR examined framework coverage of the calculator blocks as compared to all blocks. At the general content specification level, ETS researchers found that the calculator blocks were fairly representative and proportional to the framework of the corresponding grade. However, at the more specific objective level, the analysis showed that only 39 of 101 (39 percent) of grade 8 objectives were assessed in the calculator blocks (Freund, 2012). When considering all blocks, almost 80 percent of grade 8 objectives were assessed (80 of 101) (Freund, 2012).

Following a May 2012 meeting of the AICC, the general consensus of staff at NCES and the NAEP Contractor’s Alliance was that a booklet with two calculator blocks was not a viable accommodation option for the 2013 assessment (AICC, 2012). While NAGB’s goal of higher participation rates and other benefits (e.g., school testing coordinator satisfaction, less social stigma for excluded students) were achieved using the calculator accommodation strategy, lack of content coverage was a concern. In addition, there was the possibility that if states were allowed to request a calculator accommodation for students whose IEP allows them to use calculators on state assessments, then it is likely many more students would use a calculator accommodation than in the 2011 study. In some states (e.g., Virginia) students with state calculator accommodations were encouraged to participate in NAEP without the calculator as an accommodation. However, if NAEP allowed a calculator accommodation some students would be permitted to use the accommodation who previously would have participated without it. Another concern was the differential impact of a calculator accommodation on state results based on the range of SD exclusion rates at the state level.

Based on the results of the calculator booklet special study and increases in inclusion of SD, NCES approved the use of a calculator accommodation – assignment of an accommodation booklet with two calculator blocks – beginning in 2013. To make the calculator accommodation a more viable option, additional booklets were designated as calculator accommodation booklets. In the 2011 special study, there were only three booklets available at grade 8. The Alliance contractors developed additional calculator booklets to increase framework coverage among students requiring a calculator accommodation. To create additional accommodation booklets allowing use of a calculator, some current non-calculator operational blocks were selected for inclusion in the calculator accommodation booklets. SD assigned the calculator accommodation booklets were allowed to use a calculator on the blocks that are assessed operationally without calculators. Any items in the “new” calculator accommodation booklets assessing objectives that prohibit the use of a calculator will be dropped from analysis. In 2017 and beyond, changes to the Mathematics Framework and administration mode (i.e., computer-based assessment) may affect calculator use.
Research Questions

A number of questions could be asked and examined related to allowing a calculator accommodation on NAEP. Given the decision to offer a calculator accommodation in 2013 and available data from the 2011 inclusion study, we focus on questions that can be answered with currently available data and will inform reporting of future results where some students are permitted a calculator accommodation. We addressed the following research questions.

1. How will administration of calculator accommodation booklets to previously excluded students with disabilities affect student performance on NAEP mathematics items?
2. How does use of calculator accommodation booklets affect reporting of state and subgroup results, using 2009 as a baseline comparison to 2011 results?
3. How do profiles of students using a calculator accommodation compare with profiles of other SD and excluded students?
4. Are there differences in completion rates for calculator and non-calculator booklets?
5. Are there differences in student performance by calculator block item types (calculator active, inactive, and neutral) for students requiring a calculator accommodation?

Method

This study included three phases:

- Gather and Review Data and Prior Results
- Conduct Analyses
- Report Results and Recommendations

Gather and Review Data and Prior Results

HumRRO obtained NAEP data for the 2009 and 2011 operational assessments. The 2011 data include special variables that identify students excluded from the reporting sample but who participated in the calculator accommodation and inclusion studies, along with their plausible value scores. Data for public school students, where the majority of SD attend, were used. In addition, the authors gathered information about the calculator booklet study, including memoranda from ETS documenting the study and results. In addition, HumRRO received a file from ETS containing information tagging items in 2011 calculator blocks by type of calculator use – active, inactive, or neutral. It is expected that a calculator will be helpful in responding to calculator active items. Calculator inactive items do not benefit from the use of a calculator; these items do not require computation. Calculator neutral items require basic calculation, but it is not expected that a calculator will be helpful to most students in responding to these items.

Student Subgroups

To conduct analyses, NAEP participants were sorted into different groups:

- SD who would have been excluded because NAEP did not offer a calculator accommodation and subsequently participated in the calculator study (SD CalcStudy). These students received two calculator blocks.
• SD who use a calculator accommodation on their state assessment but participated in NAEP without a calculator accommodation (SD Calc). These students were randomly assigned either zero, one, or two calculator blocks.

• SD who do not use a calculator accommodation on their state assessment and participated in NAEP (SD NoCalc). These students were randomly assigned either zero, one, or two calculator blocks.

• Students without disabilities who participated in NAEP (NonSD). These students were randomly assigned either zero, one, or two calculator blocks.

Students in the calculator study (SD CalcStudy) were neither part of operational NAEP nor the reporting sample, but had NAEP scores available from the calculator booklets they used. Excluded students were sampled for NAEP but did not participate in the operational assessment or the special study. There are some demographic, disability, and accommodation data on excluded students, but no NAEP scores. Table 1 provides a list of all student groups by disability status and number of calculator blocks with sample sizes. Figure 1 illustrates the different groups of interest sampled to participate in NAEP.

Table 1. Number of Students by Disability Status and Calculator Block Assignment

<table>
<thead>
<tr>
<th>Student Subgroup</th>
<th>Uwt n</th>
<th>Wt n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students without disabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No calculator block</td>
<td>67,980</td>
<td>1,413,084</td>
</tr>
<tr>
<td>1 calculator block</td>
<td>70,647</td>
<td>1,471,271</td>
</tr>
<tr>
<td>2 calculator blocks</td>
<td>8,889</td>
<td>183,036</td>
</tr>
<tr>
<td>Students with disabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use calc. accomm. on state assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No calculator block</td>
<td>425</td>
<td>9,009</td>
</tr>
<tr>
<td>1 calculator block</td>
<td>339</td>
<td>6,864</td>
</tr>
<tr>
<td>2 calculator blocks</td>
<td>141</td>
<td>3,155</td>
</tr>
<tr>
<td>Do not use calc. accomm. on state assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No calculator block</td>
<td>8,090</td>
<td>163,730</td>
</tr>
<tr>
<td>1 calculator block</td>
<td>6,340</td>
<td>132,334</td>
</tr>
<tr>
<td>2 calculator blocks</td>
<td>1,552</td>
<td>32,961</td>
</tr>
<tr>
<td>Excluded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No calculator block</td>
<td>8,090</td>
<td>163,730</td>
</tr>
<tr>
<td>1 calculator block</td>
<td>6,340</td>
<td>132,334</td>
</tr>
<tr>
<td>2 calculator blocks</td>
<td>1,552</td>
<td>32,961</td>
</tr>
</tbody>
</table>

Excluded SD |

<table>
<thead>
<tr>
<th>Excluded SD</th>
<th>Uwt n</th>
<th>Wt n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3,739</td>
<td>67,263</td>
</tr>
</tbody>
</table>
Figure 1. Student Groups in 2011 NAEP Mathematics Sample
Conduct Analyses

We conducted analyses to provide empirical evidence related to each of the study questions using the existing 2009 and 2011 data. In addition, we looked at differences in performance on items by expected calculator activity. We conducted the following analyses:

1. Comparisons of student performance on 2011 NAEP mathematics items for each group of students, including CalcStudy, SD Calc, SD NoCalc, and NonSD.
2. Comparisons of overall and subgroup results from 2009 and 2011 for selected states to examine the effect of including previously excluded students by offering a calculator accommodation.
3. Comparisons of different groups of students on demographic, disability, and accommodation variables based on their participation in NAEP in 2011, including CalcStudy, SD Calc, SD NoCalc, Excluded, and NonSD.
4. Comparisons of the percentage of omitted items and number of items not reached for different groups of students on 2011 NAEP mathematics blocks.
5. Comparisons of percentage correct by expected calculator activity (calculator active, calculator inactive, and calculator neutral) for different types of students.

Results

Effect of Calculator Accommodation on Student Performance

To examine the effect on student performance of offering calculator accommodation booklets to a special subpopulation, we compared mean scale scores (Table 1) and percent correct (Tables 2 and 3) for SD CalcStudy, SD Calc, SD NoCalc, and NonSD.

As shown in Table 2, SD CalcStudy participants achieved a mean scale score of 237.8. SD Calc were assigned booklets containing no calculator blocks, one calculator block (either the first or second block), or two calculator blocks. Their mean scale scores ranged from 235.7 for those who received a calculator block first and a non-calculator block second to 240.9 for those who received two calculator blocks. Some of these subgroups are small and the standard errors in the mean scale scores are large, so we examined the range of mean scale scores for the lower and upper confidence intervals to make comparisons between groups. There is considerable overlap in the confidence intervals of SD CalcStudy and SD Calc. SD NoCalc scored 246.0, slightly higher than SD Calc. There was less overlap between confidence intervals for the SD CalcStudy and SD NoCalc. Students with no disabilities received much higher scores (286.1) than SD, echoing the achievement gap between SD and students without disabilities (Frieden, 2004).

A similar pattern emerged when examining percent correct for the same subgroups of students (see Table 3). SD CalcStudy participants received a p-value of 27.0, compared to a range of 24.5 to 28.4 for SD Calc. Figure 1 illustrates the considerable overlap of confidence intervals for all groups of SD requiring calculator accommodations. There was no overlap of confidence intervals for SD NoCalc (29.5-32.1) and those who participated in the calculator study (SD CalcStudy; 25.1-29.0). As seen in the mean scale score comparisons, students without disabilities (NonSD) achieved significantly greater p-values than those with disabilities. Results at the block level showed similar patterns (see Table 4).
### Table 2. Mean Scale Scores for Student Subgroups on 2011 NAEP Mathematics (Grade 8)

<table>
<thead>
<tr>
<th>Student Subgroup</th>
<th>Calculator Use</th>
<th>Uwt n</th>
<th>Mean</th>
<th>S.E.</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>Block 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculator study (SD CalcStudy)</td>
<td>Y</td>
<td>Y 764</td>
<td>237.8</td>
<td>2.6</td>
<td>[232.7, 242.8]</td>
</tr>
<tr>
<td>SD with calculator accommodation--operational NAEP (SD Calc)</td>
<td>N</td>
<td>N 425</td>
<td>239.2</td>
<td>2.8</td>
<td>[233.8, 244.6]</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Y 138</td>
<td>235.8</td>
<td>4.1</td>
<td>[227.9, 243.8]</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Y 201</td>
<td>235.7</td>
<td>3.6</td>
<td>[228.7, 242.7]</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Y 141</td>
<td>240.9</td>
<td>5.0</td>
<td>[231.2, 250.7]</td>
</tr>
<tr>
<td>SD without calculator accommodation--operational NAEP (SD NoCalc)</td>
<td>Y</td>
<td>Y 1,552</td>
<td>246.0</td>
<td>2.4</td>
<td>[241.3, 250.6]</td>
</tr>
<tr>
<td>Non-SD--operational NAEP (NonSD)</td>
<td>Y</td>
<td>Y 8,889</td>
<td>286.1</td>
<td>0.7</td>
<td>[284.8, 287.4]</td>
</tr>
</tbody>
</table>

### Table 3. Overall Percent Correct for Student Subgroups on 2011 NAEP Mathematics (Grade 8)

<table>
<thead>
<tr>
<th>Student Subgroup</th>
<th>Calculator Use</th>
<th>Uwt n</th>
<th>Percent Correct</th>
<th>S.E.</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>Block 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculator study (SD CalcStudy)</td>
<td>Y</td>
<td>Y 764</td>
<td>27.0</td>
<td>1.0</td>
<td>[25.1, 29.0]</td>
</tr>
<tr>
<td>SD with calculator accommodation--operational NAEP (SD Calc)</td>
<td>N</td>
<td>N 425</td>
<td>26.3</td>
<td>1.0</td>
<td>[24.3, 28.3]</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Y 138</td>
<td>24.5</td>
<td>1.3</td>
<td>[22.0, 27.0]</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>N 201</td>
<td>24.6</td>
<td>1.1</td>
<td>[22.4, 26.8]</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Y 141</td>
<td>28.4</td>
<td>1.5</td>
<td>[25.4, 31.4]</td>
</tr>
<tr>
<td>SD without calculator accommodation--operational NAEP (SD NoCalc)</td>
<td>Y</td>
<td>Y 1,552</td>
<td>30.8</td>
<td>0.7</td>
<td>[29.5, 32.1]</td>
</tr>
<tr>
<td>Non-SD--operational NAEP (NonSD)</td>
<td>Y</td>
<td>Y 8,889</td>
<td>48.4</td>
<td>0.4</td>
<td>[47.7, 49.1]</td>
</tr>
</tbody>
</table>
Table 4. Percent Correct by Block for Student Subgroups on 2011 NAEP Mathematics (Grade 8)

<table>
<thead>
<tr>
<th>Student Subgroup</th>
<th>Calculator Use</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Percent Correct</th>
<th>S.E.</th>
<th>95% C.I.</th>
<th>Percent Correct</th>
<th>S.E.</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculator study (SD CalcStudy)</td>
<td>Y</td>
<td>28.6</td>
<td>1.2</td>
<td>[26.2, 31.0]</td>
<td>25.3</td>
<td>1.0</td>
<td>[23.4, 27.2]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD with calculator accommodation--operational NAEP (SD Calc)</td>
<td>N</td>
<td>27.9</td>
<td>1.0</td>
<td>[25.9, 29.9]</td>
<td>24.8</td>
<td>1.2</td>
<td>[22.4, 27.1]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>25.1</td>
<td>1.5</td>
<td>[22.3, 28.0]</td>
<td>24.3</td>
<td>1.6</td>
<td>[21.1, 27.5]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>23.2</td>
<td>1.2</td>
<td>[20.8, 25.6]</td>
<td>25.9</td>
<td>1.5</td>
<td>[23.0, 28.9]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>27.9</td>
<td>1.7</td>
<td>[24.6, 31.1]</td>
<td>29.0</td>
<td>2.2</td>
<td>[24.7, 33.3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD without calculator accommodation--operational NAEP (SD NoCalc)</td>
<td>Y</td>
<td>30.2</td>
<td>0.6</td>
<td>[29.1, 31.4]</td>
<td>31.5</td>
<td>0.8</td>
<td>[29.9, 33.1]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SD--operational NAEP (NonSD)</td>
<td>Y</td>
<td>48.2</td>
<td>0.4</td>
<td>[47.5, 49.0]</td>
<td>48.6</td>
<td>0.4</td>
<td>[47.9, 49.3]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Percent Correct on 2011 NAEP Mathematics (Grade 8) by Block for SD Requiring a Calculator Accommodation.
**Effect of Calculator Accommodation on State Reporting**

**Exclusion Rates**

To examine the effect on reporting of including previously excluded students by offering a calculator accommodation on NAEP mathematics, we compared overall and subgroup results from 2009 and 2011. These analyses were conducted for the four states with the greatest number of students participating in the calculator special study (listed in order by initial exclusion rate, starting with the largest): Oklahoma, Maryland, Texas, and New Jersey. By providing a calculator accommodation on NAEP, these states would have been able to reduce their exclusion rates on grade 8 mathematics in 2011:

- Oklahoma – from 9.8 percent to 3.3 percent
- Maryland – from 6.3 percent to 2.7 percent
- Texas – from 5.2 percent to 3.5 percent
- New Jersey – from 4.2 percent to 1.4 percent

**State and Subgroup Performance**

The state-level results in Tables 5 through 8 show the potential for differences in state and state-level subgroup reporting when a large number of previously excluded SD are included in NAEP reports. In Oklahoma, as shown by the green shading in Table 5a, comparisons of 2009 and 2011 grade 8 results indicate significant gains for all students as well as females and Whites. When including students with calculator accommodations (SD CalcStudy) in 2011, the differences between 2009 and 2011 for all students, females, and Whites are not statistically significant (as shown by the yellow shading). Other subgroups, including males, Blacks, and Hispanics, show non-significant gains – essentially unchanged – when comparing 2009 to 2011 operational results. When SD CalcStudy participants are included, each of these subgroups (males, Blacks, and Hispanics) shows decreases in performance that are not statistically significant.

Including the calculator accommodation students in the 2011 results increased the number of students with disabilities, therefore we adjusted the 2009 results to include a similar proportion of SD who would have participated in NAEP if a calculator accommodation had been available in 2009.\(^5\) These adjustments are shown in Tables 5b through 8b. In Oklahoma, when comparing 2011 data with calculator accommodations students included to 2009 adjusted data, females showed significant gains, but there were no significant changes overall or in other subgroups.

In Maryland, as shown in Table 6a, there were no significant differences between 2009 and 2011 grade 8 results with or without students with calculator accommodations. When comparing 2011 data with calculator accommodations students included to 2009 adjusted data (Table 6b), there were still no significant changes overall or in any subgroup.

---

\(^5\) The 2009 mean scores were adjusted by assuming that the percentage of excluded SD that would have participated if a calculator accommodation was available in 2009 is equal to the percentage of excluded SD who participated in the calculator study in 2011. This was carried out by increasing the weights of SD included in the reporting sample in 2009 so that their total would reflect the number of additional SD from those previously excluded who would have participated with a calculator accommodation. The adjustment assumes that performance of included and excluded SD are comparable.
### Table 5a. Mean scale score comparisons for Oklahoma grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009</th>
<th>NAEP 2011 No Calculator Accommodation</th>
<th>NAEP 2011 With Calculator Accommodation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wt n</td>
<td>Mean</td>
<td>S.E.</td>
</tr>
<tr>
<td>ALL</td>
<td>40,910</td>
<td>275.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Male</td>
<td>20,771</td>
<td>277.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Female</td>
<td>20,139</td>
<td>273.8</td>
<td>1.2</td>
</tr>
<tr>
<td>White</td>
<td>23,875</td>
<td>282.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Black</td>
<td>3,890</td>
<td>260.6</td>
<td>2.9</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4,415</td>
<td>262.6</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.


### Table 5b. Mean scale score comparisons for Oklahoma grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009 (Adjusted)</th>
<th>NAEP 2011 With Calculator Accommodation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wt n</td>
<td>Mean</td>
</tr>
<tr>
<td>ALL</td>
<td>42,725</td>
<td>274.2</td>
</tr>
<tr>
<td>Male</td>
<td>21,909</td>
<td>275.9</td>
</tr>
<tr>
<td>Female</td>
<td>20,817</td>
<td>272.4</td>
</tr>
<tr>
<td>White</td>
<td>24,882</td>
<td>280.6</td>
</tr>
<tr>
<td>Black</td>
<td>4,107</td>
<td>259.0</td>
</tr>
<tr>
<td>Hispanic</td>
<td>4,567</td>
<td>261.4</td>
</tr>
</tbody>
</table>
### Table 6a. Mean scale score comparisons for Maryland grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009</th>
<th>NAEP 2011 No Calculator Accommodation</th>
<th>NAEP 2011 With Calculator Accommodation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wt n</td>
<td>Mean</td>
<td>S.E.</td>
</tr>
<tr>
<td>ALL</td>
<td>53,951</td>
<td>288.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Male</td>
<td>27,043</td>
<td>289.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Female</td>
<td>26,907</td>
<td>286.9</td>
<td>1.3</td>
</tr>
<tr>
<td>White</td>
<td>26,267</td>
<td>302.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Black</td>
<td>18,847</td>
<td>265.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5,155</td>
<td>274.9</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.


### Table 6b. Mean scale score comparisons for Maryland grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009 (Adjusted)</th>
<th>NAEP 2011 With Calculator Accommodation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wt n</td>
<td>Mean</td>
</tr>
<tr>
<td>ALL</td>
<td>56,192</td>
<td>287.4</td>
</tr>
<tr>
<td>Male</td>
<td>28,587</td>
<td>288.5</td>
</tr>
<tr>
<td>Female</td>
<td>27,605</td>
<td>286.3</td>
</tr>
<tr>
<td>White</td>
<td>27,468</td>
<td>302.0</td>
</tr>
<tr>
<td>Black</td>
<td>19,698</td>
<td>264.9</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5,324</td>
<td>273.9</td>
</tr>
</tbody>
</table>

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.

In Texas, as shown in Table 7a, comparisons of 2009 and 2011 grade 8 results indicated significant gains for all students as well as females and Hispanics. When including students with calculator accommodations in 2011, the difference between 2009 and 2011 for all students was no longer statistically significant. However, females and Hispanics still showed significant gains. When comparing 2011 data with calculator accommodations students included to 2009 adjusted data (Table 7b), there were significant gains overall and for females and Hispanics.

In New Jersey, as shown in Table 8a, there were no significant changes between 2009 and 2011 grade 8 results for all students as well as all subgroups. All students and subgroups, except males, showed non-significant gains – essentially unchanged – when comparing 2009 to 2011 operational results. There were non-significant declines for males. When students who participated in the calculator accommodation study were included the results were essentially the same. When comparing 2011 data with calculator accommodations students included to 2009 adjusted data (Table 8b), there were no significant changes overall or in any subgroup.
### Table 7a. Mean scale score comparisons for Texas grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009 (Wt n, Mean, S.E.)</th>
<th>NAEP 2011 No Calculator Accommodation (Wt n, Mean, S.E., Difference*, t, p-value)</th>
<th>NAEP 2011 With Calculator Accommodation (Wt n, Mean, S.E., Difference*, t, p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>306,243 (286.7, 1.3)</td>
<td>323,200 (290.5, 0.9) 3.8 2.5 0.01</td>
<td>328,946 (289.6, 0.9) 2.9 1.9 0.05</td>
</tr>
<tr>
<td>Male</td>
<td>151,897 (287.4, 1.4)</td>
<td>162,059 (290.9, 1.2) 3.6 1.9 0.06</td>
<td>166,075 (289.8, 1.2) 2.4 1.3 0.20</td>
</tr>
<tr>
<td>Female</td>
<td>154,345 (286.0, 1.4)</td>
<td>161,141 (290.0, 0.9) 4.0 2.3 0.02</td>
<td>162,870 (289.5, 0.9) 3.5 2.0 0.04</td>
</tr>
<tr>
<td>White</td>
<td>112,194 (300.8, 1.6)</td>
<td>102,792 (303.7, 1.2) 2.9 1.5 0.15</td>
<td>104,305 (302.9, 1.2) 2.1 1.0 0.30</td>
</tr>
<tr>
<td>Black</td>
<td>41,356 (272.3, 1.8)</td>
<td>40,946 (277.0, 1.8) 4.7 1.8 0.07</td>
<td>42,702 (275.3, 1.9) 3.0 1.2 0.25</td>
</tr>
<tr>
<td>Hispanic</td>
<td>139,625 (277.2, 1.5)</td>
<td>163,463 (283.4, 1.0) 6.2 3.5 0.00</td>
<td>165,937 (282.8, 1.0) 5.6 3.1 0.00</td>
</tr>
</tbody>
</table>


Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.

### Table 7b. Mean scale score comparisons for Texas grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009 (Adjusted) (Wt n, Mean, S.E.)</th>
<th>NAEP 2011 With Calculator Accommodation (Wt n, Mean, S.E., Difference*, t, p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>311,066 (286.2, 1.3)</td>
<td>328,946 (289.6, 0.9) 3.5 2.2 0.03</td>
</tr>
<tr>
<td>Male</td>
<td>154,854 (286.9, 1.5)</td>
<td>166,075 (289.8, 1.2) 2.9 1.5 0.12</td>
</tr>
<tr>
<td>Female</td>
<td>156,212 (285.5, 1.5)</td>
<td>162,870 (289.5, 0.9) 4.0 2.3 0.02</td>
</tr>
<tr>
<td>White</td>
<td>113,829 (300.3, 1.6)</td>
<td>104,305 (302.9, 1.2) 2.6 1.3 0.21</td>
</tr>
<tr>
<td>Black</td>
<td>42,277 (271.7, 1.8)</td>
<td>42,702 (275.3, 1.9) 3.6 1.4 0.18</td>
</tr>
<tr>
<td>Hispanic</td>
<td>141,777 (276.8, 1.5)</td>
<td>165,937 (282.8, 1.0) 6.0 3.3 0.00</td>
</tr>
</tbody>
</table>

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.

### Table 8a. Mean scale score comparisons for New Jersey grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009</th>
<th>NAEP 2011 No Calculator Accommodation</th>
<th>NAEP 2011 With Calculator Accommodation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wt n</td>
<td>Mean</td>
<td>S.E.</td>
</tr>
<tr>
<td>ALL</td>
<td>97,691</td>
<td>292.7</td>
<td>1.4</td>
</tr>
<tr>
<td>Male</td>
<td>49,178</td>
<td>294.8</td>
<td>1.7</td>
</tr>
<tr>
<td>Female</td>
<td>48,513</td>
<td>290.5</td>
<td>1.5</td>
</tr>
<tr>
<td>White</td>
<td>57,578</td>
<td>301.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Black</td>
<td>15,572</td>
<td>267.2</td>
<td>2.6</td>
</tr>
<tr>
<td>Hispanic</td>
<td>16,658</td>
<td>271.8</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.  

### Table 8b. Mean scale score comparisons for New Jersey grade 8 students on NAEP mathematics items (2009 and 2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>NAEP 2009 (Adjusted)</th>
<th>NAEP 2011 With Calculator Accommodation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wt n</td>
<td>Mean</td>
</tr>
<tr>
<td>ALL</td>
<td>98,853</td>
<td>292.3</td>
</tr>
<tr>
<td>Male</td>
<td>49,900</td>
<td>294.4</td>
</tr>
<tr>
<td>Female</td>
<td>48,953</td>
<td>290.1</td>
</tr>
<tr>
<td>White</td>
<td>58,280</td>
<td>301.2</td>
</tr>
<tr>
<td>Black</td>
<td>15,811</td>
<td>266.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>16,845</td>
<td>271.4</td>
</tr>
</tbody>
</table>

Note: Green shading indicates a statistically significant difference. Yellow shading indicates a comparison that is not statistically significant. Yellow shading shows changes from significant gains to non-significant gains.  
Figures 3 through 6 graphically summarize the differences between 2009 and 2011 mean scale scores. For each subgroup, three error bars are shown representing the 95 percent confidence interval of differences in mean scale scores based on three calculations: (a) using reporting sample results for both 2009 and 2011, (b) including students with a calculator accommodation in 2011 without adjusting 2009 results, and (c) including students with a calculator accommodation in 2011 while at the same time adjusting 2009 results. The lower and upper endpoints of the error bars represent the bounds of the confidence interval, with the midpoint representing the mean scale score difference. Error bars entirely above or below the zero horizontal reference line indicate differences between 2009 and 2011 results that were statistically significant.

As shown in Figure 3, the error bars representing operational scale score differences for Oklahoma were above the zero reference line for all students, females, and Whites, indicating statistically significant gains. All three error bars then shifted downward, enclosing the reference line, indicating not statistically significant gains when including students with calculator accommodations. After adjusting the 2009 results, all three error bars shifted upward but only the error bar representing females was above the reference line. For all other groups, the error bars followed a similar pattern of shifting downward when including students with a calculator accommodation, and then shifting upward after adjusting the 2009 results, but all indicated not statistically significant gains.

![Figure 3. Subgroup Mean Differences (2011-2009) for Oklahoma grade 8 students.](image)

Figure 3 graphically summarizes the differences between 2009 and 2011 mean scale scores for Maryland. Again, all error bars followed the general pattern of shifting downward when including students with calculator accommodations, and then shifting upward after adjusting the 2009 results across all subgroups. However, all error bars indicated not statistically significant loses or gains. Less dramatic changes were observed in Maryland compared to those observed in Figure 3 for Oklahoma.
Figure 4. Subgroup Mean Differences (2011-2009) for Maryland grade 8 students.

Figure 5 graphically summarizes the differences between 2009 and 2011 mean scale scores described above for Texas. Changes in the location of error bars relative to the reference line were very minimal across all subgroups. Note that error bars for Hispanic students were narrower compared to corresponding error bars in Figures 2 and 3, as expected, since Texas has a relatively higher proportion of Hispanic students than Oklahoma and Maryland.

Figure 5. Subgroup Mean Differences (2011-2009) for Texas grade 8 students.
Figure 6 graphically summarizes the differences between 2009 and 2011 mean scale scores for New Jersey. All error bars indicated not statistically significant gains or losses.

![Figure 6](image)

**Figure 6. Subgroup Mean Differences (2011-2009) for New Jersey grade 8 students.**

**Profiles of Students**

To put differences in student performance and the effect of a calculator accommodation on reporting in context, we examined characteristics of students to develop descriptive profiles using demographic variables – gender, race, ethnicity, and participation in the National School Lunch Program (NSLP). In addition, distributions of disability and accommodations were examined. To describe students with disabilities, we used five measures:

- Degree of disability – mild, moderate, and profound.
- Grade level of performance – performing at or above grade level, one year below grade level, and two or more years below grade level.
- Type of disability – five most frequent disabilities: specific learning disability, speech impediment, emotional disturbance, autism, and mental retardation.
- Type of accommodation – most frequently used accommodations: extended time, small group, read all in English, read directions in English, read occasionally in English, breaks, cueing, school staff administer test, one on one, special equipment/preferential seating, and scribe.
- Number of accommodations – three or more.

**Demographic Characteristics**

As shown in Table 9, students participating in the special study were more likely to be male and Black than non-SD. They also were more likely to participate in the NSLP than SD and non-SD included in operational NAEP.
### Table 9. Demographic profiles of grade 8 students participating in NAEP mathematics items (2011)

<table>
<thead>
<tr>
<th>Student Group</th>
<th>Calculator Study (SD CalcStudy)</th>
<th>Included SD (SD Calc and SD NoCalc)</th>
<th>Excluded SD</th>
<th>Included Non-SD (NonSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>S.E.</td>
<td>Percent</td>
<td>S.E.</td>
</tr>
<tr>
<td>ALL</td>
<td>100.0</td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Male</td>
<td>65.4</td>
<td>2.7</td>
<td>64.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Female</td>
<td>34.6</td>
<td>2.7</td>
<td>35.2</td>
<td>0.6</td>
</tr>
<tr>
<td>White</td>
<td>43.3</td>
<td>2.8</td>
<td>53.0</td>
<td>0.6</td>
</tr>
<tr>
<td>Black</td>
<td>25.8</td>
<td>2.9</td>
<td>18.8</td>
<td>0.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>23.3</td>
<td>3.5</td>
<td>22.3</td>
<td>0.6</td>
</tr>
<tr>
<td>NSLP</td>
<td>64.6</td>
<td>3.0</td>
<td>58.0</td>
<td>0.6</td>
</tr>
</tbody>
</table>

*Difference with SD Calculator Accommodation
Disability Characteristics

Data in Table 10 indicate most students with disabilities have a mild or moderate disability. Of those participating in operational NAEP in 2011, half had a mild disability compared to 46 percent of students participating in the calculator study. More students in the calculator study were considered working at one year below grade level (30 percent) than SD in NAEP (26 percent). The gap was even greater for students performing two or more years below grade level. Nearly 37 percent of SD receiving calculator accommodations were two or more years below grade level compared to approximately one-third of SD participating in operational NAEP without a calculator accommodation.

Table 10. Degree of disability of grade 8 students participating in NAEP mathematics (2011)

<table>
<thead>
<tr>
<th>Degree of Disability</th>
<th>Calculator Study (SD CalcStudy)</th>
<th>Included SD (SD Calc and SD NoCalc)</th>
<th>Excluded SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>SE</td>
<td>Percent</td>
</tr>
<tr>
<td>Profound</td>
<td>5.7</td>
<td>1.2</td>
<td>6.1</td>
</tr>
<tr>
<td>Moderate</td>
<td>35.2</td>
<td>3.5</td>
<td>34.7</td>
</tr>
<tr>
<td>Mild</td>
<td>45.7</td>
<td>2.7</td>
<td>50.0</td>
</tr>
</tbody>
</table>

Grade Level Student Performs on

<table>
<thead>
<tr>
<th></th>
<th>Included SD (SD Calc and SD NoCalc)</th>
<th>Excluded SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>At or above</td>
<td>18.5</td>
<td>3.5</td>
</tr>
<tr>
<td>1 year below</td>
<td>29.6</td>
<td>3.3</td>
</tr>
<tr>
<td>2 or more years below</td>
<td>36.9</td>
<td>3.6</td>
</tr>
</tbody>
</table>

*Difference with SD Calculator Accommodation

Figure 7 shows that SD are more likely to be male than female. As shown in Figure 8, students in the calculator study were slightly more likely to be a minority. Students in the calculator study were slightly more likely to participate in the NSLP than SD in NAEP and non-SD (Figure 9). There was little difference between excluded SD and calculator study students on participation in the NSLP.

![Figure 7. Demographic profiles by gender of grade 8 students participating in NAEP mathematics (2011)](image-url)
Figure 8. Demographic profiles by race/ethnicity of grade 8 students participating in NAEP mathematics (2011)

Figure 9. Demographic profiles by participants in the National School Lunch Program of grade 8 students participating in NAEP mathematics (2011)
Figure 10 visually shows students excluded from NAEP were much more likely to have profound disabilities and much less likely to have mild disabilities than SD who participated in NAEP, either operationally or as part of the calculator study. When looking at grade level (Figure 11), calculator study students were less likely to be at or above grade level and slightly more likely to be 2 or more years below grade level than SD in operational NAEP. SD excluded from NAEP were much more likely to be 2 or more years below grade level than other SD.

Figure 10. Degree of disability of grade 8 students participating in NAEP mathematics (2011)

Figure 11. Grade level of performance of grade 8 students participating in NAEP mathematics (2011)
In Table 11 and Figure 12 students participating in the calculator study with a calculator accommodation were similar to SD participating in operational NAEP in terms of the types of disabilities they have. Excluded students were significantly less likely to have a specific learning disability and more likely to have a specific disability such as mental retardation, autism, visual impairment, developmental delay, or brain injury than students in the calculator study.

As shown in Table 12 calculator study students were more likely to need read all or occasionally in English, preferential seating, and a scribe compared to SD participating in NAEP operationally.

Figure 13 illustrates the differences between SD in the calculator study and operational NAEP. There was overlap on many of the accommodations, showing little difference between the two groups except for “read aloud occasionally” and “read aloud all.”

Data in Table 13 show the percentage of students receiving one or more accommodations, two or more accommodations, and so forth. Students with disabilities participating in operational NAEP used significantly fewer accommodations than students in the calculator study. Note, all students in the calculator study received a calculator accommodation and three-quarters received at least two other accommodations, often small group and extended time. More than one in five students in the calculator study received five or more accommodations compared to less than 4 in 100 SD in operational NAEP.
Table 11. Type of disability of grade 8 students participating in NAEP mathematics (2011)

<table>
<thead>
<tr>
<th>Type of Disability</th>
<th>Calculator Study (SD CalcStudy)</th>
<th>Included SD (SD Calc and SD NoCalc)</th>
<th>Excluded SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent  SE</td>
<td>Percent  SE Difference*  t  p -value</td>
<td>Percent  SE Difference*  t  p -value</td>
</tr>
<tr>
<td>Any disability</td>
<td>97.4  1.5</td>
<td>100.0  0.0  2.6  1.7  0.09</td>
<td>100.0  0.0  2.6  1.7  0.09</td>
</tr>
<tr>
<td>Specific learning disability</td>
<td>64.1  2.9</td>
<td>58.2  0.7 -5.8 -1.9  0.05</td>
<td>40.3  1.4 -23.7 -7.2  0.00</td>
</tr>
<tr>
<td>Other health disability</td>
<td>15.6  2.3</td>
<td>15.1  0.5 -0.5 -0.2  0.83</td>
<td>12.3  0.9 -3.3 -1.3  0.18</td>
</tr>
<tr>
<td>Speech impairment</td>
<td>6.7  1.4</td>
<td>9.1  0.4  2.4  1.6  0.10</td>
<td>10.5  0.8  3.8  2.3  0.02</td>
</tr>
<tr>
<td>Emotional disturbance</td>
<td>5.1  1.2</td>
<td>5.2  0.3  0.1  0.1  0.94</td>
<td>6.4  0.6  1.3  1.0  0.34</td>
</tr>
<tr>
<td>Autism</td>
<td>2.9  1.2</td>
<td>3.5  0.2  0.7  0.5  0.60</td>
<td>8.5  0.7  5.6  4.0  0.00</td>
</tr>
<tr>
<td>Mental retardation</td>
<td>2.4  0.8</td>
<td>2.5  0.2  0.1  0.2  0.85</td>
<td>24.1  1.1  21.7 15.9  0.00</td>
</tr>
</tbody>
</table>

*Difference with SD calculator accommodation

Figure 12. Type of disability of grade 8 students participating in NAEP mathematics (2011)
Table 12. Types of accommodations used by grade 8 students participating in NAEP mathematics (2011)

<table>
<thead>
<tr>
<th>Type of Accommodation</th>
<th>Calculator Study (SD CalcStudy)</th>
<th>Included SD (SD Calc and SD NoCalc)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>SE</td>
<td>Percent</td>
<td>SE</td>
<td>Difference*</td>
<td>t</td>
<td>p-value</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>All accommodations</td>
<td>100.0</td>
<td>0.0</td>
<td>83.6</td>
<td>0.5</td>
<td>-16.4</td>
<td>-31.3</td>
</tr>
<tr>
<td>Extended time</td>
<td>59.7</td>
<td>4.7</td>
<td>68.3</td>
<td>0.8</td>
<td>8.6</td>
<td>1.8</td>
</tr>
<tr>
<td>Small group</td>
<td>66.1</td>
<td>3.2</td>
<td>64.5</td>
<td>0.8</td>
<td>-1.6</td>
<td>-0.5</td>
</tr>
<tr>
<td>Read all in English</td>
<td>44.6</td>
<td>4.2</td>
<td>26.5</td>
<td>0.6</td>
<td>-18.1</td>
<td>-4.3</td>
</tr>
<tr>
<td>Read directions in English</td>
<td>15.0</td>
<td>2.3</td>
<td>19.7</td>
<td>0.7</td>
<td>4.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Breaks during test</td>
<td>16.8</td>
<td>2.1</td>
<td>12.5</td>
<td>0.4</td>
<td>-4.3</td>
<td>-2.0</td>
</tr>
<tr>
<td>Read occasionally in English</td>
<td>16.4</td>
<td>3.1</td>
<td>6.6</td>
<td>0.4</td>
<td>-9.8</td>
<td>-3.2</td>
</tr>
<tr>
<td>Cueing</td>
<td>2.4</td>
<td>0.7</td>
<td>4.1</td>
<td>0.3</td>
<td>1.7</td>
<td>2.3</td>
</tr>
<tr>
<td>School staff administer test</td>
<td>3.0</td>
<td>0.9</td>
<td>2.8</td>
<td>0.2</td>
<td>-0.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>One on one</td>
<td>5.6</td>
<td>1.8</td>
<td>2.6</td>
<td>0.2</td>
<td>-3.0</td>
<td>-1.7</td>
</tr>
<tr>
<td>Preferential seating</td>
<td>7.0</td>
<td>1.8</td>
<td>2.5</td>
<td>0.2</td>
<td>-4.4</td>
<td>-2.4</td>
</tr>
<tr>
<td>Scribe</td>
<td>4.0</td>
<td>1.0</td>
<td>1.7</td>
<td>0.2</td>
<td>-2.3</td>
<td>-2.3</td>
</tr>
</tbody>
</table>

*Difference with SD calculator accommodation

Figure 13. Types and number of accommodations used by grade 8 students participating in NAEP mathematics (2011)
Table 13. Number of accommodations used by grade 8 students participating in NAEP mathematics (2011)

<table>
<thead>
<tr>
<th>Number of Accommodation</th>
<th>Calculator Study (SD CalcStudy)</th>
<th>Included SD (SD Calc and SD NoCalc)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>SE</td>
</tr>
<tr>
<td>Accommodations &gt;= 1</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Accommodations &gt;= 2</td>
<td>92.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Accommodations &gt;= 3</td>
<td>75.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Accommodations &gt;= 4</td>
<td>49.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Accommodations &gt;= 5</td>
<td>21.2</td>
<td>2.5</td>
</tr>
</tbody>
</table>

*Difference with SD calculator accommodation

Comparison of Omit Rates and Items Not Reached

We also examined the percent of items omitted and not reached. Table 14 and Figure 14 present omit rates for students in the calculator study and operational NAEP. Few items were omitted. Students in the calculator study were similar to SD (SD Calc and SD NoCalc) as well as non-SD in operational NAEP in terms of omit rate.

Table 14. Percent of items omitted by grade 8 students participating in NAEP mathematics (2011)

<table>
<thead>
<tr>
<th>Student Subgroup</th>
<th>Calculator Use</th>
<th>Booklet</th>
<th>Block One</th>
<th>Block Two</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Block 1</td>
<td>Block 2</td>
<td>Percent Omitted</td>
<td>S.E</td>
</tr>
<tr>
<td>Calculator study (SD CalcStudy)</td>
<td>Y</td>
<td>Y</td>
<td>1.8</td>
<td>0.2</td>
</tr>
<tr>
<td>SD with calculator accommodation--operational NAEP (SD Calc)</td>
<td>N</td>
<td>N</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>3.6</td>
<td>0.9</td>
<td>[1.8, 5.3]</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>2.5</td>
<td>0.7</td>
<td>[1.1, 3.9]</td>
</tr>
<tr>
<td>y</td>
<td>y</td>
<td>1.7</td>
<td>0.4</td>
<td>[0.9, 2.5]</td>
</tr>
<tr>
<td>SD without calculator accommodation--operational NAEP (SD NoCalc)</td>
<td>Y</td>
<td>Y</td>
<td>1.6</td>
<td>0.2</td>
</tr>
<tr>
<td>Non-SD--operational NAEP (NonSD)</td>
<td>Y</td>
<td>Y</td>
<td>1.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Figure 14. Percent of items omitted by grade 8 students participating in NAEP mathematics (2011)

Percent of items not reached is shown in Table 15 and Figure 15. Students receiving two calculator blocks in operational NAEP (SD and non-SD), completed slightly more items than SD in the calculator study who also received two calculator blocks.

Table 15. Percent of items not reached by grade 8 students participating in NAEP mathematics (2011)

<table>
<thead>
<tr>
<th>Student Subgroup</th>
<th>Calculator Use</th>
<th>Booklet</th>
<th>Block One</th>
<th>Block Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculator study</td>
<td>Block 1</td>
<td>Block 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SD CalcStudy)</td>
<td>Y</td>
<td>Y</td>
<td>1.4</td>
<td>0.6</td>
</tr>
<tr>
<td>SD with calculator accommodation--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>operational NAEP (SD Calc)</td>
<td>N</td>
<td>N</td>
<td>2.0</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Y</td>
<td>3.4</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>N</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>y</td>
<td>y</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>SD without calculator accommodation--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>operational NAEP (SD NoCalc)</td>
<td>Y</td>
<td>Y</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-SD--operational NAEP (NonSD)</td>
<td>Y</td>
<td>Y</td>
<td>1.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Comparison of Calculator Item Types

Items in calculator blocks may or may not require a calculator for ease in responding. Items containing larger numbers are typically easier with a calculator and are designated “calculator active.” Items for which a calculator would not be helpful, for example, identifying a geometric shape, are considered “calculator inactive.” “Calculator neutral” items are those where a calculator could be used but usually wouldn’t be needed, such as simple computations.

Calculator active items were more difficult than calculator neutral or inactive items for all students as shown in Table 16 and Figure 16. In general, there was no difference between calculator inactive and neutral items.

Table 16. Percent correct by calculator activity of grade 8 students participating in NAEP mathematics (2011)

<table>
<thead>
<tr>
<th>Student Subgroup</th>
<th>Calculator Use</th>
<th>Calculator Active</th>
<th>Calculator Inactive</th>
<th>Calculator Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Block 1</td>
<td>Block 2</td>
<td>Percent Correct</td>
<td>S.E.</td>
</tr>
<tr>
<td>Calculator study (SD CalcStudy)</td>
<td>Y</td>
<td>Y</td>
<td>16.6</td>
<td>1.1</td>
</tr>
<tr>
<td>SD with calculator accommodation--operational NAEP (SD Calc)</td>
<td>1 or 2 Calc. Blocks</td>
<td>17.1</td>
<td>1.2</td>
<td>[14.7 , 19.6]</td>
</tr>
<tr>
<td>SD without calculator accommodation--operational NAEP (SD NoCalc)</td>
<td>Y</td>
<td>Y</td>
<td>20.6</td>
<td>1.2</td>
</tr>
<tr>
<td>Non-SD--operational NAEP (NonSD)</td>
<td>Y</td>
<td>Y</td>
<td>43.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Figure 16. Percent correct by calculator activity of grade 8 students participating in NAEP mathematics (2011)

Discussion

Offering calculator accommodations for NAEP mathematics assessments increases inclusion, especially in states where a calculator accommodation is offered on the state assessment and thus is often included on students’ IEPs. The key questions are: (a) how does the increased inclusion affect student performance? (research question 1) and (b) is there an impact on reporting of results, particularly at the state level? (research question 2) In addressing these questions, it is useful to look at the students requiring calculator accommodations and to identify similarities and differences in students with and without such accommodations (research questions 3-5). In 2011, a NAEP calculator booklet special study was conducted offering a calculator accommodation to students who were to be excluded from participating in NAEP mathematics. Using data from the grade 8 special study participants as well as students in the reporting sample, both SD and non-SD, we performed analyses to address the key questions.

Effects on Student Performance

Students in the calculator study achieved mean scale scores similar to other students with disabilities (237.8 compared to 235.7 to 240.9 for SD requiring a calculator accommodation, but participating without one and 246.0 for SD not requiring a calculator accommodation). Both
calculator study and operational NAEP SD participants scored lower than non-SD (who averaged 286.1). Similarly, using overall percent correct reveals the same pattern of SD in the calculator study with 27 percent correct and other SD scoring between 25 percent and 31 percent correct. In comparison, non-SD scored between 48 percent and 50 percent correct, significantly higher than SD, depending on how many calculator blocks they received. Students with disabilities who do not require calculator accommodations tended to perform at a slightly higher level than students in the calculator study. The confidence interval of mean scale scores overlaps for SD without calculator accommodations but who received two calculator blocks and SD in the calculator study (also with two calculator blocks). Comparisons of overall percent correct showed no overlap. The similarity in mean scores for SD tested with and without a calculator accommodation suggests that providing a calculator accommodation for SD, when appropriate, should have little, if any, effect on performance of students with disabilities. However, reducing the proportion of SD excluded from NAEP will lower overall mean score estimates.

Additional measures of student performance included the percentage of items omitted and the percentage of items not reached. Students tended to omit few items; there were no differences in omit rates between those in the calculator study and other SD who required calculator accommodations. NAEP is not considered a speeded test. Thus, students generally finished the assessment, leaving very few items unanswered at the end of each block. Students with disabilities who were included in NAEP operationally without calculator accommodations, but received two calculator blocks, were slightly more likely to finish the assessment than students in the calculator study who also received two calculator blocks. This outcome was evident in the first block but not the second block.

**Effects on State Reporting**

Increasing inclusion on NAEP mathematics assessments by offering calculator accommodations is largely a state issue, driven in part by state policies of allowing the use of calculators on state assessments. The use of calculators on NAEP is likely to extend beyond states with assessment policies allowing calculators, as students in states that do not allow calculators on state exams may use them in classroom instruction and may have a calculator accommodation on their IEPs. Such students may be eligible for calculator accommodations on NAEP.

The effect on state level reporting is a function of the number of students using the accommodation. The more students that a state includes who were previously excluded from NAEP, the greater the impact on state level reporting. This was clearly seen in Oklahoma which would have had the greatest drop in exclusion rates if they had included students who participated in the calculator study. States exhibiting performance losses on NAEP will likely experience even greater losses or more subgroups losing ground, such as noted in Maryland which would have had a relatively large drop in its exclusion rate with the addition of a calculator accommodation. There were fewer impacts on state reporting in New Jersey and Texas where there would have been smaller decreases in exclusion rates.

States with a large number of students participating in NAEP with calculator accommodations will need to determine the extent of the impact on NAEP results. The calculator accommodation may be an additional accommodation for students who previously would have participated, or it may be the accommodation that allows a previously excluded student to be included. Increases in inclusion, particularly large increases, may affect NAEP state results because there is a performance gap between students with disabilities and those with no disabilities. Increases in the SD population participating in NAEP have the potential to decrease NAEP state results.
States with the largest drop in exclusion rates may need additional assistance to explain changes in NAEP policy when NAEP state data are released.

**Comparison of Students Requiring Calculator Accommodations**

With respect to student performance, individuals in the calculator study were similar to other students with disabilities, including those who use calculator accommodations on state assessments. This is true for SD who participate in NAEP without calculator accommodations as well as SD who do not need a calculator accommodation. The gap between SD and non-SD remains, with non-SD outscoring their SD peers.

We compared students in the calculator study to those participating in NAEP as well as those excluded from NAEP. The calculator study included more males than females. There was no appreciable difference in the gender distribution of students in the calculator study compared to SD included in NAEP or SD excluded from NAEP. The calculator study included somewhat more minorities than other SD groups (e.g., SD Calc and SD NoCalc), most closely resembling SD excluded from NAEP. Similarly, in the calculator study more students received free or reduced meals than other SD groups, particularly non-SD, and most closely resembling excluded students. Thus, there are some differences between students with disabilities in the calculator study and SD in the reporting sample of NAEP with appropriate accommodations (but not a calculator accommodation). Overall, calculator study students are similar to SD in operational NAEP but more likely to be minorities participating in the NSLP.

The degree or severity of disability as well as the different type of disability lead to differences in grade level performance. Students in the calculator study were generally reported as having mild or moderate disabilities as were SD in the reporting sample. In contrast, students excluded from NAEP were those with the most profound disabilities. The degree of disability is related to the grade level of performance, generally the milder the disability the closer to grade level the student performs, and the more profound the disability the farther from grade level the student performs. Students in the calculator study were slightly more likely than SD included in NAEP to be 2 or more years below grade level and less likely to be on or above grade level. However, students in the calculator study more closely resembled SD included in NAEP than excluded students. The major types of disabilities affecting SD were similar across students in the calculator study and those included in NAEP.

There are some differences in the accommodations needed by SD in the calculator study and those in the reporting sample of NAEP. Students in the calculator study were much more likely to need three or more accommodations. In addition, those in the calculator study were more likely to use “read aloud occasionally” and “read aloud” all than SD in operational NAEP.

**Conclusion**

Data from the 2011 calculator study indicate that students who were previously excluded from NAEP because calculator accommodations were not available may reasonably participate in NAEP. Performance of students requiring calculator accommodations is not significantly different from other SD, but is lower than students without disabilities. The gap between SD and non-SD contributes to the size of the effect on state-level reporting. The more students who were excluded on the basis of calculator accommodations in a state, the greater the impact on state results.

Although there are some similarities and some differences between students in the calculator study (i.e., those who were previously excluded but may be included with a calculator accommodation),
the differences are small and don’t have a significant impact on student performance or participation. Students with profound disabilities and students performing well below grade level would still be excluded even with the availability of a calculator accommodation on NAEP.
References


National Center for Education Statistics (NCES). Embargoed data.
Comparison of change in average reading scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 4: By state

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>218.6</td>
<td>220.3</td>
<td>-1.7</td>
<td>217.8</td>
<td>218.7</td>
<td>-0.9</td>
<td>-0.8</td>
</tr>
<tr>
<td>Alaska</td>
<td>209.3</td>
<td>207.9</td>
<td>1.4</td>
<td>208.4</td>
<td>206.8</td>
<td>1.6</td>
<td>-0.2</td>
</tr>
<tr>
<td>Arizona</td>
<td>213.1</td>
<td>212.4</td>
<td>0.7</td>
<td>212.4</td>
<td>211.3</td>
<td>1.1</td>
<td>-0.3</td>
</tr>
<tr>
<td>Arkansas</td>
<td>218.5</td>
<td>216.5</td>
<td>2.0</td>
<td>217.8</td>
<td>215.7</td>
<td>2.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>California</td>
<td>212.5</td>
<td>211.4</td>
<td>1.2</td>
<td>210.7</td>
<td>209.9</td>
<td>0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Colorado</td>
<td>226.7</td>
<td>223.4</td>
<td>3.2</td>
<td>225.6</td>
<td>222.4</td>
<td>3.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Connecticut</td>
<td>229.6</td>
<td>227.4</td>
<td>2.2</td>
<td>228.7</td>
<td>226.0</td>
<td>2.6</td>
<td>-0.5</td>
</tr>
<tr>
<td>Delaware</td>
<td>225.8</td>
<td>225.1</td>
<td>0.6</td>
<td>222.8</td>
<td>221.0</td>
<td>1.8</td>
<td>-1.1</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>205.6</td>
<td>200.6</td>
<td>5.0</td>
<td>204.8</td>
<td>199.0</td>
<td>5.8</td>
<td>-0.8</td>
</tr>
<tr>
<td>DoDEA</td>
<td>232.1</td>
<td>229.5</td>
<td>2.6</td>
<td>229.0</td>
<td>226.0</td>
<td>3.0</td>
<td>-0.4</td>
</tr>
<tr>
<td>Florida</td>
<td>227.5</td>
<td>224.5</td>
<td>2.9</td>
<td>226.0</td>
<td>223.3</td>
<td>2.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Georgia</td>
<td>221.8</td>
<td>220.8</td>
<td>1.0</td>
<td>219.2</td>
<td>217.2</td>
<td>2.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>Hawaii</td>
<td>214.8</td>
<td>213.6</td>
<td>1.2</td>
<td>213.6</td>
<td>212.4</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Idaho</td>
<td>219.3</td>
<td>220.8</td>
<td>-1.5</td>
<td>218.2</td>
<td>219.4</td>
<td>-1.1</td>
<td>-0.4</td>
</tr>
<tr>
<td>Illinois</td>
<td>218.5</td>
<td>220.7</td>
<td>-0.8</td>
<td>217.9</td>
<td>218.3</td>
<td>-0.3</td>
<td>-0.5</td>
</tr>
<tr>
<td>Indiana</td>
<td>225.3</td>
<td>220.7</td>
<td>4.6</td>
<td>224.1</td>
<td>220.0</td>
<td>4.1</td>
<td>0.5</td>
</tr>
<tr>
<td>Iowa</td>
<td>223.8</td>
<td>221.7</td>
<td>2.1</td>
<td>223.1</td>
<td>219.9</td>
<td>3.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Kansas</td>
<td>223.4</td>
<td>223.6</td>
<td>-0.2</td>
<td>222.3</td>
<td>222.0</td>
<td>0.3</td>
<td>-0.4</td>
</tr>
<tr>
<td>Kentucky</td>
<td>224.4</td>
<td>225.1</td>
<td>-0.7</td>
<td>222.5</td>
<td>222.0</td>
<td>2.3</td>
<td>-3.0</td>
</tr>
<tr>
<td>Louisiana</td>
<td>210.5</td>
<td>210.4</td>
<td>0.0</td>
<td>209.8</td>
<td>209.7</td>
<td>0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>Maine</td>
<td>224.8</td>
<td>222.0</td>
<td>2.8</td>
<td>223.9</td>
<td>221.3</td>
<td>2.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Maryland</td>
<td>232.1</td>
<td>230.8</td>
<td>1.3</td>
<td>224.5</td>
<td>225.6</td>
<td>-1.1</td>
<td>2.3</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>232.4</td>
<td>236.8</td>
<td>-4.4</td>
<td>230.8</td>
<td>233.9</td>
<td>-3.1</td>
<td>-1.3</td>
</tr>
<tr>
<td>Michigan</td>
<td>217.4</td>
<td>218.9</td>
<td>-1.5</td>
<td>215.0</td>
<td>216.4</td>
<td>-1.4</td>
<td>-0.1</td>
</tr>
<tr>
<td>Minnesota</td>
<td>227.0</td>
<td>222.3</td>
<td>4.7</td>
<td>225.4</td>
<td>221.2</td>
<td>4.2</td>
<td>0.6</td>
</tr>
<tr>
<td>Mississippi</td>
<td>208.5</td>
<td>209.2</td>
<td>-0.7</td>
<td>208.2</td>
<td>208.6</td>
<td>-0.4</td>
<td>-0.3</td>
</tr>
<tr>
<td>Missouri</td>
<td>222.3</td>
<td>220.3</td>
<td>2.0</td>
<td>221.6</td>
<td>219.3</td>
<td>2.2</td>
<td>-0.3</td>
</tr>
<tr>
<td>Montana</td>
<td>223.0</td>
<td>225.1</td>
<td>-2.2</td>
<td>221.2</td>
<td>222.4</td>
<td>-1.2</td>
<td>-1.0</td>
</tr>
<tr>
<td>Nebraska</td>
<td>223.3</td>
<td>223.3</td>
<td>-0.1</td>
<td>221.0</td>
<td>220.7</td>
<td>0.3</td>
<td>-0.4</td>
</tr>
<tr>
<td>Nevada</td>
<td>213.8</td>
<td>212.6</td>
<td>1.2</td>
<td>212.9</td>
<td>211.9</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>232.0</td>
<td>230.4</td>
<td>1.6</td>
<td>230.7</td>
<td>229.1</td>
<td>1.6</td>
<td>0.0</td>
</tr>
<tr>
<td>New Jersey</td>
<td>228.7</td>
<td>231.2</td>
<td>-2.5</td>
<td>227.8</td>
<td>227.1</td>
<td>0.7</td>
<td>-3.2</td>
</tr>
<tr>
<td>New Mexico</td>
<td>205.8</td>
<td>208.0</td>
<td>-2.3</td>
<td>205.2</td>
<td>205.1</td>
<td>0.1</td>
<td>-2.3</td>
</tr>
<tr>
<td>New York</td>
<td>223.8</td>
<td>222.5</td>
<td>1.3</td>
<td>223.1</td>
<td>221.3</td>
<td>1.7</td>
<td>-0.4</td>
</tr>
<tr>
<td>North Carolina</td>
<td>222.2</td>
<td>221.4</td>
<td>0.9</td>
<td>221.0</td>
<td>219.9</td>
<td>1.1</td>
<td>-0.3</td>
</tr>
<tr>
<td>North Dakota</td>
<td>244.1</td>
<td>225.6</td>
<td>-1.5</td>
<td>221.5</td>
<td>222.1</td>
<td>-0.6</td>
<td>-0.9</td>
</tr>
<tr>
<td>Ohio</td>
<td>239.3</td>
<td>233.8</td>
<td>0.1</td>
<td>222.5</td>
<td>220.5</td>
<td>2.0</td>
<td>-1.9</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>217.0</td>
<td>215.5</td>
<td>1.5</td>
<td>215.9</td>
<td>212.4</td>
<td>3.4</td>
<td>-1.9</td>
</tr>
<tr>
<td>Oregon</td>
<td>219.1</td>
<td>216.4</td>
<td>2.7</td>
<td>217.2</td>
<td>214.4</td>
<td>2.9</td>
<td>-0.2</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>226.4</td>
<td>227.2</td>
<td>-0.8</td>
<td>225.1</td>
<td>225.2</td>
<td>-0.1</td>
<td>-0.8</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>222.8</td>
<td>222.5</td>
<td>0.3</td>
<td>221.9</td>
<td>221.3</td>
<td>0.6</td>
<td>-0.3</td>
</tr>
<tr>
<td>South Carolina</td>
<td>213.6</td>
<td>214.9</td>
<td>-1.3</td>
<td>212.4</td>
<td>212.8</td>
<td>-0.5</td>
<td>-0.8</td>
</tr>
<tr>
<td>South Dakota</td>
<td>217.9</td>
<td>218.6</td>
<td>-1.9</td>
<td>216.1</td>
<td>217.8</td>
<td>-1.6</td>
<td>-0.2</td>
</tr>
<tr>
<td>Tennessee</td>
<td>219.7</td>
<td>214.6</td>
<td>5.1</td>
<td>217.8</td>
<td>210.3</td>
<td>7.5</td>
<td>-2.4</td>
</tr>
<tr>
<td>Texas</td>
<td>216.9</td>
<td>218.3</td>
<td>-1.4</td>
<td>214.1</td>
<td>213.2</td>
<td>1.0</td>
<td>-2.4</td>
</tr>
<tr>
<td>Utah</td>
<td>222.8</td>
<td>220.4</td>
<td>2.4</td>
<td>220.6</td>
<td>217.7</td>
<td>2.9</td>
<td>-0.5</td>
</tr>
<tr>
<td>Vermont</td>
<td>228.0</td>
<td>226.8</td>
<td>1.2</td>
<td>227.3</td>
<td>225.3</td>
<td>1.9</td>
<td>-0.7</td>
</tr>
<tr>
<td>Virginia</td>
<td>228.6</td>
<td>226.4</td>
<td>2.2</td>
<td>227.4</td>
<td>224.4</td>
<td>3.1</td>
<td>-0.9</td>
</tr>
<tr>
<td>Washington</td>
<td>225.0</td>
<td>220.5</td>
<td>4.5</td>
<td>223.1</td>
<td>218.7</td>
<td>4.4</td>
<td>0.1</td>
</tr>
<tr>
<td>West Virginia</td>
<td>214.7</td>
<td>214.4</td>
<td>0.3</td>
<td>213.5</td>
<td>213.1</td>
<td>0.4</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>2013</td>
<td>Change</td>
<td>2011</td>
<td>2013</td>
<td>Change</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>220.8</td>
<td>221.2</td>
<td>-0.4</td>
<td>219.7</td>
<td>219.9</td>
<td>-0.3</td>
<td></td>
</tr>
<tr>
<td>Wyoming¹</td>
<td>225.8</td>
<td>224.1</td>
<td>1.8</td>
<td>225.1</td>
<td>222.9</td>
<td>2.1*</td>
<td></td>
</tr>
</tbody>
</table>

¹Statistically significant change.

The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.

²Department of Defense Education Activity (domestic and overseas schools).


Read more about research on exclusion rates.

__________________________________________________________

Last updated 05 November 2013 (FW)

__________________________________________________________

U.S. Department of Education
Comparison of change in average mathematics scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 4: By state

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>232.9</td>
<td>231.3</td>
<td>1.6</td>
<td>232.3</td>
<td>230.6</td>
<td>1.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Alaska</td>
<td>236.1</td>
<td>236.4</td>
<td>-0.2</td>
<td>235.7</td>
<td>235.5</td>
<td>0.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>Arizona</td>
<td>240.3</td>
<td>235.2</td>
<td>5.1</td>
<td>239.7</td>
<td>234.6</td>
<td>5.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Arkansas</td>
<td>239.9</td>
<td>237.8</td>
<td>2.1</td>
<td>239.3</td>
<td>237.4</td>
<td>1.9</td>
<td>0.2</td>
</tr>
<tr>
<td>California</td>
<td>233.7</td>
<td>234.2</td>
<td>-0.5</td>
<td>232.6</td>
<td>233.4</td>
<td>-0.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Colorado</td>
<td>247.0</td>
<td>244.5</td>
<td>2.5</td>
<td>246.3</td>
<td>243.9</td>
<td>2.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Connecticut</td>
<td>243.4</td>
<td>242.4</td>
<td>1.0</td>
<td>242.8</td>
<td>241.9</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td>Delaware</td>
<td>243.1</td>
<td>240.4</td>
<td>2.7</td>
<td>242.1</td>
<td>239.0</td>
<td>3.1</td>
<td>-0.4</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>228.6</td>
<td>221.8</td>
<td>6.8</td>
<td>228.0</td>
<td>219.8</td>
<td>8.2</td>
<td>-1.5</td>
</tr>
<tr>
<td>DoDEA</td>
<td>245.0</td>
<td>240.8</td>
<td>4.2</td>
<td>244.3</td>
<td>239.9</td>
<td>4.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>Florida</td>
<td>241.7</td>
<td>239.9</td>
<td>1.8</td>
<td>240.9</td>
<td>239.2</td>
<td>1.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Georgia</td>
<td>240.0</td>
<td>238.4</td>
<td>1.7</td>
<td>239.3</td>
<td>237.6</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Hawaii</td>
<td>243.3</td>
<td>238.8</td>
<td>4.5</td>
<td>242.6</td>
<td>237.8</td>
<td>4.8</td>
<td>-0.3</td>
</tr>
<tr>
<td>Idaho</td>
<td>240.7</td>
<td>240.3</td>
<td>0.4</td>
<td>240.0</td>
<td>239.7</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Illinois</td>
<td>239.0</td>
<td>238.9</td>
<td>0.2</td>
<td>238.6</td>
<td>237.8</td>
<td>0.8</td>
<td>-0.6</td>
</tr>
<tr>
<td>Indiana</td>
<td>248.6</td>
<td>243.8</td>
<td>4.8</td>
<td>248.0</td>
<td>242.9</td>
<td>5.1</td>
<td>-0.3</td>
</tr>
<tr>
<td>Iowa</td>
<td>245.8</td>
<td>242.6</td>
<td>3.2</td>
<td>245.4</td>
<td>242.0</td>
<td>3.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>Kansas</td>
<td>246.2</td>
<td>246.3</td>
<td>-0.1</td>
<td>245.5</td>
<td>245.5</td>
<td>0.0</td>
<td>-0.1</td>
</tr>
<tr>
<td>Kentucky</td>
<td>241.5</td>
<td>240.8</td>
<td>0.6</td>
<td>240.8</td>
<td>239.7</td>
<td>1.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>Louisiana</td>
<td>231.4</td>
<td>230.8</td>
<td>0.6</td>
<td>230.9</td>
<td>230.3</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Maine</td>
<td>245.8</td>
<td>244.3</td>
<td>1.6</td>
<td>244.8</td>
<td>243.6</td>
<td>1.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Maryland</td>
<td>245.2</td>
<td>247.1</td>
<td>-1.9</td>
<td>244.7</td>
<td>245.2</td>
<td>-0.6</td>
<td>-1.4</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>253.0</td>
<td>253.4</td>
<td>-0.4</td>
<td>252.1</td>
<td>252.2</td>
<td>-0.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>Michigan</td>
<td>236.8</td>
<td>236.4</td>
<td>0.4</td>
<td>236.0</td>
<td>235.4</td>
<td>0.6</td>
<td>-0.2</td>
</tr>
<tr>
<td>Minnesota</td>
<td>253.4</td>
<td>249.2</td>
<td>4.2</td>
<td>252.8</td>
<td>248.6</td>
<td>4.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Mississippi</td>
<td>231.1</td>
<td>229.9</td>
<td>1.2</td>
<td>230.8</td>
<td>229.5</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Missouri</td>
<td>239.5</td>
<td>240.5</td>
<td>-0.9</td>
<td>238.8</td>
<td>239.8</td>
<td>-1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Montana</td>
<td>243.7</td>
<td>243.8</td>
<td>-0.1</td>
<td>242.7</td>
<td>243.0</td>
<td>-0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>Nebraska</td>
<td>243.2</td>
<td>239.8</td>
<td>3.4</td>
<td>242.4</td>
<td>239.2</td>
<td>3.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Nevada</td>
<td>236.3</td>
<td>237.0</td>
<td>-0.8</td>
<td>235.7</td>
<td>236.2</td>
<td>-0.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>253.0</td>
<td>251.8</td>
<td>1.2</td>
<td>252.5</td>
<td>251.1</td>
<td>1.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>New Jersey</td>
<td>246.9</td>
<td>248.0</td>
<td>-1.1</td>
<td>246.3</td>
<td>246.9</td>
<td>-0.5</td>
<td>-0.6</td>
</tr>
<tr>
<td>New Mexico</td>
<td>232.8</td>
<td>232.8</td>
<td>-0.1</td>
<td>232.3</td>
<td>232.1</td>
<td>0.2</td>
<td>-0.3</td>
</tr>
<tr>
<td>New York</td>
<td>240.3</td>
<td>237.5</td>
<td>2.8</td>
<td>239.8</td>
<td>237.1</td>
<td>2.8</td>
<td>0.1</td>
</tr>
<tr>
<td>North Carolina</td>
<td>244.8</td>
<td>244.5</td>
<td>0.3</td>
<td>244.2</td>
<td>243.7</td>
<td>0.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>North Dakota</td>
<td>246.4</td>
<td>245.2</td>
<td>1.3</td>
<td>245.1</td>
<td>243.6</td>
<td>1.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>Ohio</td>
<td>245.5</td>
<td>244.0</td>
<td>1.5</td>
<td>244.9</td>
<td>243.0</td>
<td>1.9</td>
<td>-0.4</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>238.9</td>
<td>237.4</td>
<td>1.5</td>
<td>238.1</td>
<td>234.1</td>
<td>4.0</td>
<td>-2.5</td>
</tr>
<tr>
<td>Oregon</td>
<td>240.1</td>
<td>236.9</td>
<td>3.2</td>
<td>239.1</td>
<td>235.6</td>
<td>3.5</td>
<td>-0.3</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>244.0</td>
<td>245.7</td>
<td>-1.6</td>
<td>243.2</td>
<td>244.9</td>
<td>-1.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>241.4</td>
<td>241.6</td>
<td>-0.2</td>
<td>240.9</td>
<td>241.3</td>
<td>-0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>South Carolina</td>
<td>236.6</td>
<td>237.3</td>
<td>-0.7</td>
<td>236.0</td>
<td>236.7</td>
<td>-0.6</td>
<td>-0.1</td>
</tr>
<tr>
<td>South Dakota</td>
<td>241.0</td>
<td>241.0</td>
<td>0.0</td>
<td>240.2</td>
<td>240.0</td>
<td>0.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>Tennessee</td>
<td>239.8</td>
<td>232.9</td>
<td>6.9</td>
<td>239.1</td>
<td>231.5</td>
<td>7.6</td>
<td>-0.7</td>
</tr>
<tr>
<td>Texas</td>
<td>241.9</td>
<td>241.1</td>
<td>0.8</td>
<td>241.1</td>
<td>239.5</td>
<td>1.6</td>
<td>-0.8</td>
</tr>
<tr>
<td>Utah</td>
<td>242.8</td>
<td>242.5</td>
<td>0.3</td>
<td>242.2</td>
<td>241.8</td>
<td>0.4</td>
<td>-0.1</td>
</tr>
<tr>
<td>Vermont</td>
<td>247.8</td>
<td>246.6</td>
<td>1.2</td>
<td>247.1</td>
<td>246.0</td>
<td>1.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>Virginia</td>
<td>246.2</td>
<td>245.3</td>
<td>0.8</td>
<td>245.5</td>
<td>244.4</td>
<td>1.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>Washington</td>
<td>246.3</td>
<td>243.2</td>
<td>3.1</td>
<td>245.2</td>
<td>242.4</td>
<td>2.9</td>
<td>0.2</td>
</tr>
<tr>
<td>West Virginia</td>
<td>237.4</td>
<td>234.7</td>
<td>2.8</td>
<td>236.7</td>
<td>234.0</td>
<td>2.7</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Read more about research on exclusion rates.

<table>
<thead>
<tr>
<th></th>
<th>2011</th>
<th>2013</th>
<th>Diff</th>
<th>2011</th>
<th>2013</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wisconsin</td>
<td>244.7</td>
<td>244.7</td>
<td>0.0</td>
<td>243.8</td>
<td>243.9</td>
<td>-0.1</td>
</tr>
<tr>
<td>Wyoming</td>
<td>246.5</td>
<td>243.9</td>
<td>2.6 *</td>
<td>246.1</td>
<td>243.3</td>
<td>2.8 *</td>
</tr>
</tbody>
</table>

* Statistically significant change.

1The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.

2Department of Defense Education Activity (domestic and overseas schools).


Last updated 05 November 2013 (FW)

U.S. Department of Education
## Comparison of change in average reading scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 8: By state

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>257.4</td>
<td>258.4</td>
<td>-1.0</td>
<td>256.6</td>
<td>257.1</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>Alaska</td>
<td>261.3</td>
<td>261.3</td>
<td>0.0</td>
<td>260.6</td>
<td>260.4</td>
<td>0.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>Arizona</td>
<td>260.4</td>
<td>260.1</td>
<td>0.3</td>
<td>259.5</td>
<td>259.5</td>
<td>0.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Arkansas</td>
<td>262.0</td>
<td>259.1</td>
<td>2.8 *</td>
<td>260.7</td>
<td>258.2</td>
<td>2.6</td>
<td>0.3</td>
</tr>
<tr>
<td>California</td>
<td>261.5</td>
<td>254.9</td>
<td>6.6 *</td>
<td>260.1</td>
<td>253.6</td>
<td>6.5 *</td>
<td>0.1</td>
</tr>
<tr>
<td>Colorado</td>
<td>271.0</td>
<td>270.6</td>
<td>0.4</td>
<td>270.3</td>
<td>269.6</td>
<td>0.8</td>
<td>-0.4</td>
</tr>
<tr>
<td>Connecticut</td>
<td>274.5</td>
<td>274.7</td>
<td>-0.2</td>
<td>273.2</td>
<td>273.4</td>
<td>-0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>Delaware</td>
<td>266.0</td>
<td>265.8</td>
<td>0.1</td>
<td>264.2</td>
<td>263.2</td>
<td>0.9</td>
<td>-0.8</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>247.7</td>
<td>242.1</td>
<td>5.7 *</td>
<td>246.8</td>
<td>240.8</td>
<td>6.0 *</td>
<td>-0.3</td>
</tr>
<tr>
<td>DoDEA</td>
<td>277.0</td>
<td>272.3</td>
<td>4.6 *</td>
<td>275.0</td>
<td>270.6</td>
<td>4.4 *</td>
<td>0.2</td>
</tr>
<tr>
<td>Florida</td>
<td>265.8</td>
<td>262.1</td>
<td>3.7 *</td>
<td>264.9</td>
<td>260.8</td>
<td>4.1 *</td>
<td>-0.3</td>
</tr>
<tr>
<td>Georgia</td>
<td>264.6</td>
<td>262.4</td>
<td>2.3</td>
<td>262.5</td>
<td>260.3</td>
<td>2.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Hawaii</td>
<td>260.0</td>
<td>257.2</td>
<td>2.8 *</td>
<td>259.8</td>
<td>255.9</td>
<td>3.0 *</td>
<td>-0.2</td>
</tr>
<tr>
<td>Idaho</td>
<td>270.2</td>
<td>268.0</td>
<td>2.2</td>
<td>269.1</td>
<td>267.0</td>
<td>2.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Illinois</td>
<td>266.9</td>
<td>265.6</td>
<td>1.3</td>
<td>266.2</td>
<td>264.7</td>
<td>1.5</td>
<td>-0.2</td>
</tr>
<tr>
<td>Indiana</td>
<td>267.3</td>
<td>264.7</td>
<td>2.5</td>
<td>266.2</td>
<td>263.4</td>
<td>2.8</td>
<td>-0.3</td>
</tr>
<tr>
<td>Iowa</td>
<td>269.0</td>
<td>264.6</td>
<td>4.4 *</td>
<td>268.3</td>
<td>264.1</td>
<td>4.2 *</td>
<td>0.3</td>
</tr>
<tr>
<td>Kansas</td>
<td>266.9</td>
<td>267.3</td>
<td>-0.4</td>
<td>265.9</td>
<td>266.2</td>
<td>-0.3</td>
<td>-0.2</td>
</tr>
<tr>
<td>Kentucky</td>
<td>269.6</td>
<td>268.8</td>
<td>0.8</td>
<td>267.8</td>
<td>265.2</td>
<td>2.5 *</td>
<td>-1.8</td>
</tr>
<tr>
<td>Louisiana</td>
<td>257.4</td>
<td>254.7</td>
<td>2.7</td>
<td>256.8</td>
<td>254.3</td>
<td>2.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Maine</td>
<td>269.2</td>
<td>269.9</td>
<td>-0.7</td>
<td>268.5</td>
<td>269.0</td>
<td>-0.5</td>
<td>-0.1</td>
</tr>
<tr>
<td>Maryland</td>
<td>273.8</td>
<td>271.2</td>
<td>2.6</td>
<td>269.0</td>
<td>266.9</td>
<td>2.1</td>
<td>0.5</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>277.0</td>
<td>275.4</td>
<td>1.6</td>
<td>275.7</td>
<td>272.3</td>
<td>3.4 *</td>
<td>-1.8</td>
</tr>
<tr>
<td>Michigan</td>
<td>265.9</td>
<td>265.2</td>
<td>0.7</td>
<td>264.2</td>
<td>262.3</td>
<td>2.0</td>
<td>-1.2</td>
</tr>
<tr>
<td>Minnesota</td>
<td>271.0</td>
<td>270.1</td>
<td>0.9</td>
<td>269.7</td>
<td>268.4</td>
<td>1.3</td>
<td>-0.4</td>
</tr>
<tr>
<td>Mississippi</td>
<td>253.1</td>
<td>253.8</td>
<td>-0.7</td>
<td>252.8</td>
<td>253.2</td>
<td>-0.4</td>
<td>-0.3</td>
</tr>
<tr>
<td>Missouri</td>
<td>267.2</td>
<td>266.8</td>
<td>0.5</td>
<td>266.6</td>
<td>265.8</td>
<td>0.8</td>
<td>-0.4</td>
</tr>
<tr>
<td>Montana</td>
<td>271.8</td>
<td>272.9</td>
<td>-1.1</td>
<td>270.5</td>
<td>270.6</td>
<td>-0.1</td>
<td>-1.0</td>
</tr>
<tr>
<td>Nebraska</td>
<td>269.2</td>
<td>267.7</td>
<td>1.5</td>
<td>267.6</td>
<td>264.9</td>
<td>2.7 *</td>
<td>-1.2</td>
</tr>
<tr>
<td>Nevada</td>
<td>261.7</td>
<td>258.2</td>
<td>3.5 *</td>
<td>261.1</td>
<td>257.2</td>
<td>4.0 *</td>
<td>-0.5</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>274.3</td>
<td>272.1</td>
<td>2.2 *</td>
<td>272.9</td>
<td>270.2</td>
<td>2.7 *</td>
<td>-0.5</td>
</tr>
<tr>
<td>New Jersey</td>
<td>276.4</td>
<td>275.2</td>
<td>1.2</td>
<td>275.3</td>
<td>272.2</td>
<td>3.1</td>
<td>-1.9</td>
</tr>
<tr>
<td>New Mexico</td>
<td>255.9</td>
<td>255.9</td>
<td>0.0</td>
<td>255.1</td>
<td>253.3</td>
<td>1.7</td>
<td>-1.7</td>
</tr>
<tr>
<td>New York</td>
<td>266.3</td>
<td>265.7</td>
<td>0.6</td>
<td>265.8</td>
<td>264.3</td>
<td>1.5</td>
<td>-0.9</td>
</tr>
<tr>
<td>North Carolina</td>
<td>264.5</td>
<td>262.9</td>
<td>1.6</td>
<td>263.5</td>
<td>261.8</td>
<td>1.8</td>
<td>-0.2</td>
</tr>
<tr>
<td>North Dakota</td>
<td>267.8</td>
<td>268.7</td>
<td>-0.9</td>
<td>265.2</td>
<td>264.9</td>
<td>0.3</td>
<td>-1.2</td>
</tr>
<tr>
<td>Ohio</td>
<td>269.1</td>
<td>268.3</td>
<td>0.8</td>
<td>267.8</td>
<td>265.5</td>
<td>2.4</td>
<td>-1.6</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>261.9</td>
<td>260.1</td>
<td>1.8</td>
<td>261.1</td>
<td>257.8</td>
<td>3.3 *</td>
<td>-1.5</td>
</tr>
<tr>
<td>Oregon</td>
<td>268.3</td>
<td>264.2</td>
<td>4.1 *</td>
<td>267.4</td>
<td>262.9</td>
<td>4.5 *</td>
<td>-0.5</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>272.1</td>
<td>267.8</td>
<td>4.3 *</td>
<td>271.1</td>
<td>266.1</td>
<td>5.0 *</td>
<td>-0.6</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>266.7</td>
<td>265.1</td>
<td>1.6</td>
<td>265.9</td>
<td>264.4</td>
<td>1.5</td>
<td>0.1</td>
</tr>
<tr>
<td>South Carolina</td>
<td>261.4</td>
<td>260.3</td>
<td>1.1</td>
<td>260.3</td>
<td>257.6</td>
<td>2.7</td>
<td>-1.6</td>
</tr>
<tr>
<td>South Dakota</td>
<td>268.1</td>
<td>268.9</td>
<td>-0.8</td>
<td>266.4</td>
<td>266.9</td>
<td>-0.5</td>
<td>-0.3</td>
</tr>
<tr>
<td>Tennessee</td>
<td>265.4</td>
<td>259.2</td>
<td>6.2 *</td>
<td>263.7</td>
<td>255.8</td>
<td>7.9 *</td>
<td>-1.7</td>
</tr>
<tr>
<td>Texas</td>
<td>263.7</td>
<td>261.4</td>
<td>2.3</td>
<td>261.7</td>
<td>257.7</td>
<td>4.0 *</td>
<td>-1.8</td>
</tr>
<tr>
<td>Utah</td>
<td>270.0</td>
<td>267.1</td>
<td>3.0 *</td>
<td>268.4</td>
<td>264.8</td>
<td>3.5 *</td>
<td>-0.6</td>
</tr>
<tr>
<td>Vermont</td>
<td>274.4</td>
<td>273.8</td>
<td>0.5</td>
<td>273.9</td>
<td>272.2</td>
<td>1.7</td>
<td>-1.2</td>
</tr>
<tr>
<td>Virginia</td>
<td>267.6</td>
<td>267.3</td>
<td>0.3</td>
<td>266.9</td>
<td>265.3</td>
<td>1.6</td>
<td>-1.3</td>
</tr>
<tr>
<td>Washington</td>
<td>272.0</td>
<td>267.6</td>
<td>4.5 *</td>
<td>270.6</td>
<td>266.2</td>
<td>4.4 *</td>
<td>0.1</td>
</tr>
<tr>
<td>West Virginia</td>
<td>257.4</td>
<td>256.1</td>
<td>1.3</td>
<td>256.3</td>
<td>255.1</td>
<td>1.2</td>
<td>0.1</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>267.5</td>
<td>267.2</td>
<td>0.4</td>
<td>266.6</td>
<td>265.8</td>
<td>0.8</td>
<td>-0.4</td>
</tr>
<tr>
<td>Wyoming</td>
<td>271.0</td>
<td>269.6</td>
<td>1.4</td>
<td>270.5</td>
<td>268.5</td>
<td>2.0</td>
<td>-0.6</td>
</tr>
</tbody>
</table>

* Statistically significant change.
1 The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.
2 Department of Defense Education Activity (domestic and overseas schools).


Read more about research on exclusion rates.

Last updated 05 November 2013 (FW)

U.S. Department of Education
Comparison of change in average mathematics scores from 2011 to 2013 in the reported NAEP estimates and the full population estimates (FPE), grade 8: By state

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>269.2</td>
<td>269.1</td>
<td>0.1</td>
<td>268.4</td>
<td>268.3</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Alaska</td>
<td>281.6</td>
<td>283.3</td>
<td>-1.7</td>
<td>281.0</td>
<td>281.8</td>
<td>-0.8</td>
<td>-0.9</td>
</tr>
<tr>
<td>Arizona</td>
<td>279.7</td>
<td>279.0</td>
<td>0.7</td>
<td>278.9</td>
<td>278.4</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Arkansas</td>
<td>277.9</td>
<td>279.1</td>
<td>-1.2</td>
<td>276.6</td>
<td>278.2</td>
<td>-1.6</td>
<td>0.4</td>
</tr>
<tr>
<td>California</td>
<td>275.9</td>
<td>272.8</td>
<td>3.1</td>
<td>275.1</td>
<td>272.1</td>
<td>2.9</td>
<td>0.2</td>
</tr>
<tr>
<td>Colorado</td>
<td>289.7</td>
<td>291.7</td>
<td>-2.1</td>
<td>288.8</td>
<td>291.2</td>
<td>-2.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Connecticut</td>
<td>285.2</td>
<td>287.0</td>
<td>-1.8</td>
<td>284.0</td>
<td>286.2</td>
<td>-2.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Delaware</td>
<td>282.3</td>
<td>282.8</td>
<td>-0.4</td>
<td>281.7</td>
<td>281.2</td>
<td>0.5</td>
<td>-0.9</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>265.3</td>
<td>260.5</td>
<td>4.8</td>
<td>264.8</td>
<td>258.6</td>
<td>6.2</td>
<td>-1.4</td>
</tr>
<tr>
<td>DoDEAî</td>
<td>290.4</td>
<td>287.9</td>
<td>2.6</td>
<td>289.8</td>
<td>286.4</td>
<td>3.3</td>
<td>-0.8</td>
</tr>
<tr>
<td>Florida</td>
<td>280.9</td>
<td>277.8</td>
<td>3.0</td>
<td>279.9</td>
<td>276.9</td>
<td>3.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Georgia</td>
<td>279.2</td>
<td>278.5</td>
<td>0.7</td>
<td>278.2</td>
<td>277.0</td>
<td>1.2</td>
<td>-0.5</td>
</tr>
<tr>
<td>Hawaii</td>
<td>281.4</td>
<td>277.8</td>
<td>3.6</td>
<td>280.5</td>
<td>276.7</td>
<td>3.8</td>
<td>-0.2</td>
</tr>
<tr>
<td>Idaho</td>
<td>286.4</td>
<td>286.7</td>
<td>-0.3</td>
<td>285.7</td>
<td>285.9</td>
<td>-0.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>Illinois</td>
<td>284.9</td>
<td>283.2</td>
<td>1.7</td>
<td>284.3</td>
<td>282.0</td>
<td>2.3</td>
<td>-0.6</td>
</tr>
<tr>
<td>Indiana</td>
<td>287.8</td>
<td>285.0</td>
<td>2.8</td>
<td>286.8</td>
<td>283.9</td>
<td>2.9</td>
<td>-0.2</td>
</tr>
<tr>
<td>Iowa</td>
<td>285.1</td>
<td>284.9</td>
<td>0.1</td>
<td>284.5</td>
<td>284.1</td>
<td>0.4</td>
<td>-0.3</td>
</tr>
<tr>
<td>Kansas</td>
<td>289.5</td>
<td>289.6</td>
<td>-0.1</td>
<td>288.4</td>
<td>288.9</td>
<td>-0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>Kentucky</td>
<td>280.6</td>
<td>281.6</td>
<td>-1.0</td>
<td>279.3</td>
<td>280.2</td>
<td>-0.9</td>
<td>-0.1</td>
</tr>
<tr>
<td>Louisiana</td>
<td>272.8</td>
<td>272.8</td>
<td>-0.1</td>
<td>272.1</td>
<td>272.2</td>
<td>-0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Maine</td>
<td>288.7</td>
<td>288.7</td>
<td>0.0</td>
<td>288.0</td>
<td>287.9</td>
<td>0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>Maryland</td>
<td>286.6</td>
<td>288.0</td>
<td>-1.4</td>
<td>285.7</td>
<td>284.5</td>
<td>1.2</td>
<td>-2.5</td>
</tr>
<tr>
<td>Massachusettsî</td>
<td>300.6</td>
<td>298.5</td>
<td>2.1</td>
<td>299.4</td>
<td>296.5</td>
<td>2.9</td>
<td>-0.9</td>
</tr>
<tr>
<td>Michigan</td>
<td>280.1</td>
<td>280.2</td>
<td>0.0</td>
<td>278.8</td>
<td>278.2</td>
<td>0.6</td>
<td>-0.6</td>
</tr>
<tr>
<td>Minnesota</td>
<td>294.6</td>
<td>294.9</td>
<td>-0.4</td>
<td>293.4</td>
<td>293.7</td>
<td>-0.3</td>
<td>-0.1</td>
</tr>
<tr>
<td>Mississippi</td>
<td>271.2</td>
<td>269.2</td>
<td>1.9</td>
<td>270.7</td>
<td>268.6</td>
<td>2.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>Missouri</td>
<td>283.0</td>
<td>281.9</td>
<td>1.0</td>
<td>282.2</td>
<td>281.2</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Montana</td>
<td>289.2</td>
<td>292.9</td>
<td>-3.7</td>
<td>288.3</td>
<td>291.8</td>
<td>-3.5</td>
<td>-0.1</td>
</tr>
<tr>
<td>Nebraskaî</td>
<td>285.1</td>
<td>283.2</td>
<td>1.9</td>
<td>283.9</td>
<td>281.2</td>
<td>2.7</td>
<td>-0.8</td>
</tr>
<tr>
<td>Nevada</td>
<td>278.3</td>
<td>278.1</td>
<td>0.2</td>
<td>277.7</td>
<td>276.6</td>
<td>1.1</td>
<td>-0.9</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>295.7</td>
<td>292.1</td>
<td>3.6</td>
<td>295.1</td>
<td>291.2</td>
<td>3.9</td>
<td>-0.3</td>
</tr>
<tr>
<td>New Jersey</td>
<td>296.1</td>
<td>294.1</td>
<td>1.9</td>
<td>295.0</td>
<td>292.4</td>
<td>2.5</td>
<td>-0.6</td>
</tr>
<tr>
<td>New Mexico</td>
<td>272.8</td>
<td>274.5</td>
<td>-1.7</td>
<td>271.8</td>
<td>273.6</td>
<td>-1.8</td>
<td>0.1</td>
</tr>
<tr>
<td>New York</td>
<td>281.8</td>
<td>280.5</td>
<td>1.4</td>
<td>280.9</td>
<td>279.9</td>
<td>1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>North Carolina</td>
<td>285.6</td>
<td>288.3</td>
<td>-0.6</td>
<td>284.8</td>
<td>285.2</td>
<td>-0.4</td>
<td>-0.2</td>
</tr>
<tr>
<td>North Dakota</td>
<td>290.5</td>
<td>292.0</td>
<td>-1.5</td>
<td>288.7</td>
<td>290.0</td>
<td>-1.3</td>
<td>-0.2</td>
</tr>
<tr>
<td>Ohio</td>
<td>289.5</td>
<td>288.6</td>
<td>0.9</td>
<td>288.6</td>
<td>286.0</td>
<td>2.6</td>
<td>-1.7</td>
</tr>
<tr>
<td>Oklahomaî</td>
<td>275.5</td>
<td>279.2</td>
<td>-3.7</td>
<td>274.5</td>
<td>274.0</td>
<td>0.6</td>
<td>-4.2</td>
</tr>
<tr>
<td>Oregon</td>
<td>283.5</td>
<td>282.5</td>
<td>1.0</td>
<td>282.6</td>
<td>281.7</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>289.6</td>
<td>286.1</td>
<td>3.5</td>
<td>288.7</td>
<td>284.6</td>
<td>4.1</td>
<td>-0.6</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>284.1</td>
<td>282.9</td>
<td>1.2</td>
<td>283.5</td>
<td>282.2</td>
<td>1.3</td>
<td>-0.1</td>
</tr>
<tr>
<td>South Carolina</td>
<td>279.8</td>
<td>281.0</td>
<td>-1.2</td>
<td>279.0</td>
<td>278.9</td>
<td>0.1</td>
<td>-1.2</td>
</tr>
<tr>
<td>South Dakota</td>
<td>287.3</td>
<td>290.6</td>
<td>-3.3</td>
<td>286.4</td>
<td>289.4</td>
<td>-3.0</td>
<td>-0.3</td>
</tr>
<tr>
<td>Tennessee</td>
<td>277.7</td>
<td>274.0</td>
<td>3.7</td>
<td>276.6</td>
<td>272.1</td>
<td>4.5</td>
<td>-0.8</td>
</tr>
<tr>
<td>Texas</td>
<td>288.2</td>
<td>290.3</td>
<td>-2.1</td>
<td>287.1</td>
<td>287.9</td>
<td>-0.8</td>
<td>-1.4</td>
</tr>
<tr>
<td>Utah</td>
<td>284.3</td>
<td>283.3</td>
<td>1.0</td>
<td>283.3</td>
<td>281.7</td>
<td>1.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>Vermont</td>
<td>295.5</td>
<td>293.9</td>
<td>1.6</td>
<td>294.9</td>
<td>293.2</td>
<td>1.7</td>
<td>-0.1</td>
</tr>
<tr>
<td>Virginia</td>
<td>288.1</td>
<td>289.3</td>
<td>-1.1</td>
<td>287.5</td>
<td>287.5</td>
<td>0.0</td>
<td>-1.1</td>
</tr>
<tr>
<td>Washington</td>
<td>290.0</td>
<td>288.1</td>
<td>1.8</td>
<td>288.7</td>
<td>287.1</td>
<td>1.6</td>
<td>0.3</td>
</tr>
<tr>
<td>West Virginia</td>
<td>274.4</td>
<td>273.3</td>
<td>1.2</td>
<td>273.3</td>
<td>272.4</td>
<td>0.9</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Read more about research on exclusion rates.

* Statistically significant change.
1The significance of the trend results from the officially reported sample for this jurisdiction would be different under the scenario.
2Department of Defense Education Activity (domestic and overseas schools).
A Validity Study of the NAEP Full Population Estimates

Larry V. Hedges
Northwestern University

Victor Bandeira de Mello
American Institutes for Research

August 2013
Commissioned by the NAEP Validity Studies (NVS) Panel

EXCERPTS

George W. Bohrnstedt, Panel Chair
Frances B. Stancavage, Project Director

This report was prepared for the National Center for Education Statistics under Contract No. ED-04-CO-0025/0012 with the American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.
**The NAEP Validity Studies (NVS) Panel** was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

**Panel Members:**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter Behuniak</td>
<td>University of Connecticut</td>
</tr>
<tr>
<td>Gerunda Hughes</td>
<td>Howard University</td>
</tr>
<tr>
<td>George W. Bohrnstedt</td>
<td>American Institutes for Research</td>
</tr>
<tr>
<td>Robert Linn</td>
<td>University of Colorado Boulder</td>
</tr>
<tr>
<td>James R. Chromy</td>
<td>Research Triangle Institute</td>
</tr>
<tr>
<td>Ina V.S. Mullis</td>
<td>Boston College</td>
</tr>
<tr>
<td>Phil Daro</td>
<td>Strategic Education Research Partnership (SERP) Institute</td>
</tr>
<tr>
<td>Scott Norton</td>
<td>Council of Chief State School Officers</td>
</tr>
<tr>
<td>Lizanne DeStefano</td>
<td>University of Illinois</td>
</tr>
<tr>
<td>Gary Phillips</td>
<td>American Institutes for Research</td>
</tr>
<tr>
<td>Richard P. Durán</td>
<td>University of California, Santa Barbara</td>
</tr>
<tr>
<td>Lorrie Shepard</td>
<td>University of Colorado Boulder</td>
</tr>
<tr>
<td>David Grissmer</td>
<td>University of Virginia</td>
</tr>
<tr>
<td>David Thissen</td>
<td>University of North Carolina, Chapel Hill</td>
</tr>
<tr>
<td>Larry Hedges</td>
<td>Northwestern University</td>
</tr>
<tr>
<td>Karen Wixson</td>
<td>University of North Carolina, Greensboro</td>
</tr>
</tbody>
</table>

**Project Director:**

Frances B. Stancavage  
*American Institutes for Research*

**Project Officer:**

Janis Brown  
*National Center for Education Statistics*

**For Information:**

NAEP Validity Studies (NVS)  
American Institutes for Research  
2800 Campus Drive, Suite 200  
San Mateo, CA 94403  
Phone: 650/ 843-8100  
Fax: 650/ 843-8200
Executive Summary

To support an internal evaluation of the impact of changing exclusion rates on reports of statistically significant gains in National Assessment of Educational Progress (NAEP) scores across states, the National Center for Education Statistics (NCES) sponsored research on imputation procedures used to calculate NAEP scores for the excluded students and provided adjusted or full population estimates (FPEs) for the 1996 to 2000 NAEP mathematics gains. The FPE methodology developed by McLaughlin (2005) makes use of information in the student-level NAEP data file, which includes data for students with disabilities (SDs) and English language learners (ELLs) generated from questionnaire responses completed by school staff. In 2009, the task force on FPEs formed by the National Institute of Statistical Sciences and the NAEP-Education Statistics Services Institute (NAEP-ESSI) found that methods used to calculate FPEs were sufficiently sound that there was no identified need for drastic modifications. The task force also recommended that NCES support studies to extend and further validate the methodology for imputing plausible values. The occasion of two special inclusion studies conducted in conjunction with the 2011 NAEP Mathematics Assessment presented just such an opportunity for additional validity research.

Both studies focused on the assessment of otherwise-excluded students by offering accommodations that are not allowed in operational NAEP. One study allowed the use of calculators as an accommodation (in states that permit this accommodation on their state assessments). The other provided students with an inclusion booklet made up of Knowledge and Skill Appropriate (KaSA) blocks that were somewhat easier than standard NAEP blocks. In some states, there were students included in both studies (that is, some students included because of the calculator accommodation and other students included because of the inclusion blocks). In other states, only the inclusion block was offered because the states do not allow a calculator accommodation on their state assessments. After school personnel had finalized their exclusion decisions for the operational assessment, they were asked to reconsider whether excluded students could participate using the calculator or KaSA blocks. If they agreed, these students became participants in the special studies. The data from the special studies were scaled with the data from the operational NAEP assessment, and plausible values were created for the participants in the special studies.

Because these 2011 special inclusion studies yielded a sample of students excluded from operational NAEP for whom both NAEP scaled plausible values and FPEs were available, they provided an opportunity to conduct a validity study of the FPEs. The logic was to compare results from an assessment that included the actual scaled scores for some otherwise excluded students (those who could be included with the special accommodations) with results based on the FPEs.

The total number of operationally excluded students in the 2011 NAEP Mathematics Assessment was 5,049 out of a total sample of 169,452 public school
students (about 3.0 percent). Only 1,197 (23.4 percent of the excluded students) participated in the validity study (891 in the special calculator booklet study and 307 in the inclusion booklet study). This was a much smaller sample size than had been expected. Moreover, the special studies sample differed somewhat from the group of excluded students as a whole in ways that are likely to be related to performance on the assessment. In particular, the students in the special studies sample were rated by school personnel as tending to be among the more able of the excluded student group.

Because of the small sample sizes, the differences between the means of the FPEs and of estimates based on scaled plausible values for the otherwise excluded students (overall and for 14 subgroups) resulted in only one significant difference. However, when 95 percent confidence intervals were constructed to examine for possible bias, the resulting intervals ran from 0 to 10 NAEP points, suggesting that the FPEs may tend to overestimate the actual population parameter. This overestimation is not surprising (and indeed was hypothesized to be the case) because the achievement information on which the FPEs are based is only from assessed students.

It is not clear that FPEs have to be unbiased to be useful, however. Unbiased estimation of unobservable assessment scores is probably an impossible goal in any event. A principled method that leads to smaller bias in estimating a group that is undercovered in a population may be highly desirable. Excluding a population subgroup because it cannot be assessed is roughly equivalent (for estimating population averages) to imputing the mean of the assessed population. The special studies samples investigated here scored, on the average, at about the 10th percentile of the assessed population. If we interpret the difference between the average FPEs and scaled plausible values from the special studies as bias, then the results presented here suggest that the bias in imputing the mean of the assessed population is approximately 10 times as large as that in using the FPEs.

When one considers the possibility of improving NAEP population estimates by expanding the pool of tested students, the study also offers some insights. First, because of the small numbers of students successfully recruited into the special studies (and the characteristics of these students, who tended to be rated by their schools as among the most able of the excluded students), the studies suggest that offering the calculator block and KaSA booklet accommodations, by themselves, would not have a substantial impact on national parameter estimates. However, results for the FPE estimates on the entire excluded population do show nonnegligible impacts on national parameter estimates. This suggests that if accommodations to include more of the currently excluded students could be found, such accommodations could have a nonnegligible impact on national parameter estimates.

Finally, one can question whether the concept of full population estimates is sensible. The reason is that the concept of full population estimates presupposes that there is (at least in theory) an assessment score for every student, including those who are currently excluded from the assessment. If there are students
whom we could not conceive of as participating in the assessment under any
conditions, then the concept of “the assessment score they would have obtained
if they had participated” may not make sense. One might therefore argue that a
group that could never be assessed should be excluded from the definition of the
population used to draw inferences. By redefining the population, efforts could
focus on developing methods to include as many members of the (newly defined)
population as possible in operational assessments and on developing methods to
impute scores for those excluded.
Validity Considerations of the Validity Study

One threat to what might be called the statistical validity of the study is that the sample size may not be large enough to provide adequate statistical power or precision for the estimates compared. To evaluate this, it is important, as a first step, to examine the sample sizes obtained and determine if the estimates for the operationally excluded subgroup are precise enough to draw conclusions. The logical framework of this study is that of an equivalence (not superiority) study. That is, we conclude that FPEs are valid if the estimates based on them do not differ from those based on scaled plausible values for students in the inclusion samples. Consequently, we must set the smallest difference that is meaningful and determine whether the sample size will yield adequate statistical power to detect such differences. As a guideline, we suggest using the convention of 80 percent power at the 5 percent significance level.

A crude precision analysis can be done by computing the standard error of the difference between the estimate of a population parameter (e.g., the mean) based on full population methods \( Y_{FPE} \) and the same estimate based on scaled plausible values \( Y_{SPV} \)

\[
S = \sqrt{S_{FPE}^2 + S_{SPV}^2 - 2S_{FPE}S_{SPV}r},
\]

where \( S_{FPE}^2 \) and \( S_{SPV}^2 \) are the variances of \( Y_{FPE} \) and \( Y_{SPV} \), and \( r \) is an estimate of the correlation between them. A crude estimate of the power to detect a true difference between \( Y_{FPE} \) and \( Y_{INC} \) of size \( \delta \) is

\[
p = 1 - \Phi(c - \delta/S) + \Phi(-c - \delta/S),
\]

where \( c \) is the appropriate critical value of the standard normal distribution and \( \Phi(x) \) is the standard normal cumulative distribution function.

The question of how large the difference \( \delta \) must be to be meaningful is more difficult. We used the approach of studying how large \( \delta \) must be to produce a consequential difference in assessment scores, as in Table 7. Using the standard errors from Table 6, we evaluate the power to detect the smallest bias that would lead to a change in national or subgroup means by 0.5 and 1.0 NAEP scale-score points in Table 10. For the nation and all of the subgroups considered, the power to detect a bias large enough to change the average estimate by 1.0 NAEP scale-score points is essentially 1.0; thus, these studies appear adequately powered to detect biases large enough to produce a change of 1.0 NAEP scale-score point.

The situation is somewhat different with respect to biases large enough to produce a change of 0.5 NAEP scale-score points. In the black and Hispanic subgroups, the power to detect such a change is only about 70 percent. Thus, the special studies cannot be considered definitive in ruling out such biases in the black and Hispanic reporting subgroups. Note also that, although the point estimates of bias for these two subgroups were less than 5 NAEP scale-score points, the upper ends of the
95 percent confidence intervals for the bias estimates in these two groups (17.2 and 20.8, respectively) do exceed the threshold for bias that could cause a 0.5 NAEP scale-score change in national estimates for each of those groups. In other words, the power of these validity studies is not high enough to rule out biases that could change national estimates of the mean in the black and Hispanic reporting subgroups by as much as 0.5 NAEP scale-score points.

Table 10. Power to Detect a Bias in FPEs That Could Produce a Change in Overall Averages of 0.5 or 1.0 Points in Various Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>To Detect Overall Bias of 0.5</th>
<th>To Detect Overall Bias of 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FPE Bias</td>
<td>Power</td>
</tr>
<tr>
<td>Nation</td>
<td>19</td>
<td>1.00</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
<td>0.98</td>
</tr>
<tr>
<td>Female</td>
<td>26</td>
<td>1.00</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>22</td>
<td>1.00</td>
</tr>
<tr>
<td>Black</td>
<td>13</td>
<td>0.71</td>
</tr>
<tr>
<td>Hispanic</td>
<td>18</td>
<td>0.70</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>18</td>
<td>0.95</td>
</tr>
<tr>
<td>Southeast</td>
<td>22</td>
<td>0.99</td>
</tr>
<tr>
<td>Central</td>
<td>18</td>
<td>0.99</td>
</tr>
<tr>
<td>West</td>
<td>19</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Note: These computations assume a two-sided 5 percent nonsimultaneous significance test.

There also are two threats to internal validity that can be characterized as selection threats. If the school personnel making exclusion decisions know that this is part of a special study, then biases might arise because of experimenter demand characteristics (see Orne, 1962) or Hawthorne effects (Mayo, 1949).⁶ We believe that the data collection plan did not explicitly characterize this as part of a special study, which should minimize that validity threat.

The second selection threat is that the school personnel might be motivated to exclude from the operational assessment the students that they believe will perform most poorly. Because they are told that the initially excluded students will not be part of the operational assessment, they have no incentive to exclude students they believe will perform most poorly from the validity study. However, any tendency to exclude students whom they believe will perform most poorly from the operational assessment could produce a bias in the validity study estimates.

---

⁶ Experimenter effects refer to experimental results that are biased as a result of the study participants’ desire to please the researcher. Hawthorne effects are similar. In a classic study of worker productivity at the Western Electric Hawthorne factory, it was shown that the results were less due to the interventions that were put in place than the fact that the workers were being studied, which seemed to increase productivity in and of itself.
assessment could mean that the validity study sample may include students who could have been included in the operational assessment, but who were systematically excluded because they were expected to have poorer performance than those included.

The implications for performance of the sample of operationally excluded students who are in the validity study are unclear because these two factors work in opposite directions. Assuming that, in general, students who could properly participate in the operational assessment will perform better than those who could not, adding these excluded students to the validity study sample might artifactually elevate the performance of the students in the validity study. However, if school personnel are correct that the operationally excluded students perform more poorly than included students, they may also perform more poorly than the properly excluded students, which would artifactually reduce the performance of the students in the validity study sample.

The basic validity question is whether the excluded students who participated in the special studies differ from other excluded students in unobservable (or at least unmeasured) ways that are correlated with achievement (holding constant the observables used in creating the FPEs). The fact that the results in Table 6 suggest that estimates of average achievement based on scaled plausible values are slightly smaller than those based on FPEs suggests that this may be the case.

More elaborate statistical modeling to estimate the expected performance of the excluded students also would be possible. For example, suppose that the excluded students are modeled to be the lower tail (the lowest x% of the distribution, where x is the exclusion rate) of the achievement distribution. We could use the assumption of a distribution shape (e.g., normal) to obtain the expected average (and even standard deviation) of the excluded group. Such an analysis would not, however, resolve whether the poorer performance of excluded students was a consequence of proper exclusion (which is consistent with excluded students having poorer performance) or improper exclusion (excluding students who could have participated but who were excluded because they were expected to have poorer performance).

Conclusions

The special inclusions studied here are disappointing in that they made it possible to include in the assessment only about a quarter of the excluded students and less than 1 percent of the total sample. Moreover, the students they made it possible to include appear to be among the most able of the excluded students—those who were “almost able” to be included without the special accommodations. The cost of these special accommodations seems relatively large for the potential benefit achieved.

In general, it appears that the FPEs may tend to overestimate the results based on scaled plausible values in the special studies, although these differences are far from statistically significant. This is not surprising (and indeed was hypothesized to be the case) because the achievement information on which the FPEs are based is from assessed students. Presumably, there are reasons that students are not assessed, and
not all of these depend on observable (or at least observed) characteristics. Thus, any assessed student whose observed characteristics are equivalent to a student who is not assessed differs on some characteristics that are not observed. If this is so, and if these unobserved characteristics are correlated with (also unobserved) assessment scores, then the FPEs would be biased estimates of the assessment scores. More specifically, it is plausible that the unobserved information leading to exclusion is negatively related to assessment scores. If so, then FPEs would overestimate the performance of excluded students.

It is not clear that FPEs have to be unbiased to be useful, however. Unbiased estimation of unobservable assessment scores is probably an impossible goal in any event. A principled method that leads to smaller bias in estimating a group that is undercovered in a population may be highly desirable. Excluding a population subgroup because it cannot be assessed is roughly equivalent (for estimating population averages) to imputing the mean of the assessed population. The special studies sample investigated here scored, on the average, at about the 10th percentile of the assessed population. If we interpret the difference between the average FPEs and scaled plausible values from the special studies as bias, then the results presented here suggest that the bias in imputing the mean of the assessed population is approximately 10 times as large as that in using the FPEs.

The composition of the special studies sample appears to include more able students than the average of the excluded student population. If this is true, then the difference between the (unobserved) ability of the entire excluded population and the FPEs (the bias in the FPEs) could be larger for the entire exclusion population than for the special studies sample. Although the special studies provide no empirical evidence about the size of that bias, it is difficult to imagine that it could be larger than the bias implied by imputing the mean of the assessed population for these values.

These studies suggest that the calculator block and KaSA booklet accommodations, by themselves, will not change the number of included students enough to have a substantial impact on national parameter estimates. However, results for the FPE estimates on the entire excluded population do show nonnegligible impacts on national parameter estimates. This suggests that if accommodations to include more of the currently excluded students could be found, such accommodations could have a nonnegligible impact on national parameter estimates. Moreover, because FPEs appear to overestimate estimates based on scaled plausible values, the impact of including currently excluded students would likely be even larger than that estimated by the FPEs.

It is important to remember that these special studies are relatively small, and consequently their results have considerable sampling uncertainty that makes it difficult to draw sharp conclusions. The sampling uncertainty made it infeasible to carry out many analyses that would have been desirable. For example, it would be useful to see if patterns of bias were reasonably constant across states and across all reporting groups, but it was not meaningful to conduct these analyses. A fair conclusion is that the sampling uncertainty is so large that any conclusions drawn from this study must be done with extreme care.
It may be useful to question whether the concept of full population estimates is sensible. The reason is that the concept of full population estimates presupposes that there is (at least in theory) an assessment score for every student, including those who are currently excluded from the assessment. If there are students whom we could not conceive as participating in the assessment under any conditions, then the concept of “the assessment score they would have obtained if they had participated” may not make sense. Moreover, it is impossible that any empirical methods could be developed to impute assessment scores for a group that could never have assessment scores—no empirical information about assessment scores could exist for that group. Consequently, it will never be possible to validate methods of imputing assessment scores for a group that could never be assessed. One might therefore argue that a group that could never be assessed should be excluded from the definition of the population used to draw inferences. By redefining the population, efforts could focus on developing methods to include as many members of the (newly defined) population as possible in operational assessments and developing methods to impute scores for those excluded. Of course, there is a problem in identifying the group that should be defined as not (ever) assessable. Nevertheless, it may be worth attempting to develop at least provisional definitions of such a group.

This suggests a concept of expanded population estimates (rather than full population estimates) that corresponds to estimating the assessment scores that could be obtained by all students who could participate in the assessment under conditions of special accommodations. One virtue of this definition is that every student in the inference population could be assessed under some accommodations (including accommodations that might be infeasible under operational conditions because of time or cost constraints). Because it would be possible to obtain assessment scores for every student in the population, empirical methods could, in principle, be used to develop imputations for any students in this population who are excluded from the operational assessment (perhaps in special studies involving extensive accommodations). Moreover, it would be possible to empirically validate such methods.
Appendix A. Procedures for Calculating Full Population Estimates

McLaughlin (2000) introduced a method to estimate the achievement of the subset of the students with disabilities (SDs) and English language learners (ELLs) excluded by NAEP. The method relies on the NAEP SD and ELL questionnaires, descriptive surveys that are filled out by a teacher or knowledgeable staff person for each student with a disability and each English language learner selected to participate in NAEP—whether or not these students actually participate in NAEP or are excluded on the grounds that NAEP testing would be inappropriate for them.

The basic assumption is that excluded students in a given state with a particular profile based on student and school demographic characteristics and information from the SD and ELL questionnaires will, on average, be at the same achievement level as students with disabilities and English language learners in that state who participated in NAEP and had the same profile of demographic characteristics and information on the SD and ELL questionnaires. McLaughlin called this the profile matching method. Since the scores resulting from this procedure provide estimates that now include all of a state’s SDs and ELLs, they are called **full population estimates** (FPEs).

No student takes the entire NAEP assessment. Instead, a student takes a random sample of blocks of items drawn from the entire item set for a given assessment. The items each student takes are used to compute five sets of what are called **plausible values**. These are then used to compute estimates of performance for the entire population of students as well as congressionally mandated subgroups of students (e.g., males and females).

In computing the FPEs, plausible values for the composite NAEP scale in each grade and subject are computed first for all excluded ELLs in the NAEP public school sample, and second, separately, for all excluded SDs in the sample who are not also ELLs. Data for students who are neither ELL nor SD are not used in the process. The plausible values are constructed in three steps.

1. **Predictor preparation.** Predictive demographic information and questionnaire responses, which are available for both included and excluded ELLs and SDs, are extracted from the NAEP file and recoded to maximize predictive power. Stepwise regression is used to remove predictors possessing no significant power in predicting plausible values for included ELLs (or SDs) and to remove predictors that are too highly correlated with other predictors.

2. **Estimation of the mean expected score for each excluded student.** A single pooled within-state linear regression is carried out to estimate the coefficient for each of the predictors created in step 1 in predicting the scores of included ELLs.

---

7 The procedures in this paper used five plausible values, but the estimation procedure has been changed for the 2013 NAEP assessments and now generates 20 plausible values. Future versions of the software for generating FPEs will be updated to reflect this change.
The regression intercept is adjusted separately for each state so that the mean predicted score for included ELLs (or SDs) matches their observed mean in each state. The resulting coefficients are used to impute an estimate for each excluded ELL (or SD).

3. **Estimation of imputation error variance and generation of five random plausible values for each excluded student.** Five plausible values are generated for each excluded student by adding to the estimate obtained in step 2 random normal deviates with three components of variance: (1) average variation among the five NAEP plausible values for included ELLs (or SDs), (2) average regression error due to the imperfect linear regression prediction in step 2, and (3) sampling error introduced in matching the included ELL (or SD) mean in the state.

One of the difficulties that the FPE procedure has had to deal with is that the set of questions that comprise the NAEP SD and ELL questionnaires have changed from year to year. As a result, the prediction equations change from NAEP administration to administration. While this fact does not diminish the utility of the FPE procedure, it does mean that the fit of regression results to the data can vary over time. Table A-1 below lists separately the variables used in the NAEP 2011 Grade 8 reading and mathematics FPE regressions for ELLs and SDs.

---

8 A student’s “score” is defined as the mean of the five plausible values for that student.
Table A-1. Variables Used in the Linear Regressions for Grade 8 Reading and Mathematics: 2011

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mathematics Grade 8</th>
<th>Reading Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>ELL</td>
</tr>
<tr>
<td><strong>Items from the ELL questionnaire</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XL04501</td>
<td>What is this student’s ELL classification?</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03801</td>
<td>How is student included in state assessment?</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03901</td>
<td>Extended time (allowed for all subjects)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03902</td>
<td>Small group (allowed for all subjects)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03908</td>
<td>Test items read aloud in English</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03905</td>
<td>Breaks during testing (allowed for all subjects)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03909</td>
<td>Must have an aide administer test</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03910</td>
<td>Cueing to stay on task</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03906</td>
<td>Bilingual dictionary w/out definitions in any language</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03911</td>
<td>Read directions aloud in Spanish</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03912</td>
<td>Test items read aloud in Spanish (math &amp; science)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03913</td>
<td>Spanish/English version of the test (math &amp; science)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL03914</td>
<td>Student receives other accommodations</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04001</td>
<td>How should this student be included on NAEP test?</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04002</td>
<td>If student ineligible for NAEP, record admin. code</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04101</td>
<td>How long has student been receiving instruction in English?</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04201</td>
<td>Grade level of performance in NAEP subject</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04301</td>
<td>Student’s English proficiency: listening comprehension in English</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04302</td>
<td>Student’s English proficiency: Speaking English</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04303</td>
<td>Student’s English proficiency: Reading English</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XL04304</td>
<td>Student’s English proficiency: Writing English</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td><strong>Items from the SD questionnaire</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS04701</td>
<td>Why is this student classified as SD?</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04801</td>
<td>How is student included in state assessment?</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04901</td>
<td>Extended time (allowed for all subjects)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04902</td>
<td>Small group (allowed for all subjects)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04907</td>
<td>Test items read aloud in English</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04905</td>
<td>Breaks during testing (allowed for all subjects)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04908</td>
<td>Must have an aide administer test</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04909</td>
<td>Responds orally to a scribe</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04910</td>
<td>Large-print version of the test</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04911</td>
<td>Magnification equipment</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04912</td>
<td>Uses a calculator for all sections (math only)</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS04913</td>
<td>Uses template/special equip./preferential seating</td>
<td>•</td>
<td></td>
</tr>
</tbody>
</table>
## Variable Description

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mathematics Grade 8</th>
<th>Reading Grade 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>XS04914</td>
<td>Cueing to stay on task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS04915</td>
<td>Presentation or response in braille</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS04916</td>
<td>Presentation or response in sign language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS04917</td>
<td>Student receives other accommodations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS05001</td>
<td>How should this student be included on NAEP test?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS05002</td>
<td>If student ineligible for NAEP, record admin. code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS050101</td>
<td>Student’s identified disability: Specific learning</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>XS050102</td>
<td>Student’s identified disability: Hearing impairment</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS050103</td>
<td>Student’s identified disability: Visual impairment</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS050105</td>
<td>Student’s identified disability: Mental retardation</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>XS050106</td>
<td>Student’s identified disability: Emotional disturbance</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS050107</td>
<td>Student’s identified disability: Orthopedic impairment</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>XS050108</td>
<td>Student’s identified disability: Brain injury</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS050109</td>
<td>Student’s identified disability: Autism</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>XS050110</td>
<td>Student’s identified disability: Developmental delay</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS050111</td>
<td>Student’s identified disability: Other health</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>XS050104</td>
<td>Student’s identified disability: Speech impairment</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>XS05012</td>
<td>Student’s identified disability: Other-write-in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS05201</td>
<td>Degree of student’s disability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS05301</td>
<td>Grade level student performs in the NAEP subject</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>IEP</td>
<td>Student classified as having a disability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMIN</td>
<td>Student is not white</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>DSEX</td>
<td>Student gender</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>SLUNCH</td>
<td>National School Lunch Program eligibility</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>PCTBLK</td>
<td>School-level percentage of black students</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>PCTIND</td>
<td>School-level percentage of American Indian students</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>PCTHISP</td>
<td>School-level percentage of Hispanic students</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>READVAR</td>
<td>School-level state test scores—Reading</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>MATHVAR</td>
<td>School-level state test scores—Math</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>SENROL8</td>
<td>School enrollment</td>
<td>•</td>
<td></td>
</tr>
</tbody>
</table>

### Student and school characteristics

Attachment A8 - NAEP 2013 Operations Report: Students Assessed, Accommodated, Excluded, and Absent, including total participation rates (available in both PDF and Excel formats)
### NAEP 2013 Operations Report Table 4

#### Number of Students Assessed, Accommodated, Excluded, and Absent

<table>
<thead>
<tr>
<th></th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Total Participation Rate</th>
<th>Not Assessed</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
<td>Accommodated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>831,930</td>
<td>792,560</td>
<td>631,303</td>
<td>97,349</td>
<td>16,082</td>
<td>2.0%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>214,934</td>
<td>205,142</td>
<td>162,475</td>
<td>28,679</td>
<td>3,116</td>
<td>1.5%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>216,426</td>
<td>206,683</td>
<td>164,835</td>
<td>25,752</td>
<td>5,596</td>
<td>2.7%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>201,461</td>
<td>191,407</td>
<td>152,232</td>
<td>23,047</td>
<td>2,912</td>
<td>1.5%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>199,109</td>
<td>189,328</td>
<td>151,761</td>
<td>20,051</td>
<td>4,458</td>
<td>2.4%</td>
</tr>
<tr>
<td>Total TUDA</td>
<td>149,257</td>
<td>141,849</td>
<td>105,663</td>
<td>23,255</td>
<td>4,139</td>
<td>2.9%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>38,864</td>
<td>37,044</td>
<td>27,356</td>
<td>6,969</td>
<td>767</td>
<td>2.1%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>40,094</td>
<td>38,178</td>
<td>28,349</td>
<td>6,259</td>
<td>1,571</td>
<td>4.1%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>34,758</td>
<td>32,942</td>
<td>24,684</td>
<td>5,124</td>
<td>656</td>
<td>2.0%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>35,541</td>
<td>33,685</td>
<td>25,274</td>
<td>4,903</td>
<td>1,145</td>
<td>3.4%</td>
</tr>
<tr>
<td>Total Public</td>
<td>818,369</td>
<td>779,146</td>
<td>619,153</td>
<td>96,739</td>
<td>16,047</td>
<td>2.1%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>211,657</td>
<td>201,896</td>
<td>159,511</td>
<td>28,542</td>
<td>3,114</td>
<td>1.5%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>213,046</td>
<td>203,340</td>
<td>161,770</td>
<td>25,454</td>
<td>4,458</td>
<td>2.7%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>198,049</td>
<td>188,035</td>
<td>149,217</td>
<td>22,869</td>
<td>2,904</td>
<td>1.5%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>195,617</td>
<td>185,875</td>
<td>148,655</td>
<td>19,874</td>
<td>4,448</td>
<td>2.4%</td>
</tr>
<tr>
<td>Total Nonpublic</td>
<td>13,561</td>
<td>13,414</td>
<td>12,150</td>
<td>610</td>
<td>35</td>
<td>0.3%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>3,277</td>
<td>3,246</td>
<td>2,964</td>
<td>137</td>
<td>2</td>
<td>0.1%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>3,380</td>
<td>3,343</td>
<td>3,065</td>
<td>118</td>
<td>15</td>
<td>0.4%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>3,412</td>
<td>3,372</td>
<td>3,015</td>
<td>178</td>
<td>8</td>
<td>0.2%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>3,492</td>
<td>3,453</td>
<td>3,106</td>
<td>177</td>
<td>10</td>
<td>0.3%</td>
</tr>
<tr>
<td>Alabama</td>
<td>12,683</td>
<td>11,964</td>
<td>10,630</td>
<td>578</td>
<td>131</td>
<td>1.1%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>3,219</td>
<td>3,050</td>
<td>2,724</td>
<td>140</td>
<td>32</td>
<td>1.0%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>3,351</td>
<td>3,175</td>
<td>2,828</td>
<td>169</td>
<td>36</td>
<td>1.1%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>3,035</td>
<td>2,838</td>
<td>2,512</td>
<td>135</td>
<td>29</td>
<td>1.0%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>3,078</td>
<td>2,901</td>
<td>2,566</td>
<td>134</td>
<td>34</td>
<td>1.2%</td>
</tr>
<tr>
<td>Alabama</td>
<td>12,478</td>
<td>11,694</td>
<td>10,630</td>
<td>578</td>
<td>131</td>
<td>1.1%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>3,131</td>
<td>2,946</td>
<td>2,173</td>
<td>535</td>
<td>34</td>
<td>1.2%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>3,257</td>
<td>3,061</td>
<td>2,259</td>
<td>573</td>
<td>44</td>
<td>1.4%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>3,002</td>
<td>2,821</td>
<td>2,095</td>
<td>470</td>
<td>34</td>
<td>1.2%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>3,088</td>
<td>2,866</td>
<td>2,189</td>
<td>418</td>
<td>38</td>
<td>1.3%</td>
</tr>
<tr>
<td>Arizona</td>
<td>13,394</td>
<td>12,559</td>
<td>10,359</td>
<td>1,358</td>
<td>162</td>
<td>1.3%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>3,427</td>
<td>3,228</td>
<td>2,650</td>
<td>388</td>
<td>38</td>
<td>1.2%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>3,531</td>
<td>3,338</td>
<td>2,738</td>
<td>414</td>
<td>38</td>
<td>1.1%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>3,180</td>
<td>2,957</td>
<td>2,441</td>
<td>282</td>
<td>41</td>
<td>1.4%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>3,256</td>
<td>3,036</td>
<td>2,530</td>
<td>274</td>
<td>45</td>
<td>1.5%</td>
</tr>
<tr>
<td>Arkansas</td>
<td>13,327</td>
<td>12,475</td>
<td>9,903</td>
<td>1,717</td>
<td>192</td>
<td>1.5%</td>
</tr>
<tr>
<td>Grade 4 Mathematics</td>
<td>3,423</td>
<td>3,223</td>
<td>2,554</td>
<td>464</td>
<td>39</td>
<td>1.2%</td>
</tr>
<tr>
<td>Grade 4 Reading</td>
<td>3,560</td>
<td>3,348</td>
<td>2,646</td>
<td>507</td>
<td>36</td>
<td>1.1%</td>
</tr>
<tr>
<td>Grade 8 Mathematics</td>
<td>3,161</td>
<td>2,939</td>
<td>2,350</td>
<td>388</td>
<td>58</td>
<td>2.0%</td>
</tr>
<tr>
<td>Grade 8 Reading</td>
<td>3,183</td>
<td>2,965</td>
<td>2,353</td>
<td>358</td>
<td>59</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
## NAEP 2013 Operations Report Table 4

### Number of Students Assessed, Accommodated, Excluded, and Absent

<table>
<thead>
<tr>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Total Participation Rate</th>
<th>Not Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standard</td>
<td>Accommodated</td>
<td>Excluded</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State</th>
<th>Grade</th>
<th>Subject</th>
<th>Original Student Sample</th>
<th>Assessed</th>
<th>%Assessed</th>
<th>Total Participaiton Rate</th>
<th>Excluded</th>
<th>%Excluded</th>
<th>Parent Refusal</th>
<th>Other Absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>35,116</td>
<td>33,587</td>
<td>92.3%</td>
<td>703</td>
<td>2.1%</td>
<td>173</td>
<td>1,721</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>Grade 4</td>
<td>Reading</td>
<td>9,000</td>
<td>8,598</td>
<td>93.4%</td>
<td>135</td>
<td>1.6%</td>
<td>32</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>8,353</td>
<td>7,981</td>
<td>92.5%</td>
<td>661</td>
<td>1.8%</td>
<td>58</td>
<td>453</td>
<td></td>
</tr>
<tr>
<td>California</td>
<td>Grade 8</td>
<td>Reading</td>
<td>8,505</td>
<td>8,152</td>
<td>91.2%</td>
<td>611</td>
<td>2.7%</td>
<td>46</td>
<td>451</td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>13,221</td>
<td>12,658</td>
<td>92.1%</td>
<td>157</td>
<td>1.2%</td>
<td>112</td>
<td>732</td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,400</td>
<td>3,276</td>
<td>91.3%</td>
<td>365</td>
<td>1.1%</td>
<td>26</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>Colorado</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,126</td>
<td>2,968</td>
<td>92.4%</td>
<td>661</td>
<td>1.8%</td>
<td>40</td>
<td>363</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>12,799</td>
<td>12,455</td>
<td>91.4%</td>
<td>222</td>
<td>1.8%</td>
<td>181</td>
<td>670</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,234</td>
<td>3,147</td>
<td>92.7%</td>
<td>44</td>
<td>1.4%</td>
<td>35</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,091</td>
<td>3,043</td>
<td>90.4%</td>
<td>64</td>
<td>2.1%</td>
<td>63</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>Connecticut</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,190</td>
<td>3,140</td>
<td>92.5%</td>
<td>34</td>
<td>1.1%</td>
<td>36</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>13,362</td>
<td>12,992</td>
<td>90.2%</td>
<td>381</td>
<td>2.9%</td>
<td>53</td>
<td>844</td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,419</td>
<td>3,333</td>
<td>92.3%</td>
<td>44</td>
<td>1.4%</td>
<td>14</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,172</td>
<td>3,080</td>
<td>92.4%</td>
<td>43</td>
<td>1.3%</td>
<td>36</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>Delaware</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,247</td>
<td>3,147</td>
<td>88.4%</td>
<td>110</td>
<td>3.5%</td>
<td>12</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>8,811</td>
<td>8,485</td>
<td>91.8%</td>
<td>123</td>
<td>1.4%</td>
<td>15</td>
<td>559</td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>Grade 4</td>
<td>Reading</td>
<td>2,279</td>
<td>2,209</td>
<td>93.8%</td>
<td>31</td>
<td>1.4%</td>
<td>5</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>District of Columbia</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>2,065</td>
<td>1,986</td>
<td>90.4%</td>
<td>19</td>
<td>1.0%</td>
<td>3</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>20,344</td>
<td>19,088</td>
<td>92.1%</td>
<td>418</td>
<td>2.2%</td>
<td>122</td>
<td>970</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 4</td>
<td>Reading</td>
<td>6,919</td>
<td>6,556</td>
<td>93.0%</td>
<td>127</td>
<td>1.9%</td>
<td>67</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>6,432</td>
<td>6,073</td>
<td>90.4%</td>
<td>111</td>
<td>1.8%</td>
<td>91</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 8</td>
<td>Reading</td>
<td>6,542</td>
<td>6,179</td>
<td>90.5%</td>
<td>136</td>
<td>2.2%</td>
<td>69</td>
<td>384</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>26,966</td>
<td>25,547</td>
<td>91.4%</td>
<td>576</td>
<td>2.3%</td>
<td>286</td>
<td>1,324</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>Grade 4</td>
<td>Reading</td>
<td>6,191</td>
<td>5,656</td>
<td>93.0%</td>
<td>127</td>
<td>1.9%</td>
<td>67</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>6,582</td>
<td>6,179</td>
<td>90.5%</td>
<td>136</td>
<td>2.2%</td>
<td>69</td>
<td>384</td>
<td></td>
</tr>
<tr>
<td>Florida</td>
<td>Grade 8</td>
<td>Reading</td>
<td>6,542</td>
<td>6,179</td>
<td>90.5%</td>
<td>136</td>
<td>2.2%</td>
<td>69</td>
<td>384</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>3,679</td>
<td>3,264</td>
<td>93.5%</td>
<td>54</td>
<td>1.7%</td>
<td>10</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,777</td>
<td>3,383</td>
<td>89.8%</td>
<td>200</td>
<td>5.9%</td>
<td>13</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>2,623</td>
<td>2,397</td>
<td>93.4%</td>
<td>27</td>
<td>1.1%</td>
<td>9</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td>Grade 8</td>
<td>Reading</td>
<td>2,629</td>
<td>2,391</td>
<td>90.5%</td>
<td>93</td>
<td>3.9%</td>
<td>4</td>
<td>131</td>
<td></td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
<table>
<thead>
<tr>
<th>State</th>
<th>Grade</th>
<th>Subject</th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Excluded</th>
<th>% Excluded</th>
<th>Other Reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawaii</td>
<td>4</td>
<td>Mathematics</td>
<td>13,546</td>
<td>12,894</td>
<td>12,894</td>
<td>1,358</td>
<td>91.2%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Hawaii</td>
<td>4</td>
<td>Reading</td>
<td>3,482</td>
<td>3,301</td>
<td>3,301</td>
<td>1,358</td>
<td>93.5%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Hawaii</td>
<td>8</td>
<td>Mathematics</td>
<td>3,593</td>
<td>3,401</td>
<td>3,401</td>
<td>1,358</td>
<td>92.3%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Hawaii</td>
<td>8</td>
<td>Reading</td>
<td>3,212</td>
<td>3,078</td>
<td>3,078</td>
<td>1,358</td>
<td>89.1%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Hawaii</td>
<td>8</td>
<td>Reading</td>
<td>3,259</td>
<td>3,114</td>
<td>3,114</td>
<td>1,358</td>
<td>89.4%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Idaho</td>
<td>4</td>
<td>Mathematics</td>
<td>13,412</td>
<td>12,552</td>
<td>12,552</td>
<td>1,358</td>
<td>93.2%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Idaho</td>
<td>4</td>
<td>Reading</td>
<td>3,479</td>
<td>3,269</td>
<td>3,269</td>
<td>1,358</td>
<td>93.9%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Idaho</td>
<td>8</td>
<td>Mathematics</td>
<td>3,576</td>
<td>3,377</td>
<td>3,377</td>
<td>1,358</td>
<td>93.5%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Idaho</td>
<td>8</td>
<td>Reading</td>
<td>3,147</td>
<td>2,922</td>
<td>2,922</td>
<td>1,358</td>
<td>93.3%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Idaho</td>
<td>8</td>
<td>Reading</td>
<td>3,210</td>
<td>2,984</td>
<td>2,984</td>
<td>1,358</td>
<td>92.2%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Illinois</td>
<td>4</td>
<td>Mathematics</td>
<td>20,005</td>
<td>19,387</td>
<td>19,387</td>
<td>1,358</td>
<td>93.4%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Illinois</td>
<td>4</td>
<td>Reading</td>
<td>5,081</td>
<td>4,910</td>
<td>4,910</td>
<td>1,358</td>
<td>93.6%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Illinois</td>
<td>8</td>
<td>Mathematics</td>
<td>4,780</td>
<td>4,636</td>
<td>4,636</td>
<td>1,358</td>
<td>93.5%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Illinois</td>
<td>8</td>
<td>Reading</td>
<td>4,896</td>
<td>4,761</td>
<td>4,761</td>
<td>1,358</td>
<td>92.7%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Indiana</td>
<td>4</td>
<td>Mathematics</td>
<td>12,849</td>
<td>12,243</td>
<td>12,243</td>
<td>1,358</td>
<td>92.8%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Indiana</td>
<td>4</td>
<td>Reading</td>
<td>3,340</td>
<td>3,202</td>
<td>3,202</td>
<td>1,358</td>
<td>93.8%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Indiana</td>
<td>8</td>
<td>Mathematics</td>
<td>3,456</td>
<td>3,306</td>
<td>3,306</td>
<td>1,358</td>
<td>92.0%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Indiana</td>
<td>8</td>
<td>Reading</td>
<td>2,994</td>
<td>2,830</td>
<td>2,830</td>
<td>1,358</td>
<td>91.1%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Indiana</td>
<td>8</td>
<td>Reading</td>
<td>3,059</td>
<td>2,905</td>
<td>2,905</td>
<td>1,358</td>
<td>91.3%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Iowa</td>
<td>4</td>
<td>Mathematics</td>
<td>12,464</td>
<td>11,998</td>
<td>11,998</td>
<td>1,358</td>
<td>92.1%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Iowa</td>
<td>4</td>
<td>Reading</td>
<td>3,092</td>
<td>2,989</td>
<td>2,989</td>
<td>1,358</td>
<td>94.4%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Iowa</td>
<td>8</td>
<td>Mathematics</td>
<td>3,064</td>
<td>2,942</td>
<td>2,942</td>
<td>1,358</td>
<td>93.0%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Iowa</td>
<td>8</td>
<td>Reading</td>
<td>3,116</td>
<td>2,975</td>
<td>2,975</td>
<td>1,358</td>
<td>92.2%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kansas</td>
<td>4</td>
<td>Mathematics</td>
<td>13,582</td>
<td>12,971</td>
<td>12,971</td>
<td>1,358</td>
<td>92.7%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kansas</td>
<td>4</td>
<td>Reading</td>
<td>3,439</td>
<td>3,283</td>
<td>3,283</td>
<td>1,358</td>
<td>93.4%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kansas</td>
<td>8</td>
<td>Mathematics</td>
<td>3,541</td>
<td>3,395</td>
<td>3,395</td>
<td>1,358</td>
<td>93.2%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kansas</td>
<td>8</td>
<td>Reading</td>
<td>3,269</td>
<td>3,121</td>
<td>3,121</td>
<td>1,358</td>
<td>92.5%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kentucky</td>
<td>4</td>
<td>Mathematics</td>
<td>18,070</td>
<td>17,228</td>
<td>17,228</td>
<td>1,358</td>
<td>92.0%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kentucky</td>
<td>4</td>
<td>Reading</td>
<td>4,653</td>
<td>4,450</td>
<td>4,450</td>
<td>1,358</td>
<td>93.3%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kentucky</td>
<td>8</td>
<td>Mathematics</td>
<td>4,798</td>
<td>4,584</td>
<td>4,584</td>
<td>1,358</td>
<td>91.7%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Kentucky</td>
<td>8</td>
<td>Reading</td>
<td>4,285</td>
<td>4,076</td>
<td>4,076</td>
<td>1,358</td>
<td>92.3%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Louisiana</td>
<td>4</td>
<td>Mathematics</td>
<td>13,274</td>
<td>12,189</td>
<td>12,189</td>
<td>1,358</td>
<td>93.2%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Louisiana</td>
<td>4</td>
<td>Reading</td>
<td>3,331</td>
<td>3,117</td>
<td>3,117</td>
<td>1,358</td>
<td>93.5%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Louisiana</td>
<td>8</td>
<td>Mathematics</td>
<td>3,442</td>
<td>3,192</td>
<td>3,192</td>
<td>1,358</td>
<td>93.6%</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Louisiana</td>
<td>8</td>
<td>Reading</td>
<td>3,237</td>
<td>2,924</td>
<td>2,924</td>
<td>1,358</td>
<td>92.9%</td>
<td>Parent Refusal</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
NAEP 2013 Operations Report Table 4
Number of Students Assessed, Accommodated, Excluded, and Absent

<table>
<thead>
<tr>
<th>State</th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Total Participation Rate</th>
<th>Not Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
<td>Accommodated</td>
<td>%Excluded</td>
<td>Parent Refusal</td>
<td>Other Absence</td>
</tr>
<tr>
<td>Maine</td>
<td>12,742</td>
<td>12,276</td>
<td>9,413</td>
<td>505</td>
<td>206</td>
</tr>
<tr>
<td>Grade 4</td>
<td>3,376</td>
<td>3,279</td>
<td>2,502</td>
<td>2,502</td>
<td>91.7%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>3,474</td>
<td>3,370</td>
<td>2,548</td>
<td>2,548</td>
<td>92.0%</td>
</tr>
<tr>
<td>Reading</td>
<td>2,918</td>
<td>2,790</td>
<td>2,164</td>
<td>2,164</td>
<td>91.3%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>2,974</td>
<td>2,837</td>
<td>2,199</td>
<td>2,199</td>
<td>90.9%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>3,212</td>
<td>3,068</td>
<td>2,540</td>
<td>2,540</td>
<td>90.8%</td>
</tr>
<tr>
<td>Reading</td>
<td>18,436</td>
<td>17,699</td>
<td>13,666</td>
<td>1,749</td>
<td>86.5%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>4,748</td>
<td>4,545</td>
<td>3,461</td>
<td>773</td>
<td>93.2%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>4,754</td>
<td>4,702</td>
<td>3,957</td>
<td>231</td>
<td>81.4%</td>
</tr>
<tr>
<td>Reading</td>
<td>4,544</td>
<td>4,686</td>
<td>3,579</td>
<td>237</td>
<td>89.6%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>4,581</td>
<td>4,624</td>
<td>3,597</td>
<td>239</td>
<td>89.9%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>20,140</td>
<td>19,612</td>
<td>14,747</td>
<td>2,996</td>
<td>90.5%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>5,155</td>
<td>5,012</td>
<td>3,710</td>
<td>862</td>
<td>91.2%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>5,318</td>
<td>5,171</td>
<td>3,978</td>
<td>721</td>
<td>90.0%</td>
</tr>
<tr>
<td>Reading</td>
<td>4,766</td>
<td>4,665</td>
<td>3,462</td>
<td>730</td>
<td>89.9%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>4,681</td>
<td>4,745</td>
<td>3,597</td>
<td>683</td>
<td>89.8%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>17,938</td>
<td>16,818</td>
<td>13,377</td>
<td>1,703</td>
<td>89.7%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>4,620</td>
<td>4,340</td>
<td>3,455</td>
<td>474</td>
<td>90.5%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>4,753</td>
<td>4,450</td>
<td>3,573</td>
<td>412</td>
<td>89.6%</td>
</tr>
<tr>
<td>Reading</td>
<td>4,578</td>
<td>4,492</td>
<td>3,557</td>
<td>371</td>
<td>89.9%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>4,517</td>
<td>4,342</td>
<td>3,216</td>
<td>403</td>
<td>88.8%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>13,038</td>
<td>12,452</td>
<td>10,297</td>
<td>1,091</td>
<td>91.5%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>3,491</td>
<td>3,324</td>
<td>2,780</td>
<td>333</td>
<td>93.7%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>3,609</td>
<td>3,455</td>
<td>2,867</td>
<td>323</td>
<td>92.3%</td>
</tr>
<tr>
<td>Reading</td>
<td>5,299</td>
<td>5,070</td>
<td>3,866</td>
<td>153</td>
<td>88.7%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>2,999</td>
<td>2,867</td>
<td>2,160</td>
<td>203</td>
<td>90.4%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>13,183</td>
<td>12,512</td>
<td>10,980</td>
<td>768</td>
<td>93.9%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>3,336</td>
<td>3,193</td>
<td>2,827</td>
<td>198</td>
<td>94.7%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>3,447</td>
<td>3,297</td>
<td>2,905</td>
<td>210</td>
<td>94.5%</td>
</tr>
<tr>
<td>Reading</td>
<td>3,177</td>
<td>2,982</td>
<td>2,596</td>
<td>180</td>
<td>93.1%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>3,223</td>
<td>3,040</td>
<td>2,652</td>
<td>180</td>
<td>93.2%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>13,498</td>
<td>12,463</td>
<td>10,382</td>
<td>1,247</td>
<td>93.3%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>3,612</td>
<td>3,294</td>
<td>2,775</td>
<td>329</td>
<td>94.2%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>3,712</td>
<td>3,401</td>
<td>2,833</td>
<td>367</td>
<td>94.1%</td>
</tr>
<tr>
<td>Reading</td>
<td>3,063</td>
<td>2,860</td>
<td>2,377</td>
<td>283</td>
<td>93.0%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>3,111</td>
<td>2,908</td>
<td>2,397</td>
<td>268</td>
<td>91.6%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>13,202</td>
<td>12,666</td>
<td>10,608</td>
<td>903</td>
<td>90.9%</td>
</tr>
<tr>
<td>Grade 4</td>
<td>3,350</td>
<td>3,216</td>
<td>2,709</td>
<td>246</td>
<td>91.9%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>3,470</td>
<td>3,354</td>
<td>2,844</td>
<td>222</td>
<td>91.4%</td>
</tr>
<tr>
<td>Reading</td>
<td>3,170</td>
<td>3,028</td>
<td>2,515</td>
<td>234</td>
<td>90.8%</td>
</tr>
<tr>
<td>Grade 8</td>
<td>3,212</td>
<td>3,068</td>
<td>2,540</td>
<td>201</td>
<td>89.3%</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
<table>
<thead>
<tr>
<th>State</th>
<th>Grade</th>
<th>Subject</th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed Standard</th>
<th>Assessed Accommodated</th>
<th>Total Participation Rate</th>
<th>Excluded</th>
<th>%Excluded</th>
<th>Parent Refusal</th>
<th>Other Absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nebraska</td>
<td>4</td>
<td>Mathematics</td>
<td>13,283</td>
<td>12,763</td>
<td>10,239</td>
<td>1,492</td>
<td>91.9%</td>
<td>326</td>
<td>2.6%</td>
<td>54</td>
<td>652</td>
</tr>
<tr>
<td>Nebraska</td>
<td>4</td>
<td>Reading</td>
<td>3,468</td>
<td>3,327</td>
<td>2,680</td>
<td>442</td>
<td>93.8%</td>
<td>56</td>
<td>1.7%</td>
<td>10</td>
<td>139</td>
</tr>
<tr>
<td>Nebraska</td>
<td>8</td>
<td>Mathematics</td>
<td>3,091</td>
<td>2,965</td>
<td>2,395</td>
<td>323</td>
<td>91.7%</td>
<td>55</td>
<td>1.9%</td>
<td>17</td>
<td>175</td>
</tr>
<tr>
<td>Nebraska</td>
<td>8</td>
<td>Reading</td>
<td>3,160</td>
<td>3,045</td>
<td>2,452</td>
<td>275</td>
<td>89.6%</td>
<td>92</td>
<td>3.0%</td>
<td>17</td>
<td>209</td>
</tr>
<tr>
<td>Nevada</td>
<td>4</td>
<td>Mathematics</td>
<td>13,950</td>
<td>13,042</td>
<td>9,987</td>
<td>2,127</td>
<td>92.9%</td>
<td>158</td>
<td>1.2%</td>
<td>93</td>
<td>677</td>
</tr>
<tr>
<td>Nevada</td>
<td>4</td>
<td>Reading</td>
<td>3,541</td>
<td>3,306</td>
<td>2,376</td>
<td>745</td>
<td>94.4%</td>
<td>47</td>
<td>1.4%</td>
<td>13</td>
<td>125</td>
</tr>
<tr>
<td>Nevada</td>
<td>8</td>
<td>Mathematics</td>
<td>3,336</td>
<td>3,127</td>
<td>2,555</td>
<td>319</td>
<td>91.9%</td>
<td>31</td>
<td>1.0%</td>
<td>27</td>
<td>195</td>
</tr>
<tr>
<td>Nevada</td>
<td>8</td>
<td>Reading</td>
<td>3,419</td>
<td>3,196</td>
<td>2,579</td>
<td>337</td>
<td>91.2%</td>
<td>34</td>
<td>1.1%</td>
<td>40</td>
<td>206</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>4</td>
<td>Mathematics</td>
<td>13,223</td>
<td>12,906</td>
<td>10,033</td>
<td>1,704</td>
<td>90.9%</td>
<td>245</td>
<td>1.9%</td>
<td>240</td>
<td>684</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>4</td>
<td>Reading</td>
<td>3,357</td>
<td>3,283</td>
<td>2,585</td>
<td>456</td>
<td>92.6%</td>
<td>40</td>
<td>1.2%</td>
<td>50</td>
<td>152</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>8</td>
<td>Mathematics</td>
<td>3,177</td>
<td>3,090</td>
<td>2,345</td>
<td>459</td>
<td>90.7%</td>
<td>33</td>
<td>1.1%</td>
<td>71</td>
<td>182</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>8</td>
<td>Reading</td>
<td>3,229</td>
<td>3,153</td>
<td>2,422</td>
<td>384</td>
<td>89.0%</td>
<td>90</td>
<td>2.9%</td>
<td>66</td>
<td>191</td>
</tr>
<tr>
<td>New Jersey</td>
<td>4</td>
<td>Mathematics</td>
<td>13,124</td>
<td>12,809</td>
<td>9,936</td>
<td>1,828</td>
<td>91.8%</td>
<td>221</td>
<td>1.7%</td>
<td>256</td>
<td>566</td>
</tr>
<tr>
<td>New Jersey</td>
<td>4</td>
<td>Reading</td>
<td>3,328</td>
<td>3,249</td>
<td>2,554</td>
<td>490</td>
<td>93.7%</td>
<td>39</td>
<td>1.2%</td>
<td>33</td>
<td>133</td>
</tr>
<tr>
<td>New Jersey</td>
<td>8</td>
<td>Mathematics</td>
<td>3,449</td>
<td>3,361</td>
<td>2,668</td>
<td>466</td>
<td>93.2%</td>
<td>55</td>
<td>1.6%</td>
<td>38</td>
<td>134</td>
</tr>
<tr>
<td>New Jersey</td>
<td>8</td>
<td>Reading</td>
<td>3,147</td>
<td>3,070</td>
<td>2,292</td>
<td>488</td>
<td>90.6%</td>
<td>52</td>
<td>1.7%</td>
<td>83</td>
<td>155</td>
</tr>
<tr>
<td>New Jersey</td>
<td>8</td>
<td>Reading</td>
<td>3,200</td>
<td>3,129</td>
<td>2,422</td>
<td>384</td>
<td>89.7%</td>
<td>75</td>
<td>2.4%</td>
<td>102</td>
<td>146</td>
</tr>
<tr>
<td>New Mexico</td>
<td>4</td>
<td>Mathematics</td>
<td>16,480</td>
<td>15,406</td>
<td>12,241</td>
<td>2,025</td>
<td>92.6%</td>
<td>213</td>
<td>1.4%</td>
<td>87</td>
<td>840</td>
</tr>
<tr>
<td>New Mexico</td>
<td>4</td>
<td>Reading</td>
<td>3,172</td>
<td>3,098</td>
<td>2,990</td>
<td>673</td>
<td>93.7%</td>
<td>51</td>
<td>1.3%</td>
<td>12</td>
<td>182</td>
</tr>
<tr>
<td>New Mexico</td>
<td>4</td>
<td>Reading</td>
<td>4,300</td>
<td>4,049</td>
<td>3,200</td>
<td>578</td>
<td>93.3%</td>
<td>40</td>
<td>1.0%</td>
<td>17</td>
<td>214</td>
</tr>
<tr>
<td>New Mexico</td>
<td>8</td>
<td>Mathematics</td>
<td>3,965</td>
<td>3,695</td>
<td>2,942</td>
<td>435</td>
<td>91.4%</td>
<td>57</td>
<td>1.5%</td>
<td>28</td>
<td>233</td>
</tr>
<tr>
<td>New Mexico</td>
<td>8</td>
<td>Reading</td>
<td>4,043</td>
<td>3,754</td>
<td>3,109</td>
<td>339</td>
<td>91.8%</td>
<td>65</td>
<td>1.7%</td>
<td>30</td>
<td>211</td>
</tr>
<tr>
<td>New York</td>
<td>4</td>
<td>Mathematics</td>
<td>17,873</td>
<td>17,470</td>
<td>12,426</td>
<td>3,391</td>
<td>90.5%</td>
<td>238</td>
<td>1.4%</td>
<td>433</td>
<td>982</td>
</tr>
<tr>
<td>New York</td>
<td>4</td>
<td>Reading</td>
<td>4,509</td>
<td>4,423</td>
<td>3,138</td>
<td>891</td>
<td>91.1%</td>
<td>56</td>
<td>1.3%</td>
<td>128</td>
<td>210</td>
</tr>
<tr>
<td>New York</td>
<td>4</td>
<td>Reading</td>
<td>4,644</td>
<td>4,546</td>
<td>3,277</td>
<td>888</td>
<td>91.6%</td>
<td>63</td>
<td>1.4%</td>
<td>119</td>
<td>199</td>
</tr>
<tr>
<td>New York</td>
<td>8</td>
<td>Mathematics</td>
<td>4,336</td>
<td>4,233</td>
<td>3,034</td>
<td>766</td>
<td>89.8%</td>
<td>70</td>
<td>1.7%</td>
<td>86</td>
<td>277</td>
</tr>
<tr>
<td>New York</td>
<td>8</td>
<td>Reading</td>
<td>4,384</td>
<td>4,268</td>
<td>2,977</td>
<td>846</td>
<td>89.6%</td>
<td>49</td>
<td>1.1%</td>
<td>100</td>
<td>296</td>
</tr>
<tr>
<td>North Carolina</td>
<td>4</td>
<td>Mathematics</td>
<td>18,874</td>
<td>18,076</td>
<td>14,528</td>
<td>2,133</td>
<td>92.2%</td>
<td>261</td>
<td>1.4%</td>
<td>210</td>
<td>944</td>
</tr>
<tr>
<td>North Carolina</td>
<td>4</td>
<td>Reading</td>
<td>4,839</td>
<td>4,621</td>
<td>3,688</td>
<td>609</td>
<td>93.0%</td>
<td>62</td>
<td>1.3%</td>
<td>54</td>
<td>208</td>
</tr>
<tr>
<td>North Carolina</td>
<td>4</td>
<td>Reading</td>
<td>4,982</td>
<td>4,787</td>
<td>3,897</td>
<td>567</td>
<td>93.3%</td>
<td>73</td>
<td>1.5%</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>North Carolina</td>
<td>8</td>
<td>Mathematics</td>
<td>4,480</td>
<td>4,296</td>
<td>3,420</td>
<td>502</td>
<td>91.3%</td>
<td>56</td>
<td>1.3%</td>
<td>53</td>
<td>265</td>
</tr>
<tr>
<td>North Carolina</td>
<td>8</td>
<td>Reading</td>
<td>4,573</td>
<td>4,372</td>
<td>3,523</td>
<td>455</td>
<td>91.0%</td>
<td>70</td>
<td>1.6%</td>
<td>53</td>
<td>271</td>
</tr>
<tr>
<td>North Dakota</td>
<td>4</td>
<td>Mathematics</td>
<td>15,046</td>
<td>14,203</td>
<td>11,759</td>
<td>1,300</td>
<td>91.9%</td>
<td>490</td>
<td>3.4%</td>
<td>42</td>
<td>612</td>
</tr>
<tr>
<td>North Dakota</td>
<td>4</td>
<td>Reading</td>
<td>3,700</td>
<td>3,502</td>
<td>2,942</td>
<td>319</td>
<td>93.1%</td>
<td>90</td>
<td>2.6%</td>
<td>17</td>
<td>134</td>
</tr>
<tr>
<td>North Dakota</td>
<td>4</td>
<td>Reading</td>
<td>3,845</td>
<td>3,637</td>
<td>3,055</td>
<td>306</td>
<td>92.4%</td>
<td>146</td>
<td>4.0%</td>
<td>7</td>
<td>123</td>
</tr>
<tr>
<td>North Dakota</td>
<td>8</td>
<td>Mathematics</td>
<td>3,716</td>
<td>3,492</td>
<td>2,828</td>
<td>386</td>
<td>92.0%</td>
<td>108</td>
<td>3.1%</td>
<td>7</td>
<td>163</td>
</tr>
<tr>
<td>North Dakota</td>
<td>8</td>
<td>Reading</td>
<td>3,785</td>
<td>3,572</td>
<td>2,934</td>
<td>289</td>
<td>90.2%</td>
<td>146</td>
<td>4.1%</td>
<td>11</td>
<td>192</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
<table>
<thead>
<tr>
<th>State</th>
<th>Grade</th>
<th>Subject</th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Standard</th>
<th>Accommodated</th>
<th>Total Participation Rate</th>
<th>Excluded</th>
<th>% Excluded</th>
<th>Parent Refusal</th>
<th>Other Absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohio</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>18,579</td>
<td>17,470</td>
<td>13,246</td>
<td>2,684</td>
<td>91.2%</td>
<td>439</td>
<td>2.5%</td>
<td>110</td>
<td>991</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>Grade 4</td>
<td>Reading</td>
<td>4,682</td>
<td>4,409</td>
<td>3,403</td>
<td>653</td>
<td>92.0%</td>
<td>98</td>
<td>2.2%</td>
<td>17</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>4,491</td>
<td>4,205</td>
<td>3,157</td>
<td>671</td>
<td>91.0%</td>
<td>77</td>
<td>1.8%</td>
<td>37</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>Grade 8</td>
<td>Reading</td>
<td>4,597</td>
<td>4,304</td>
<td>3,193</td>
<td>693</td>
<td>90.3%</td>
<td>114</td>
<td>2.6%</td>
<td>36</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>13,603</td>
<td>12,713</td>
<td>10,141</td>
<td>1,610</td>
<td>92.4%</td>
<td>208</td>
<td>1.6%</td>
<td>98</td>
<td>656</td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,595</td>
<td>3,384</td>
<td>2,688</td>
<td>448</td>
<td>92.7%</td>
<td>62</td>
<td>1.8%</td>
<td>14</td>
<td>172</td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,126</td>
<td>2,903</td>
<td>2,347</td>
<td>550</td>
<td>91.6%</td>
<td>48</td>
<td>1.7%</td>
<td>35</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>Oklahoma</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,193</td>
<td>2,962</td>
<td>2,410</td>
<td>525</td>
<td>92.3%</td>
<td>39</td>
<td>1.3%</td>
<td>33</td>
<td>155</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>13,495</td>
<td>12,780</td>
<td>10,244</td>
<td>1,465</td>
<td>91.6%</td>
<td>252</td>
<td>2.0%</td>
<td>138</td>
<td>681</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,549</td>
<td>3,372</td>
<td>2,605</td>
<td>496</td>
<td>92.0%</td>
<td>80</td>
<td>2.4%</td>
<td>25</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,123</td>
<td>2,915</td>
<td>2,363</td>
<td>283</td>
<td>91.5%</td>
<td>45</td>
<td>1.5%</td>
<td>41</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>Oregon</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,158</td>
<td>2,988</td>
<td>2,424</td>
<td>302</td>
<td>91.2%</td>
<td>43</td>
<td>1.4%</td>
<td>49</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>17,781</td>
<td>17,098</td>
<td>13,269</td>
<td>2,331</td>
<td>91.2%</td>
<td>398</td>
<td>2.3%</td>
<td>152</td>
<td>948</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Grade 4</td>
<td>Reading</td>
<td>4,531</td>
<td>4,375</td>
<td>3,480</td>
<td>566</td>
<td>92.5%</td>
<td>93</td>
<td>2.1%</td>
<td>23</td>
<td>213</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>4,648</td>
<td>4,478</td>
<td>3,565</td>
<td>560</td>
<td>92.1%</td>
<td>121</td>
<td>2.7%</td>
<td>21</td>
<td>211</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Grade 8</td>
<td>Reading</td>
<td>4,260</td>
<td>4,077</td>
<td>3,073</td>
<td>612</td>
<td>90.4%</td>
<td>86</td>
<td>2.1%</td>
<td>46</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>10,968</td>
<td>10,530</td>
<td>7,407</td>
<td>2,419</td>
<td>93.3%</td>
<td>13</td>
<td>0.1%</td>
<td>130</td>
<td>561</td>
<td></td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>Grade 4</td>
<td>Reading</td>
<td>5,086</td>
<td>4,906</td>
<td>3,381</td>
<td>1,239</td>
<td>94.2%</td>
<td>11</td>
<td>0.2%</td>
<td>55</td>
<td>220</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>5,882</td>
<td>5,624</td>
<td>4,026</td>
<td>1,610</td>
<td>92.6%</td>
<td>2</td>
<td>0.0%</td>
<td>75</td>
<td>341</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,393</td>
<td>3,291</td>
<td>2,617</td>
<td>475</td>
<td>94.0%</td>
<td>38</td>
<td>1.2%</td>
<td>10</td>
<td>151</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,483</td>
<td>3,368</td>
<td>2,747</td>
<td>409</td>
<td>93.7%</td>
<td>43</td>
<td>1.3%</td>
<td>8</td>
<td>161</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,242</td>
<td>3,155</td>
<td>2,494</td>
<td>437</td>
<td>92.9%</td>
<td>36</td>
<td>1.1%</td>
<td>15</td>
<td>173</td>
<td></td>
</tr>
<tr>
<td>Rhode Island</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,292</td>
<td>3,189</td>
<td>2,469</td>
<td>454</td>
<td>91.7%</td>
<td>43</td>
<td>1.3%</td>
<td>21</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>12,903</td>
<td>12,158</td>
<td>10,229</td>
<td>1,121</td>
<td>93.4%</td>
<td>183</td>
<td>1.5%</td>
<td>74</td>
<td>551</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,214</td>
<td>3,053</td>
<td>2,575</td>
<td>327</td>
<td>95.1%</td>
<td>31</td>
<td>1.0%</td>
<td>12</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,320</td>
<td>3,150</td>
<td>2,618</td>
<td>309</td>
<td>92.9%</td>
<td>58</td>
<td>1.8%</td>
<td>17</td>
<td>148</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,159</td>
<td>2,954</td>
<td>2,478</td>
<td>274</td>
<td>93.2%</td>
<td>37</td>
<td>1.3%</td>
<td>24</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>South Carolina</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,210</td>
<td>3,001</td>
<td>2,558</td>
<td>211</td>
<td>92.3%</td>
<td>57</td>
<td>1.9%</td>
<td>21</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>13,430</td>
<td>12,862</td>
<td>10,942</td>
<td>1,045</td>
<td>93.2%</td>
<td>253</td>
<td>2.0%</td>
<td>20</td>
<td>602</td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,438</td>
<td>3,301</td>
<td>2,779</td>
<td>321</td>
<td>93.9%</td>
<td>50</td>
<td>1.5%</td>
<td>2</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,531</td>
<td>3,388</td>
<td>2,867</td>
<td>295</td>
<td>93.3%</td>
<td>78</td>
<td>2.3%</td>
<td>7</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>Grade 8</td>
<td>Mathematics</td>
<td>3,194</td>
<td>3,040</td>
<td>2,569</td>
<td>260</td>
<td>93.1%</td>
<td>42</td>
<td>1.4%</td>
<td>7</td>
<td>162</td>
<td></td>
</tr>
<tr>
<td>South Dakota</td>
<td>Grade 8</td>
<td>Reading</td>
<td>3,267</td>
<td>3,133</td>
<td>2,727</td>
<td>169</td>
<td>92.4%</td>
<td>83</td>
<td>2.6%</td>
<td>4</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>Grade 4</td>
<td>Mathematics</td>
<td>13,280</td>
<td>12,539</td>
<td>10,266</td>
<td>1,240</td>
<td>91.8%</td>
<td>299</td>
<td>2.4%</td>
<td>115</td>
<td>619</td>
<td></td>
</tr>
<tr>
<td>Tennessee</td>
<td>Grade 4</td>
<td>Reading</td>
<td>3,371</td>
<td>3,193</td>
<td>2,592</td>
<td>378</td>
<td>93.0%</td>
<td>42</td>
<td>1.3%</td>
<td>31</td>
<td>150</td>
<td></td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
## Table 4
Number of Students Assessed, Accommodated, Excluded, and Absent

<table>
<thead>
<tr>
<th>State</th>
<th>Grade</th>
<th>Subject</th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Accommodated</th>
<th>Total Participation Rate</th>
<th>Excluded</th>
<th>% Excluded</th>
<th>Parent Refusal</th>
<th>Other Absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tennessee</td>
<td>4</td>
<td>Reading</td>
<td>3,504</td>
<td>3,305</td>
<td></td>
<td>2,654</td>
<td>396</td>
<td>92.3%</td>
<td>105</td>
<td>3.2%</td>
<td>27</td>
</tr>
<tr>
<td>Tennessee</td>
<td>8</td>
<td>Mathematics</td>
<td>3,194</td>
<td>3,012</td>
<td></td>
<td>2,485</td>
<td>258</td>
<td>91.1%</td>
<td>56</td>
<td>1.9%</td>
<td>34</td>
</tr>
<tr>
<td>Tennessee</td>
<td>8</td>
<td>Reading</td>
<td>3,211</td>
<td>3,029</td>
<td></td>
<td>2,535</td>
<td>208</td>
<td>90.6%</td>
<td>96</td>
<td>3.2%</td>
<td>23</td>
</tr>
<tr>
<td>Texas</td>
<td>4</td>
<td>Mathematics</td>
<td>9,201</td>
<td>8,765</td>
<td></td>
<td>6,084</td>
<td>2,132</td>
<td>93.7%</td>
<td>169</td>
<td>1.9%</td>
<td>72</td>
</tr>
<tr>
<td>Texas</td>
<td>4</td>
<td>Reading</td>
<td>9,131</td>
<td>9,043</td>
<td></td>
<td>6,227</td>
<td>1,803</td>
<td>88.8%</td>
<td>644</td>
<td>7.1%</td>
<td>60</td>
</tr>
<tr>
<td>Texas</td>
<td>8</td>
<td>Mathematics</td>
<td>8,756</td>
<td>8,223</td>
<td></td>
<td>6,365</td>
<td>1,138</td>
<td>91.2%</td>
<td>169</td>
<td>2.1%</td>
<td>58</td>
</tr>
<tr>
<td>Texas</td>
<td>8</td>
<td>Reading</td>
<td>8,902</td>
<td>8,394</td>
<td></td>
<td>6,551</td>
<td>982</td>
<td>89.7%</td>
<td>293</td>
<td>3.5%</td>
<td>55</td>
</tr>
<tr>
<td>Utah</td>
<td>4</td>
<td>Mathematics</td>
<td>13,977</td>
<td>13,353</td>
<td>10,914</td>
<td>1,284</td>
<td>91.4%</td>
<td>301</td>
<td>2.3%</td>
<td>125</td>
<td>727</td>
</tr>
<tr>
<td>Utah</td>
<td>4</td>
<td>Reading</td>
<td>3,551</td>
<td>3,376</td>
<td></td>
<td>2,753</td>
<td>406</td>
<td>93.6%</td>
<td>43</td>
<td>1.3%</td>
<td>17</td>
</tr>
<tr>
<td>Utah</td>
<td>8</td>
<td>Mathematics</td>
<td>3,474</td>
<td>3,198</td>
<td></td>
<td>2,612</td>
<td>295</td>
<td>90.9%</td>
<td>47</td>
<td>1.5%</td>
<td>45</td>
</tr>
<tr>
<td>Utah</td>
<td>8</td>
<td>Reading</td>
<td>3,424</td>
<td>3,295</td>
<td></td>
<td>2,710</td>
<td>261</td>
<td>90.2%</td>
<td>101</td>
<td>3.1%</td>
<td>37</td>
</tr>
<tr>
<td>Vermont</td>
<td></td>
<td></td>
<td>12,209</td>
<td>11,842</td>
<td>9,372</td>
<td>1,666</td>
<td>93.2%</td>
<td>128</td>
<td>1.1%</td>
<td>64</td>
<td>612</td>
</tr>
<tr>
<td>Vermont</td>
<td>4</td>
<td>Mathematics</td>
<td>3,000</td>
<td>2,903</td>
<td></td>
<td>2,300</td>
<td>421</td>
<td>93.7%</td>
<td>40</td>
<td>1.4%</td>
<td>14</td>
</tr>
<tr>
<td>Vermont</td>
<td>4</td>
<td>Reading</td>
<td>3,094</td>
<td>3,007</td>
<td></td>
<td>2,397</td>
<td>428</td>
<td>93.9%</td>
<td>35</td>
<td>1.2%</td>
<td>19</td>
</tr>
<tr>
<td>Vermont</td>
<td>8</td>
<td>Mathematics</td>
<td>3,025</td>
<td>2,933</td>
<td></td>
<td>2,309</td>
<td>422</td>
<td>93.1%</td>
<td>25</td>
<td>0.9%</td>
<td>10</td>
</tr>
<tr>
<td>Vermont</td>
<td>8</td>
<td>Reading</td>
<td>3,090</td>
<td>2,999</td>
<td></td>
<td>2,366</td>
<td>395</td>
<td>92.1%</td>
<td>28</td>
<td>0.9%</td>
<td>21</td>
</tr>
<tr>
<td>Virginia</td>
<td>4</td>
<td>Mathematics</td>
<td>13,289</td>
<td>12,696</td>
<td>10,357</td>
<td>1,409</td>
<td>92.7%</td>
<td>175</td>
<td>1.4%</td>
<td>128</td>
<td>627</td>
</tr>
<tr>
<td>Virginia</td>
<td>4</td>
<td>Reading</td>
<td>3,330</td>
<td>3,204</td>
<td></td>
<td>2,575</td>
<td>403</td>
<td>92.9%</td>
<td>49</td>
<td>1.5%</td>
<td>22</td>
</tr>
<tr>
<td>Virginia</td>
<td>8</td>
<td>Mathematics</td>
<td>3,227</td>
<td>3,066</td>
<td></td>
<td>2,493</td>
<td>340</td>
<td>92.4%</td>
<td>33</td>
<td>1.1%</td>
<td>43</td>
</tr>
<tr>
<td>Virginia</td>
<td>8</td>
<td>Reading</td>
<td>3,284</td>
<td>3,122</td>
<td></td>
<td>2,572</td>
<td>291</td>
<td>91.7%</td>
<td>43</td>
<td>1.4%</td>
<td>42</td>
</tr>
<tr>
<td>Washington</td>
<td></td>
<td></td>
<td>13,581</td>
<td>13,085</td>
<td>10,401</td>
<td>1,449</td>
<td>90.6%</td>
<td>311</td>
<td>2.4%</td>
<td>161</td>
<td>763</td>
</tr>
<tr>
<td>Washington</td>
<td>4</td>
<td>Mathematics</td>
<td>3,554</td>
<td>3,424</td>
<td></td>
<td>2,696</td>
<td>454</td>
<td>92.0%</td>
<td>75</td>
<td>2.2%</td>
<td>36</td>
</tr>
<tr>
<td>Washington</td>
<td>4</td>
<td>Reading</td>
<td>3,694</td>
<td>3,561</td>
<td></td>
<td>2,850</td>
<td>413</td>
<td>91.6%</td>
<td>99</td>
<td>2.8%</td>
<td>39</td>
</tr>
<tr>
<td>Washington</td>
<td>8</td>
<td>Mathematics</td>
<td>3,139</td>
<td>3,016</td>
<td></td>
<td>2,380</td>
<td>313</td>
<td>89.3%</td>
<td>62</td>
<td>2.1%</td>
<td>51</td>
</tr>
<tr>
<td>Washington</td>
<td>8</td>
<td>Reading</td>
<td>3,194</td>
<td>3,084</td>
<td></td>
<td>2,475</td>
<td>269</td>
<td>89.0%</td>
<td>75</td>
<td>2.4%</td>
<td>35</td>
</tr>
<tr>
<td>West Virginia</td>
<td>4</td>
<td>Mathematics</td>
<td>12,831</td>
<td>12,162</td>
<td>10,197</td>
<td>974</td>
<td>91.9%</td>
<td>214</td>
<td>1.8%</td>
<td>84</td>
<td>693</td>
</tr>
<tr>
<td>West Virginia</td>
<td>4</td>
<td>Reading</td>
<td>3,156</td>
<td>3,007</td>
<td></td>
<td>2,517</td>
<td>280</td>
<td>93.0%</td>
<td>54</td>
<td>1.8%</td>
<td>11</td>
</tr>
<tr>
<td>West Virginia</td>
<td>4</td>
<td>Reading</td>
<td>3,196</td>
<td>3,119</td>
<td></td>
<td>2,611</td>
<td>256</td>
<td>91.9%</td>
<td>55</td>
<td>1.8%</td>
<td>26</td>
</tr>
<tr>
<td>West Virginia</td>
<td>8</td>
<td>Mathematics</td>
<td>3,178</td>
<td>2,987</td>
<td></td>
<td>2,493</td>
<td>230</td>
<td>91.2%</td>
<td>48</td>
<td>1.6%</td>
<td>20</td>
</tr>
<tr>
<td>West Virginia</td>
<td>8</td>
<td>Reading</td>
<td>3,233</td>
<td>3,049</td>
<td></td>
<td>2,576</td>
<td>208</td>
<td>91.3%</td>
<td>57</td>
<td>1.9%</td>
<td>27</td>
</tr>
<tr>
<td>Wisconsin</td>
<td></td>
<td></td>
<td>17,764</td>
<td>17,167</td>
<td>12,930</td>
<td>2,893</td>
<td>92.2%</td>
<td>391</td>
<td>2.3%</td>
<td>92</td>
<td>861</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>4</td>
<td>Mathematics</td>
<td>4,445</td>
<td>4,297</td>
<td></td>
<td>3,246</td>
<td>755</td>
<td>93.1%</td>
<td>98</td>
<td>2.3%</td>
<td>14</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>4</td>
<td>Reading</td>
<td>4,540</td>
<td>4,407</td>
<td></td>
<td>3,287</td>
<td>788</td>
<td>92.5%</td>
<td>102</td>
<td>2.3%</td>
<td>21</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>8</td>
<td>Mathematics</td>
<td>4,346</td>
<td>4,203</td>
<td></td>
<td>3,174</td>
<td>669</td>
<td>91.4%</td>
<td>91</td>
<td>2.2%</td>
<td>28</td>
</tr>
<tr>
<td>Wisconsin</td>
<td>8</td>
<td>Reading</td>
<td>4,433</td>
<td>4,260</td>
<td></td>
<td>3,223</td>
<td>681</td>
<td>91.6%</td>
<td>100</td>
<td>2.3%</td>
<td>29</td>
</tr>
<tr>
<td>Wyoming</td>
<td></td>
<td></td>
<td>13,907</td>
<td>13,191</td>
<td>10,718</td>
<td>1,527</td>
<td>92.8%</td>
<td>161</td>
<td>1.2%</td>
<td>52</td>
<td>733</td>
</tr>
<tr>
<td>Wyoming</td>
<td>4</td>
<td>Mathematics</td>
<td>3,525</td>
<td>3,358</td>
<td></td>
<td>2,740</td>
<td>406</td>
<td>93.7%</td>
<td>34</td>
<td>1.0%</td>
<td>10</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
### NAEP 2013 Operations Report Table 4

Number of Students Assessed, Accommodated, Excluded, and Absent

<table>
<thead>
<tr>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Total Participation Rate</th>
<th>Not Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standard</td>
<td>Accommodated</td>
<td>Excluded</td>
</tr>
<tr>
<td>Wyoming</td>
<td>Grade 4 Reading</td>
<td>3,624</td>
<td>3,462</td>
<td>2,817</td>
</tr>
<tr>
<td>Wyoming</td>
<td>Grade 8 Mathematics</td>
<td>3,348</td>
<td>3,138</td>
<td>2,543</td>
</tr>
<tr>
<td>Wyoming</td>
<td>Grade 8 Reading</td>
<td>3,410</td>
<td>3,233</td>
<td>2,618</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
<table>
<thead>
<tr>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th></th>
<th></th>
<th>Not Assessed</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standard</td>
<td>Accommodated</td>
<td>Total Participation Rate</td>
<td>Excluded</td>
<td>%Excluded</td>
<td>Parent Refusal</td>
<td>Other Absence</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>831,930</td>
<td>792,560</td>
<td>91.9%</td>
<td>16,082</td>
<td>2.0%</td>
<td>6,419</td>
<td>41,407</td>
</tr>
<tr>
<td>Total Grade 4 Mathematics</td>
<td></td>
<td>214,934</td>
<td>205,142</td>
<td>93.2%</td>
<td>3,116</td>
<td>1.5%</td>
<td>1,353</td>
<td>9,519</td>
</tr>
<tr>
<td>Total Grade 4 Reading</td>
<td></td>
<td>216,426</td>
<td>206,683</td>
<td>92.1%</td>
<td>5,596</td>
<td>2.7%</td>
<td>1,375</td>
<td>9,305</td>
</tr>
<tr>
<td>Total Grade 8 Mathematics</td>
<td></td>
<td>201,461</td>
<td>191,407</td>
<td>91.6%</td>
<td>2,912</td>
<td>1.5%</td>
<td>1,908</td>
<td>11,308</td>
</tr>
<tr>
<td>Total Grade 8 Reading</td>
<td></td>
<td>199,109</td>
<td>189,328</td>
<td>90.7%</td>
<td>4,458</td>
<td>2.4%</td>
<td>1,783</td>
<td>11,275</td>
</tr>
<tr>
<td>Total Public</td>
<td></td>
<td>818,369</td>
<td>779,146</td>
<td>91.9%</td>
<td>16,047</td>
<td>2.1%</td>
<td>6,349</td>
<td>40,858</td>
</tr>
<tr>
<td>Total Public Grade 4 Mathematics</td>
<td></td>
<td>211,657</td>
<td>201,896</td>
<td>93.1%</td>
<td>3,114</td>
<td>1.5%</td>
<td>1,337</td>
<td>9,392</td>
</tr>
<tr>
<td>Total Public Grade 4 Reading</td>
<td></td>
<td>213,046</td>
<td>203,340</td>
<td>92.1%</td>
<td>5,581</td>
<td>2.7%</td>
<td>1,361</td>
<td>9,174</td>
</tr>
<tr>
<td>Total Public Grade 8 Mathematics</td>
<td></td>
<td>198,049</td>
<td>188,035</td>
<td>91.5%</td>
<td>2,904</td>
<td>1.5%</td>
<td>1,685</td>
<td>11,160</td>
</tr>
<tr>
<td>Total Public Grade 8 Reading</td>
<td></td>
<td>195,617</td>
<td>185,875</td>
<td>90.7%</td>
<td>4,448</td>
<td>2.4%</td>
<td>1,766</td>
<td>11,132</td>
</tr>
<tr>
<td>Total Nonpublic</td>
<td></td>
<td>13,561</td>
<td>13,414</td>
<td>95.1%</td>
<td>35</td>
<td>0.3%</td>
<td>70</td>
<td>549</td>
</tr>
<tr>
<td>Total Nonpublic Grade 4 Mathematics</td>
<td></td>
<td>3,277</td>
<td>3,246</td>
<td>95.5%</td>
<td>2</td>
<td>0.1%</td>
<td>16</td>
<td>127</td>
</tr>
<tr>
<td>Total Nonpublic Grade 4 Reading</td>
<td></td>
<td>3,380</td>
<td>3,343</td>
<td>95.2%</td>
<td>15</td>
<td>0.4%</td>
<td>14</td>
<td>131</td>
</tr>
<tr>
<td>Total Nonpublic Grade 8 Mathematics</td>
<td></td>
<td>3,412</td>
<td>3,372</td>
<td>94.7%</td>
<td>8</td>
<td>0.2%</td>
<td>23</td>
<td>148</td>
</tr>
<tr>
<td>Total Nonpublic Grade 8 Reading</td>
<td></td>
<td>3,492</td>
<td>3,453</td>
<td>95.1%</td>
<td>10</td>
<td>0.3%</td>
<td>17</td>
<td>143</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
## Table 4
Number of Students Assessed, Accommodated, Excluded, and Absent - by Race/Ethnicity

<table>
<thead>
<tr>
<th></th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Not Assessed</th>
<th>Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standard</td>
<td>Accommodated</td>
<td>Total Participation Rate</td>
</tr>
<tr>
<td>Total</td>
<td>831,930</td>
<td>792,560</td>
<td>631,303</td>
<td>97,349</td>
</tr>
<tr>
<td>Total</td>
<td>214,934</td>
<td>205,142</td>
<td>162,475</td>
<td>28,679</td>
</tr>
<tr>
<td>Total</td>
<td>216,426</td>
<td>206,683</td>
<td>164,835</td>
<td>25,572</td>
</tr>
<tr>
<td>Total</td>
<td>201,461</td>
<td>191,407</td>
<td>152,232</td>
<td>23,047</td>
</tr>
<tr>
<td>Total</td>
<td>199,109</td>
<td>189,328</td>
<td>151,761</td>
<td>20,587</td>
</tr>
<tr>
<td>White, not Hispanic</td>
<td>436,625</td>
<td>419,910</td>
<td>348,706</td>
<td>38,446</td>
</tr>
<tr>
<td>White, not Hispanic</td>
<td>110,037</td>
<td>105,961</td>
<td>88,535</td>
<td>10,221</td>
</tr>
<tr>
<td>White, not Hispanic</td>
<td>113,628</td>
<td>109,518</td>
<td>91,538</td>
<td>9,977</td>
</tr>
<tr>
<td>White, not Hispanic</td>
<td>105,206</td>
<td>100,919</td>
<td>83,094</td>
<td>10,943</td>
</tr>
<tr>
<td>Black or African American, not Hispanic</td>
<td>146,123</td>
<td>135,961</td>
<td>107,176</td>
<td>16,767</td>
</tr>
<tr>
<td>Black or African American, not Hispanic</td>
<td>37,044</td>
<td>34,644</td>
<td>27,698</td>
<td>4,407</td>
</tr>
<tr>
<td>Black or African American, not Hispanic</td>
<td>38,003</td>
<td>35,534</td>
<td>28,526</td>
<td>3,976</td>
</tr>
<tr>
<td>Hispanic, of any race</td>
<td>169,977</td>
<td>161,673</td>
<td>115,456</td>
<td>33,024</td>
</tr>
<tr>
<td>Hispanic, of any race</td>
<td>46,908</td>
<td>44,698</td>
<td>30,362</td>
<td>11,403</td>
</tr>
<tr>
<td>Hispanic, of any race</td>
<td>43,265</td>
<td>41,225</td>
<td>28,443</td>
<td>12,926</td>
</tr>
<tr>
<td>Hispanic, of any race</td>
<td>42,557</td>
<td>39,404</td>
<td>30,308</td>
<td>8,590</td>
</tr>
<tr>
<td>Hispanic, of any race</td>
<td>37,247</td>
<td>35,346</td>
<td>26,813</td>
<td>5,405</td>
</tr>
<tr>
<td>Asian, not Hispanic</td>
<td>33,618</td>
<td>32,702</td>
<td>26,743</td>
<td>3,774</td>
</tr>
<tr>
<td>Asian, not Hispanic</td>
<td>8,891</td>
<td>8,615</td>
<td>6,981</td>
<td>1,125</td>
</tr>
<tr>
<td>Asian, not Hispanic</td>
<td>9,169</td>
<td>8,883</td>
<td>7,207</td>
<td>1,686</td>
</tr>
<tr>
<td>Asian, not Hispanic</td>
<td>7,786</td>
<td>7,594</td>
<td>6,265</td>
<td>818</td>
</tr>
<tr>
<td>Asian, not Hispanic</td>
<td>7,772</td>
<td>7,610</td>
<td>6,290</td>
<td>735</td>
</tr>
<tr>
<td>American Indian or Alaska Native, not Hispanic</td>
<td>17,483</td>
<td>16,046</td>
<td>12,037</td>
<td>2,477</td>
</tr>
<tr>
<td>American Indian or Alaska Native, not Hispanic</td>
<td>4,359</td>
<td>4,030</td>
<td>3,008</td>
<td>696</td>
</tr>
<tr>
<td>American Indian or Alaska Native, not Hispanic</td>
<td>4,556</td>
<td>4,235</td>
<td>3,215</td>
<td>610</td>
</tr>
<tr>
<td>American Indian or Alaska Native, not Hispanic</td>
<td>4,218</td>
<td>3,839</td>
<td>2,816</td>
<td>1,037</td>
</tr>
<tr>
<td>American Indian or Alaska Native, not Hispanic</td>
<td>4,430</td>
<td>3,942</td>
<td>2,998</td>
<td>970</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander, not Hispanic</td>
<td>6,996</td>
<td>6,602</td>
<td>5,018</td>
<td>921</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander, not Hispanic</td>
<td>1,803</td>
<td>1,697</td>
<td>1,329</td>
<td>248</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander, not Hispanic</td>
<td>1,794</td>
<td>1,675</td>
<td>1,306</td>
<td>214</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander, not Hispanic</td>
<td>1,679</td>
<td>1,603</td>
<td>1,178</td>
<td>241</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander, not Hispanic</td>
<td>1,720</td>
<td>1,627</td>
<td>1,205</td>
<td>216</td>
</tr>
<tr>
<td>Two or more races, not Hispanic</td>
<td>20,850</td>
<td>19,666</td>
<td>16,167</td>
<td>1,960</td>
</tr>
<tr>
<td>Two or more races, not Hispanic</td>
<td>5,818</td>
<td>5,497</td>
<td>4,562</td>
<td>957</td>
</tr>
<tr>
<td>Two or more races, not Hispanic</td>
<td>5,932</td>
<td>5,613</td>
<td>4,600</td>
<td>924</td>
</tr>
<tr>
<td>Two or more races, not Hispanic</td>
<td>4,509</td>
<td>4,224</td>
<td>3,465</td>
<td>417</td>
</tr>
<tr>
<td>Two or more races, not Hispanic</td>
<td>4,591</td>
<td>4,332</td>
<td>3,540</td>
<td>403</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
### NAEP 2013 Operations Report Table 4
Number of Students Assessed, Accommodated, Excluded, and Absent - by NSLP

<table>
<thead>
<tr>
<th></th>
<th>Original Student Sample</th>
<th>Actual Student Sample</th>
<th>Assessed</th>
<th>Total Participation Rate</th>
<th>Not Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
<td>Accommodated</td>
<td></td>
<td>Excluded</td>
<td>Parent Refusal</td>
</tr>
<tr>
<td>Total</td>
<td>831,930</td>
<td>792,560</td>
<td>631,303</td>
<td>97,349</td>
<td>16,082</td>
</tr>
<tr>
<td>Total Grade 4 Mathematics</td>
<td>214,934</td>
<td>205,142</td>
<td>162,475</td>
<td>28,679</td>
<td>3,116</td>
</tr>
<tr>
<td>Total Grade 4 Reading</td>
<td>216,426</td>
<td>206,853</td>
<td>164,835</td>
<td>25,572</td>
<td>5,596</td>
</tr>
<tr>
<td>Total Grade 8 Mathematics</td>
<td>201,461</td>
<td>191,407</td>
<td>152,232</td>
<td>23,047</td>
<td>2,912</td>
</tr>
<tr>
<td>Total Grade 8 Reading</td>
<td>199,109</td>
<td>189,328</td>
<td>151,761</td>
<td>20,051</td>
<td>4,458</td>
</tr>
<tr>
<td>Student not eligible</td>
<td>356,447</td>
<td>346,886</td>
<td>296,699</td>
<td>26,019</td>
<td>4,490</td>
</tr>
<tr>
<td>Student not eligible Grade 4 Mathematics</td>
<td>87,476</td>
<td>85,215</td>
<td>73,175</td>
<td>6,880</td>
<td>853</td>
</tr>
<tr>
<td>Student not eligible Grade 4 Reading</td>
<td>90,213</td>
<td>87,979</td>
<td>75,728</td>
<td>6,570</td>
<td>1,395</td>
</tr>
<tr>
<td>Student not eligible Grade 8 Mathematics</td>
<td>88,491</td>
<td>85,949</td>
<td>73,095</td>
<td>6,461</td>
<td>922</td>
</tr>
<tr>
<td>Student not eligible Grade 8 Reading</td>
<td>90,267</td>
<td>87,743</td>
<td>74,701</td>
<td>6,108</td>
<td>1,320</td>
</tr>
<tr>
<td>Free lunch</td>
<td>411,397</td>
<td>385,372</td>
<td>284,841</td>
<td>65,218</td>
<td>10,402</td>
</tr>
<tr>
<td>Free lunch Grade 4 Mathematics</td>
<td>111,232</td>
<td>104,704</td>
<td>76,740</td>
<td>20,106</td>
<td>2,042</td>
</tr>
<tr>
<td>Free lunch Grade 4 Reading</td>
<td>109,428</td>
<td>102,921</td>
<td>76,065</td>
<td>17,482</td>
<td>3,732</td>
</tr>
<tr>
<td>Free lunch Grade 8 Mathematics</td>
<td>97,514</td>
<td>90,889</td>
<td>67,142</td>
<td>15,055</td>
<td>1,810</td>
</tr>
<tr>
<td>Free lunch Grade 8 Reading</td>
<td>93,223</td>
<td>86,858</td>
<td>64,894</td>
<td>12,575</td>
<td>2,818</td>
</tr>
<tr>
<td>Reduced price lunch</td>
<td>38,723</td>
<td>37,517</td>
<td>30,304</td>
<td>4,500</td>
<td>707</td>
</tr>
<tr>
<td>Reduced price lunch Grade 4 Mathematics</td>
<td>9,577</td>
<td>9,307</td>
<td>7,503</td>
<td>1,205</td>
<td>145</td>
</tr>
<tr>
<td>Reduced price lunch Grade 4 Reading</td>
<td>9,968</td>
<td>9,682</td>
<td>7,805</td>
<td>1,177</td>
<td>234</td>
</tr>
<tr>
<td>Reduced price lunch Grade 8 Mathematics</td>
<td>9,561</td>
<td>9,242</td>
<td>7,442</td>
<td>1,130</td>
<td>131</td>
</tr>
<tr>
<td>Reduced price lunch Grade 8 Reading</td>
<td>9,617</td>
<td>9,286</td>
<td>7,554</td>
<td>988</td>
<td>197</td>
</tr>
<tr>
<td>School not participating</td>
<td>21,708</td>
<td>20,260</td>
<td>17,449</td>
<td>1,387</td>
<td>398</td>
</tr>
<tr>
<td>School not participating Grade 4 Mathematics</td>
<td>5,855</td>
<td>5,400</td>
<td>4,642</td>
<td>448</td>
<td>54</td>
</tr>
<tr>
<td>School not participating Grade 4 Reading</td>
<td>6,007</td>
<td>5,564</td>
<td>4,802</td>
<td>298</td>
<td>213</td>
</tr>
<tr>
<td>School not participating Grade 8 Mathematics</td>
<td>4,889</td>
<td>4,617</td>
<td>3,985</td>
<td>331</td>
<td>34</td>
</tr>
<tr>
<td>School not participating Grade 8 Reading</td>
<td>4,957</td>
<td>4,679</td>
<td>4,020</td>
<td>310</td>
<td>97</td>
</tr>
<tr>
<td>School refused</td>
<td>165</td>
<td>163</td>
<td>150</td>
<td>8</td>
<td>96.9%</td>
</tr>
<tr>
<td>School refused Grade 4 Mathematics</td>
<td>60</td>
<td>58</td>
<td>52</td>
<td>4</td>
<td>96.6%</td>
</tr>
<tr>
<td>School refused Grade 4 Reading</td>
<td>64</td>
<td>64</td>
<td>59</td>
<td>3</td>
<td>96.9%</td>
</tr>
<tr>
<td>School refused Grade 8 Mathematics</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>100.0%</td>
</tr>
<tr>
<td>School refused Grade 8 Reading</td>
<td>21</td>
<td>21</td>
<td>19</td>
<td>1</td>
<td>95.2%</td>
</tr>
<tr>
<td>Information unavailable</td>
<td>3,490</td>
<td>2,362</td>
<td>1,860</td>
<td>217</td>
<td>87.9%</td>
</tr>
<tr>
<td>Information unavailable Grade 4 Mathematics</td>
<td>734</td>
<td>458</td>
<td>363</td>
<td>36</td>
<td>87.1%</td>
</tr>
<tr>
<td>Information unavailable Grade 4 Reading</td>
<td>746</td>
<td>473</td>
<td>376</td>
<td>42</td>
<td>88.4%</td>
</tr>
<tr>
<td>Information unavailable Grade 8 Mathematics</td>
<td>986</td>
<td>690</td>
<td>548</td>
<td>70</td>
<td>89.6%</td>
</tr>
<tr>
<td>Information unavailable Grade 8 Reading</td>
<td>1,024</td>
<td>741</td>
<td>573</td>
<td>69</td>
<td>86.6%</td>
</tr>
</tbody>
</table>

Other reasons include temporary absence; long-term absence; chronic truant; suspended or expelled; in school, did not attend; disruptive behavior; student refusal; and other.
Average scale scores and percentages for reading, grade 8 by status as English Language Learner, 3 categories [ELL3], year and jurisdiction: 2013, 2011, 2009, 2007, 2005, and 2003

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Yes</th>
<th></th>
<th>No</th>
<th></th>
<th>Formerly ELL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average scale score</td>
<td>Percentage</td>
<td>Average scale score</td>
<td>Percentage</td>
<td>Average scale score</td>
</tr>
<tr>
<td>2013</td>
<td>National public</td>
<td>225</td>
<td>5</td>
<td>269</td>
<td>92</td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>220</td>
<td>12</td>
<td>269</td>
<td>74</td>
<td>256</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>226</td>
<td>4</td>
<td>268</td>
<td>88</td>
<td>262</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>215</td>
<td>6</td>
<td>270</td>
<td>94</td>
<td>‡</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>227</td>
<td>7</td>
<td>267</td>
<td>89</td>
<td>255</td>
</tr>
<tr>
<td>2011</td>
<td>National public</td>
<td>223</td>
<td>5</td>
<td>266</td>
<td>91</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>220</td>
<td>17</td>
<td>264</td>
<td>69</td>
<td>252</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>225</td>
<td>4</td>
<td>264</td>
<td>88</td>
<td>261</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>216</td>
<td>5</td>
<td>268</td>
<td>95</td>
<td>‡</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>225</td>
<td>7</td>
<td>265</td>
<td>89</td>
<td>247</td>
</tr>
<tr>
<td>2009</td>
<td>National public</td>
<td>219</td>
<td>5</td>
<td>265</td>
<td>92</td>
<td>253</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>215</td>
<td>19</td>
<td>263</td>
<td>68</td>
<td>252</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>233</td>
<td>3</td>
<td>266</td>
<td>94</td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>213</td>
<td>3</td>
<td>266</td>
<td>97</td>
<td>‡</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>216</td>
<td>6</td>
<td>263</td>
<td>91</td>
<td>255</td>
</tr>
<tr>
<td>2007</td>
<td>National public</td>
<td>222</td>
<td>6</td>
<td>263</td>
<td>93</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>219</td>
<td>21</td>
<td>261</td>
<td>71</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>232</td>
<td>3</td>
<td>261</td>
<td>95</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>211</td>
<td>3</td>
<td>265</td>
<td>97</td>
<td>‡</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>215</td>
<td>6</td>
<td>264</td>
<td>91</td>
<td>256</td>
</tr>
<tr>
<td>2005</td>
<td>National public</td>
<td>224</td>
<td>5</td>
<td>263</td>
<td>93</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>222</td>
<td>20</td>
<td>258</td>
<td>75</td>
<td>258</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>221</td>
<td>4</td>
<td>257</td>
<td>95</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>221</td>
<td>3</td>
<td>267</td>
<td>88</td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>216</td>
<td>6</td>
<td>261</td>
<td>93</td>
<td>243</td>
</tr>
<tr>
<td>2003</td>
<td>National public</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Texas</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

— Not available.
‡ Reporting standards not met.
# Rounds to zero.

NOTE: The NAEP Reading scale ranges from 0 to 500. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.

## Average scale scores and percentages for mathematics, grade 8 by status as English Language Learner, 3 categories [ELL3], year and jurisdiction: 2013, 2011, 2009, 2007, and 2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Jurisdiction</th>
<th>Yes</th>
<th>No</th>
<th>Formerly ELL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Average scale score</td>
<td>Percentage</td>
<td>Average scale score</td>
</tr>
<tr>
<td>2013</td>
<td>National</td>
<td>246</td>
<td>5</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>235</td>
<td>12</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>243</td>
<td>5</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>241</td>
<td>6</td>
<td>285</td>
</tr>
<tr>
<td>2011</td>
<td>National</td>
<td>244</td>
<td>5</td>
<td>287</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>234</td>
<td>17</td>
<td>283</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>246</td>
<td>5</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>239</td>
<td>6</td>
<td>283</td>
</tr>
<tr>
<td>2009</td>
<td>National</td>
<td>243</td>
<td>5</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>237</td>
<td>19</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>241</td>
<td>5</td>
<td>282</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>231</td>
<td>5</td>
<td>285</td>
</tr>
<tr>
<td>2007</td>
<td>National</td>
<td>246</td>
<td>6</td>
<td>284</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>241</td>
<td>21</td>
<td>279</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>243</td>
<td>5</td>
<td>279</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>236</td>
<td>4</td>
<td>282</td>
</tr>
<tr>
<td>2005</td>
<td>National</td>
<td>244</td>
<td>5</td>
<td>281</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>241</td>
<td>20</td>
<td>275</td>
</tr>
<tr>
<td></td>
<td>Florida</td>
<td>243</td>
<td>5</td>
<td>276</td>
</tr>
<tr>
<td></td>
<td>New York</td>
<td>237</td>
<td>4</td>
<td>282</td>
</tr>
</tbody>
</table>

† Reporting standards not met.
# Rounds to zero.

NOTE: The NAEP Mathematics scale ranges from 0 to 500. Detail may not sum to totals because of rounding. Some apparent differences between estimates may not be statistically significant.

**NOTE TO Reporting and Dissemination Committee on Embargoed Pre-Release Access to NAEP Reports**

The guidelines attached were endorsed by the Reporting and Dissemination Committee in August 2011 as administrative procedures for handling news media requests for embargoed access to NAEP reports to help prepare accurate news stories before the time set for an official release. Recently, application of the guidelines by Governing Board staff has been criticized by an online education news site in New York City, which was denied pre-release access to the 2013 NAEP Reading and Mathematics Report Cards. The Committee will discuss the guidelines and their application at this meeting. It may also discuss whether to prepare guidelines for pre-release access by other persons or organizations that wish to comment on NAEP reports or provide summaries or interpretations for their members or the public.

In addition to the guidelines, this tab includes the report on the Committee discussion in August 2011, information on the recent issue involving the Board’s embargo policy, and background materials on embargo practices.

The guidelines pertain only to embargoed pre-release access to NAEP materials by news media personnel and provide for equal treatment of all news organizations, regardless of how their news product is disseminated, whether published, broadcast, or posted on the Internet. Recipients must agree not to make any information public until the time set by the Board for public release.

The guidelines do not apply to education constituency groups, such as a teachers union or school board association, advocacy groups with varying views on education issues, or non-profit think tanks that offer commentary and analysis. At present such groups are not given pre-release access to NAEP reports on the grounds that doing so would, in effect, constitute a general public release because, as a government program, the National Assessment should not pick favorites among them.

Pre-release access to NAEP reports has been given on an embargoed basis to public officials, their staffs, and the organizations representing them that are involved in authorizing, funding, or facilitating the National Assessment. These have included members of Congress, governors, the superintendents of states and urban districts participating in NAEP, senior officials of the U.S. Education Department and the White House. Pre-release briefings have been given to three organizations—the Council of Chief State School Officers (CCSSO) and the National Governors’ Association (NGA) for state and national NAEP reports and the Council of the Great City Schools for reports on the Trial Urban District Assessment (TUDA). Embargoed reports are also given to members of the Governing Board and occasionally, upon request, to former members. In addition, pre-release data has been given to NAEP advisory committees and to persons directly involved in preparing the assessments, such as subject-matter experts.
National Assessment Governing Board

News Media Embargo Guidelines

INTRODUCTION

Under law, the National Assessment Governing Board has the responsibility to “plan and execute the initial public release of National Assessment of Educational Progress (NAEP) reports.” The NAEP authorizing statute continues that NAEP data “shall not be released prior to the release of [such] reports.”

As part of pre-release activities, information is provided to the media in order to facilitate news coverage that reaches the general public. The practice for many years has been to grant access to confidential information to media representatives who have signed an embargo agreement, promising not to print or broadcast news of a report before the scheduled time of release. With the rapid evolution of the media industry bringing new and influential voices through the Internet, more requests for embargoed access are being received from those outside traditional print and broadcast news organizations.

In order for staff to make fair decisions about who should receive embargoed access, objective guidelines are needed. These guidelines establish the criteria and procedures to be used.

FUNCTION AND BENEFIT OF NEWS MEDIA EMBARGOES

Under a longstanding tradition, organizations that release news and research findings to the public have used embargoes as a way to give reporters advance access to the information while retaining control of the timing and nature of their releases. Government officials and agencies, scientific and medical journals, corporate and consumer businesses, and financial institutions often use embargoes, particularly for lengthy or complex information that requires time for thorough review and analysis before news stories are completed.

Embargo agreements can be beneficial to the releasing organization, journalists, and the public that reads the news and can lead to broad-based dissemination and fuller coverage. Embargoed access may achieve the following:

Give reporters the time to read and analyze reports, to do further research on complex information, to conduct interviews, and to write more complete, nuanced stories before the time set for release. This reduces the chances that a reporter will “dash off” a story quickly and as a result make errors in interpreting data.
Permit news organizations to print or broadcast a story or place it on the Internet as soon as an embargo is lifted, promptly spreading news of the report or research findings to their audiences.

Create interest and buy-in among journalists who are granted access, which may increase coverage. The additional time provided before stories must be written may help journalists appreciate the significance of the information and how newsworthy it is.

**RISKS OF EMBARGOES**

Embargo breaks may be committed by a news organization or individual seeking to scoop the competition, or they may happen through accident or carelessness.

For most media outlets and individual reporters, the risks of damaging a relationship with a source or attracting negative attention heavily outweigh the possible benefits of violating an embargo agreement. Such cases do happen, but they are rare.

While journalists do not take a formal oath, and need no license, journalistic ethics demand that embargoes—once agreed to—be respected. If a journalist working outside of the traditional media practices ethical journalism, he or she will not knowingly break an embargo.

**CRITERIA FOR ACCESS**

A requestor must meet one of the criteria below in order to receive embargoed access to NAEP reports:

1) **The requestor is an editor, reporter, columnist, or blogger affiliated with a print, broadcast, or online news organization.**

*Print and broadcast news organizations for which qualifying employees may receive access would include newspapers, magazines, news services, and radio and television news outlets. Some examples: Associated Press, the Bozeman Daily Chronicle, the New York Times, MSNBC, Fox 5 NY, the New Yorker, National Review, the Nation, WTOP, Education Week.*

*Examples of online general-interest news organizations that would receive access:* Huffington Post, Daily Kos, the Texas Tribune, the Daily Caller.

*Examples of print and online education trade publications and news providers that would receive access:* Education Daily, Hechinger Report of Columbia University’s Hechinger Institute for Education Journalism, Alexander Russo’s This Week in Education, Inking and Thinking on Education by Joanne Jacobs.
2) The requestor is a freelance reporter working on a story for a news organization in one of the categories above.

Requestors may be asked to provide documentation of their employment or freelance assignment.

PROCEDURE FOR REQUESTS

Information about the requirements for embargoed access to NAEP reports and embargo agreement forms shall be made available to news media prior to NAEP releases.

A separate agreement form must be signed by each person receiving embargoed information before each release.

DENIAL OF ACCESS

Reporters shall be denied embargoed access to NAEP information if they are not in one of the categories above or refuse to sign the embargo agreement. Those who knowingly break the embargo shall not be granted embargoed access to subsequent NAEP reports for up to two years.

Appeals regarding denial of access shall be determined by the Commissioner of Education Statistics in consultation with the Executive Director of the Governing Board.
4. Embargo Guidelines for NAEP Releases

The Committee reviewed a set of guidelines, prepared by Governing Board staff, on granting embargoed access to NAEP reports to members of the news media prior to the public release. The guidelines were developed in consultation with NCES, and incorporate Committee views on a previous draft that was discussed at the meeting in May 2011.

Mr. Harris said the guidelines were prompted by major changes in the news business during the past few years with the decline of newspapers, magazines, and television and radio news and the rise of news websites and bloggers on the Internet. The key principle behind the guidelines is that it is in NAEP’s interest to grant advance access to its reports to give journalists the time to write better, fuller, more accurate stories. At some point the Board may want to grant advance access to stakeholder organizations, such as teacher unions, or advocacy groups and think-tanks. But it is very difficult to do that without being accused of bias or, on the other hand, giving reports to everyone that requests them, and not having a scheduled release at all.

The proposed embargo guidelines for news media make no distinction among the vehicles used for transmitting the news. They apply equally to print, broadcast, and online media. But they provide that embargoed pre-release access will only be granted to an editor, reporter, columnist, or blogger affiliated with a news organization. The second category that would receive access is freelance reporters working for a news organization, who may be asked to provide documentation of their assignment.

Several Committee members asked how staff would define a news organization as opposed to an advocacy group. Larry Feinberg, of the Board staff, acknowledged that the lines were not always clear but said that a publication or blog that is an offshoot of a policy group or constituency organization, such as the professional association of mathematics teachers, would not be regarded as a news organization.

Gov. Sonny Purdue said staff should be very conservative in defining a news organization to avoid having interest groups shape the initial coverage of NAEP results and the conversation around their release. NCES Commissioner Jack Buckley said he felt it is important to be cautious in order to avoid spreading data widely before an official release.

The Committee endorsed the embargo guidelines for implementation by the Executive Director and Board staff as part of the initial public release of NAEP reports. These will be administrative guidelines, not an official policy, and do not require action by the full Governing Board.
CONFIDENTIALITY AGREEMENT:

The Nation’s Report Card:
2013 Mathematics and Reading

Under this agreement, you will have access to secure National Center for Education Statistics data that you agree to keep confidential as outlined below.

The data from the report 2013 Mathematics and Reading and statements, commentary, and other materials on the data may not in any way be made public—including print or Internet publication, wire, or broadcast—prior to Thursday, November 7, at 10 a.m. EST.

Neither the data nor the access information provided enabling you to view embargoed data and related materials online is to be shared with other individuals or organizations, including on a “hold for release” basis. All conversations about the embargoed data and related materials will be limited to those within your organization who need to be informed for essential work purposes only.

By signing this agreement, you are agreeing that you and your organization will abide by the terms above. Please sign below and send to Shannon Tucker via fax at (703) 299-2424 or email at stucker@reingold.com.

Accepted by:

Signature: __________________________

Print Name: __________________________

Email: ________________________________

Organization: ________________________

Date: ________________________________
The following were denied embargoed pre-release access to materials on the 2013 NAEP Reading and Mathematics Report Cards that were released on November 7, 2013:

1. Mary Tillotson, watchdog.org
2. Kim Greene, Scholastic Instructor and Administrator Magazines
3. Joy Pullman, School Reform News
4. Matt Freidman, Scholastic Math and Science Magazines
5. Tara Welty, Scholastic Instructor Magazines
6. Allison Aubuchon, Foundation for Excellence in Education
8. Matt Korobkin, Rodel Foundation
9. Ashley Inman, Education Next
10. Dolly Sullivan, Educate Maine
11. Philissa Cramer, Gotham Schools
12. Kate Schimel, Ed News Colorado
The federal agency that administers the National Assessment of Educational Progress is standing by its decision to exclude a group of online news organizations from early, embargoed access to test results this week.

The National Assessment Governing Board decided to bar Chalkbeat, Gotham Schools, and Ed News Colorado from the early access provided other news organizations to data and a background briefing on 4th and 8th grade mathematics and reading results.

"Right now reporters being briefed on new #NAEP scores. We're not b/c @GovBoard has inexplicably barred @gothamschools & @Chalkbeat," said a Twitter posting, or tweet, from GothamSchools on Wednesday.

The three sites (GothamSchools, Chalkbeat, and Ed News Colorado) are part of the same organization and are being re-branded under the Chalkbeat banner. They actually learned Monday in an email from a NAGB contractor that they would not be given the embargoed access.

"While the Board understands that these groups may have received access in the past, their current relationship with the Colorado Nonprofit Development Center is considered to be an affiliation with an outside organization and therefore outside of the embargo access policy," the contractor, Reingold Inc., said in the email. "This determination is based the Governing Board's understanding of the current relationship with the Colorado Nonprofit Development Center with the understanding that Chalkbeat may eventually be an independently funded news outlet."

Elizabeth Green, the editor of the Chalkbeat sites, said in an interview that the Colorado nonprofit center acts solely as the "fiscal sponsor" of the Chalkbeat sites because they have yet to receive their own nonprofit status from the Internal Revenue Service. The center provides back-office services such as human resources, she said.

"Its sole purpose is to incubate non-profit groups," Green said. "It's really not dissimilar from contracting with an outside HR or back-office services firm."

The Chalkbeat sites applied for the same embargoed access given many other print, broadcast, and online news outlets. That includes up to 48 hours advance access to the NAEP results, and a background briefing with officials the day before the expiration of the embargo allows news outlets to publish their stories. For this week's NAEP results, the embargo expired at 10 a.m. Eastern time on Thursday.

"We protested the decision, and I still don't understand their rationale," Green said. "I think it is important to make a distinction between who is an independent news organization and who isn't. We are. We follow professional standards just like any other news organization."

Cornelia Orr, the executive director of the National Assessment Governing Board, said in an interview that the agency adopted an informal policy about three years ago regarding access to its embargoed data. The policy requires that news organizations be independent, she said.

"We developed a policy that seems to have come right up against Chalkbeat," Orr said. "I'm sympathetic to their being caught here and not having independent status, but that's what it was."

Orr said NAGB has received requests from advocacy groups seeking the same embargoed access that independent news organizations receive. She cited the Education Trust as one example. Such groups are excluded under the
policy. (After some reflection, Orr declined to provide a copy of the informal policy, saying it was an internal document.)

Many federal agencies in Washington have variations of early access to embargoed materials, such as unemployment reports in the U.S. Department of Labor or crop forecasts in the U.S. Department of Agriculture. The pre-publication access is meant to help reporters digest complex data.

"The governing board's first and primary responsibility is to protect the NAEP data and to get the cleanest story about the NAEP data out there," Orr said.

Meanwhile, Orr noted, there have been fast-paced changes in the media in recent years, including many new online news outlets. Among the Web news organizations that did get the embargoed access this week were the Huffington Post and Stateline.org, she said.

Orr also said that in Chalkbeat's case for this NAEP release, there was a fair amount of "11th hour" back and forth before NAGB decided to exclude its sites.

"We didn't want to feel pressured at the end and let Chalkbeat in and not others who had applied earlier," she said. [At least one other online news site, EdSource Today, which focuses on California schools, said late Thursday on a listserv for the Education Writers Association that it had sought access to the embargoed NAEP materials and was turned down. I didn't have the chance to ask Orr about the particulars of that.]

Chalkbeat's Green said her organization got hold of the NAEP data through other sources on Wednesday and prepared stories that were ready to publish as soon as the embargo ended on Thursday morning. (She didn't identify the sources, but said Chalkbeat agreed to observe the embargo time.)

Orr noticed that the Chalkbeat stories had benefited from some alternative early access.

"An embargo is an embargo, and I appreciate that they abided by it," Orr said. "I don't want to come off as unsympathetic. I appreciate their frustration. We will continue to review our policy."
'Non-traditional' journalists barred from viewing tax-funded test results early

By Mary C. Tillotson | Watchdog.org

A taxpayer-funded research board has refused to share a national study of American education with Watchdog.org and other "non-traditional" news outlets.

Instead, the National Assessment Governing Board is offering early access to the report Tuesday and invitations to discuss it with the authors during a teleconference Wednesday to select media outlets only, including the Huffington Post.

The Nation's Report Card: 2013 Mathematics and Reading, Grades 4 and 8, will be available to the public (and those "non-traditional" news outlets) Thursday.

"In a world where we are all bombarded with news 24/7, to discriminate between one form [of journalism] goes against our basic Constitution," said Ginger Stanley, executive director of the Virginia Press Association.

Sharon Tucker, communications associate for Reingold, Inc., a private corporation apparently hired to oversee media relations, explained the slight in an email to Watchdog.

"Governing Board policy only permits embargo access for reporters affiliated with traditional news media outlets (e.g. The Washington Post, Chicago Tribune, National Geographic, Huffington Post) that are unassociated with outside organizations (i.e. nonprofit associations, government agencies, academic institutions, for-profit businesses)," the email said.

Tucker's note offered no other explanation, and she didn’t reply to an email or return repeated phone calls, though a receptionist said Tucker was in the office.

"Some of the most important journalism in the last decade has been done by nonprofit journalists, led by Watchdog.org," said Watchdog’s senior content editor, Mark Lisheron. "Pro Publica and Inside Climate News, both non-traditional by your definition, have won very traditional Pulitzer Prizes."
Watchdog.org reporters carry state-approved media credentials in more than half of the country’s state capitols.

"Watchdog.org vigorously objects to being denied a preview of a taxpayer-funded national study by an agency clinging to a barrier in the world of journalism as useless as the Berlin Wall," Lisheron said. "We’ve earned our place at the press table for our public service."

The National Assessment Governing Board doesn’t have the resources to handle media requests from every outlet and had to draw the line somewhere, said Stephaan Harris, public affairs specialist for the board.

"That’s the board’s policy," Harris said. "In part, it was a feeling that if you allowed, say, any kind of group or association no matter how small that had any kind of online arm, you’d be opening the floodgates for people claiming themselves as journalists. It’s a way to minimize this and manage our embargo process."

The board considered using the number of website hits to determine whether a news agency was qualified to receive early access, but thought it wouldn't be fair to those agencies just below that threshold, Harris said.

"There’s really no fair way," he said. "If we say, ‘Your website has a minimum of 100,000 original visits, then we can consider you a news organization,’ then people who get just under that would cry foul."

Discriminating against certain news agencies violates the First Amendment, Stanley said, as does allowing embargoed access at all.

"There should be no distinction between a journalist and a citizen," she said. "All of our laws are citizen laws. Freedom of the press and the First Amendment — the laws of our country, whether they’re state laws or federal laws, are citizen laws, and should create an open government for all."

As for scrapping the embargoed early-access policy entirely, Harris said he could bring the idea to the board meeting in December.

The National Assessment of Educational Progress began in 1969 and is under No Child Left Behind. States only qualify for federal funding if they administer the NAEP math and reading tests every two years to their fourth and eighth graders.

The governing board is selected by the U.S. Department of Education, but is officially independent of the department.

Contact Mary C. Tillotson at mtillotson@watchdog.org.
Here's a phrase you don't hear much: newsroom expansion.

That's the goal of a new non-profit news outlet debuting Monday that is gearing up to cover education in-depth in four states, in the process providing an alternative model for local journalism about schools, education policy and education politics.

Created by a handful of refugees from beleaguered — and in a few cases shuttered — print newspapers, the online-only Chalkbeat springs from the unlikely partnership created last January when the New York-based non-profit news site GothamSchools merged with Denver-based EdNews Colorado. Mostly foundation-funded, it gets about one-fifth of its revenue from local sponsorships and job ads for teachers and administrators.

On Monday it's expanding to two more cities with fraught school politics: Indianapolis and Memphis. The network plans to add others as funding from local philanthropists comes calling; it already plans to hire a reporter to cover Nashville schools.

In each bureau, Chalkbeat plans to cover the state legislature and state board of education, as well as the day-to-day developments of schools and districts. They're also demanding that local philanthropy help cover costs as a "public good," much as it would support an art museum or symphony. The network also wants to scale back the role of philanthropy, making each bureau more self-sustaining as it grows.

The outlet's expansion is encouraging news, said Mark Jurkowitz, associate director of the Pew Research Center's Journalism Project. "I think this is further evidence of the growth potential of the non-profit news sector," he said.

A Pew survey last June uncovered 172 non-profit news outlets, most of them tiny startups. What Chalkbeat could represent, Jurkowitz said, is the next step in their evolution as a non-profit essentially franchises its news-gathering model in different cities. "Clearly here is a place where there is a perceived need for coverage of local schools and local school systems that may not be covered as well in the legacy publications," he said.

Media critic Jeff Jarvis, director of the City University of New York's Tow-Knight Center for Entrepreneurial Journalism, sees the effort as a way of "getting back some of the reporting beats that we have lost" as traditional newsrooms shrink. "We know those beats can be businesses now."
With projected revenues this year of about $2.3 million, Chalkbeat is already bigger than most non-profit news outlets, though a few such as ProPublica and The Texas Tribune are quite a bit larger. The Pew survey found that of 77 digital non-profit news outlets willing to disclose revenue, only 14 reported incomes of more than $1 million. Of the 93 willing to reveal staffing levels, most said they had no more than five paid full-time staffers.

When its Indiana and Tennessee bureaus are fully staffed early next year, Chalkbeat will have 22 full-time employees.

That stands in stark contrast to recent trends in newspaper hiring. Though no firm figures exist on the drop in education reporting positions, the American Society of News Editors' most recent annual "newsroom census" found that for the first time since 1978, the overall number of full-time editorial jobs dropped below 40,000. In 2012, newspapers employed about 38,000 reporters, editors and other journalists, nearly one-third fewer than 2000. Just last year, they cut an estimated 2,600 editorial jobs.

Chalkbeat's expansion represents a quiet triumph for its founders, among them the editor and publisher, respectively: Gotham's Elizabeth Green and EdNews' Alan Gottlieb, two journalists bent on social justice and fascinated by education's role in making cities work. Both covered education at big-city newspapers and both have seen their beats slashed by downsizing.

Editor Elizabeth Green, 29, center, talks with Emma Sokoloff-Rubin, 24, from Manhattan, left, and Sarah Darville, 22, from Brooklyn, at Chalkbeat. (Photo: Jennifer S. Altman for USA TODAY)

Green, 29, began her journalism career as a student at Montgomery Blair High School, a top-flight school in Silver Spring, Md., where one day in 2000, she recalled, the principal announced over the loudspeaker, "You black and Hispanic students need to get your test scores up!" It was the first time she realized that her school had an achievement gap.

Green began spending her lunch hour interviewing classmates "on the other end of the cafeteria" and wrote up her findings in the school newspaper. Three years later, studying at Harvard, she wrote a nearly 5,000-word expose in Fifteen Minutes, the weekend magazine of The Harvard Crimson, that took aim at the hollowness of the university's "oft-touted commitment to diversity."

That piece, plus her unabashed wonkishness — Green wrote her senior thesis on Alabama Gov. Bob Riley's failed campaign to make the state's tax code more progressive — landed her a job at U.S. News and World Report. In 2007, she began
covering city schools for *The New York Sun*, but the paper imploded in 2008. She and partner Philissa Cramer founded GothamSchools later that year as part of an existing non-profit, OpenPlans.

Gottlieb, 57, came to Denver in 1988 to work at *The Denver Post*, where he covered the city's schools, which at the time were operating under court-ordered busing. Gottlieb became so engrossed in issues of school quality, funding and racial segregation that he finally had to quit the *Post* to write about the issues full time at a local foundation. "He soon moved to EdNews Colorado, founded in January 2008 as a daily blog on education policy in the state legislature.

As with many online news enterprises, Chalkbeat is bristling with veterans: Its Colorado capital editor is Todd Engdahl, a former *Denver Post* city editor who had hired Gottlieb in 1988. Engdahl lost his job during *Post* layoffs in 2007. Chalkbeat's Indianapolis bureau chief is Scott Elliott, who's leaving newspapers after 22 years, the last three at the *Indianapolis Star*.

But among the newsroom refugees are a few who may never know what it's like to complain about the dying news business: One of Chalkbeat Colorado's newest hires is Kate Schimel, a 23-year-old reporter who was a one-time intern. "She's never been in a traditional newsroom," Gottlieb said.
News embargo
From Wikipedia, the free encyclopedia

In journalism and public relations, a **news embargo** or **press embargo** is a request by a source that the information or news provided by that source not be published until a certain date or certain conditions have been met. The understanding is that if the embargo is broken by reporting before then, the source will retaliate by restricting access to further information by that journalist or his publication, giving them a long-term disadvantage relative to more cooperative outlets. They are often used by businesses making a product announcement, by medical journals, and by government officials announcing policy initiatives; the media is given advance knowledge of details being held secret so that reports can be prepared to coincide with the announcement date and yet still meet press time. In theory, press embargoes reduce inaccuracy in the reporting of breaking stories by reducing the incentive for journalists to cut corners in hopes of "scooping" the competition.

Embargoes are usually arranged in advance as "gentlemen's agreements." However, sometimes publicists will send embargoed press releases to newsrooms unsolicited in hopes that they will respect the embargo date without having first agreed to do so — the phrase "For Immediate Release" often found at the top of press releases indicates that the information in the release is *not* embargoed.

News organizations sometimes break embargoes and report information before the embargo expires, either accidentally (due to miscommunication in the newsroom) or intentionally (to get the jump on their competitors). Breaking an embargo is typically considered a serious breach of trust and can result in the source barring the offending news outlet from receiving advance information for a long period of time.

News embargoes are one of several ways a source can influence media presentation of the information they provide; others include providing information "on background" or "not for attribution," limiting or providing "access," or even direct government or market intervention against the reporters or media company. (See confidentiality terminology in journalism for a full discussion of these.) The manner in which journalists react to these and other attempts to influence coverage are a matter of journalistic ethics.

**Examples of embargoes**

- Biweekly press briefings from the International Monetary Fund are typically embargoed until 10:30 a.m. Washington time, 1430 GMT (for synchronised effect on global stock markets).
- Reporters who accompanied U.S. President George W. Bush on a Thanksgiving visit to Iraq in 2003 were embargoed from filing until the President left the country. They were told that, in the interests of security, the trip would be canceled if news broke before its conclusion.¹
- The Ministry of Defence in the United Kingdom informed a handful of journalism outlets that Prince Harry would be serving in Afghanistan, on condition that the information not be released until the end of his deployment. The information was leaked after about two months, and officials agreed to end the embargo. The prince was immediately removed from the battlefield, reportedly for his safety and that of his fellow soldiers.
- In Canada, Australia and other countries, prior to the release of the budget and other important government announcements, reporters are held in a "lockup" so that they can prepare stories in advance. They are not permitted to file until after the official announcement (for example, after the Minister of Finance rises to deliver the budget speech.) Lockups are particularly aimed at
preventing insider trading on the basis of leaked government announcements.\[2\][3] A similar lockup is done in the United States when the Federal Reserve Board is preparing to adjust an interest rate.

\textit{The New York Times} in 2008 prompted suppression of the story of the kidnapping of David Rohde (their reporter) in news outlets and on Wikipedia until his return in 2009. This example, in which the instigator of the embargo is not the source, may be a case of self-censorship instead.

**Embargoes on articles in scientific journals**

News embargoes are commonly applied on information of health-related news regarding upcoming medical journal articles. All major medical journals, including the \textit{New England Journal of Medicine}, the \textit{Journal of the American Medical Association}, and \textit{The Lancet}, have publication embargoes.

The \textit{JAMA} embargo probably dates back to the editorship of Morris Fishbein, from 1924 to 1949, and holds until 15:00 Central Time on the day before the cover date of the issue. Journalists who agree to not publish (in print, on television, on radio, or via Internet) until that time the information contained in a manuscript to be published by the journal receive advance copies of the journal by mail during the week before publication. For selected articles, press releases and news release videos are also prepared by science writers and released to journalists during that week.\[4\]

The reasons given for such embargoes are twofold. First, they enable journalists to produce more comprehensive and accurate coverage, as the embargo provides time in which they can research the background to a story and thus publish "backgrounders" along with the story's release. Second, they enable doctors and scientists to receive and to analyze medical studies before the general public does, enabling them to be better informed when called upon to comment or to react by journalists or by patients. However, some object to the medical news embargo system, claiming that it is driven by profit motives on the parts of the medical journals.\[4\][5]
If you must use embargoes, here’s how to do it right

IVAN ORANSKY(1)

Key words: Embargo; Science journalism; Publishing; Research; Medical news

(1) Founder, Embargo Watch (http://embargowatch.wordpress.com); Adjunct Associate Professor, Arthur L. Carter Journalism Institute, New York University. orans01@nyu.edu.

Embargoes - often used by scientific institutions such as medical societies and scientific journals to give access to reporters before material is published - can inspire heated arguments. Some journalists love them, while others say they - along with Ingelfinger Rule, which prohibits pre-publication publicity of results before they appear in a peer-reviewed journal - discourage original reporting [1]. Journals find them helpful in “choreographing” the dance of medical news [2], but some have eschewed them completely [3].

Despite all of this debate, and the fact that embargoes are becoming “less and less practical” [4], in the words of one press officer, they are here to stay, at least for some time.

And as one public relations executive said recently, 'Every manager here has a different set of rules about embargoes' [5].

With that in mind, here are some guidelines for appropriate embargo policies that actually live up to the oft-stated goals of allowing reporters enough time to report stories accurately, while avoiding needless restrictions on the flow of scientific information.

1. **Give a reasonable amount of time.**
   What’s reasonable? That’s probably a judgment call, dependent on how complicated the material is, what else is happening in the world, and other factors. Many journals that publish weekly provide embargoed material about five days before publication, which seems like enough time. And I’ve suggested that 24 hours should be a minimum, even in our millisecond news cycle world. But one thing’s for sure: 38 minutes is not long enough [6].

2. **Don’t embargo material that’s freely available online.** This might appear obvious. But based on the number of journals and scientific conferences that still try to claim that their online accepted papers, corrected proofs [7], and abstracts are embargoed until some time they determine later, it bears repeating. If it’s freely available online, it can’t be embargoed. End of story.

3. **Give a specific time for your embargoes.** This doesn’t come up very often, but saying that a paper is embargoed for a date isn’t enough - you have to also say what time, and in what time zone. Otherwise it will lift 26 times [8] for people in 26 different time zones.

4. **Don’t ask everyone to agree to an embargo, then let one news outlet go ahead with a story.** I’m as big a fan of original reporting as the next journalist, but I’m not a fan of backroom side deals. If reporters...
have agreed to an embargo on your material, don't then give an exclusive to a paper - say, The New York Times [9] - while making everyone else wait to publish.

5. **Don’t ask for a quid pro quo.** Lots of press officers believe - perhaps with some proof [10] - that embargoes increase the chance something will be covered. But don’t make that coverage a condition of your embargo agreement [11]. Reporters may go to your conference for many reasons, including becoming better-informed about a subject, and never write anything that can be pegged to that conference. Quid pro quo is unseemly, not to mention a journalistic no-no.

6. **Be consistent about sanctions and early embargo lifts.** If someone has agreed to your embargo policy, they should get the same sanctions for breaking it as anyone else does, no matter what outlet they work for. Those sanctions should be clearly spelled out in your embargo policy, and you should avoid the temptation to look the other way for repeated “inadvertent” breaks. And don’t blame someone for breaking an embargo if he never agreed to embargoes in the first place. Also: Lift the embargo once the material appears online, whether it’s an obscure blogger or a major wire service that broke it. If one reason for embargoes is to level the playing field, then keep the playing field level.

7. **Keep the number of cooks in the kitchen to a minimum.** Nowadays, for many journal studies, there are at least two press releases: one from the journal, and one from the researchers’ institution. If the research had an industry sponsor, there may be a third. Conferences can get even more complicated, and that’s where inadvertent breaks can happen. Do your best to minimize those, and confusion.

8. **If other news is coming out within a day or two of yours, move your embargo so they match.** Let’s say you’re publishing a study on a particular subject, and your embargo lifts on Thursday at 5 p.m. Eastern. You find out that a competing journal is publishing a study on the same subject at 5 p.m. Eastern on Tuesday. Move yours to Tuesday, and let your press list know. If one reason for embargoes is to allow reporters to write better-informed stories, why insist that they only cover your news if they want to publish at the embargo time [12]? Be flexible. Readers will thank you.

9. **Make sure recipients of your “embargoed” emails have actually agreed to an embargo.** Sending something and marking it “embargoed” doesn’t mean it actually is [13]. Just because someone agreed to another institution’s embargo policy doesn’t mean that she agreed to yours. There’s nothing stopping her from writing about the story, and she won’t have broken any agreements.

10. **Don’t try to restrict with whom reporters can speak.** As a number of embargo policies spell out, part of the reason to give journalists time with material before it’s published is so that they can seek outside comment. Requiring that reporters not share the material with anyone before the embargo [14] lifts turns them into stenographers [15]. At the very least, it will make people more cynical about the reasons for embargoes.

---

**References**


[5] http://www.dontgetcaught.biz/2012/10/embargoes-or-
anything-goes-10-big-myths.html
Governing Board and Committee Input in NAEP Results

The Reporting and Dissemination Committee is continuing an ongoing discussion of its role in the reporting, release, and dissemination of National Assessment of Educational Progress (NAEP) results. The Committee desires more input at the beginning, or conception, phase of report development, rather than solely providing feedback on a late-stage draft report or web site. The goal is to have input at a “big picture” level rather than provide detailed edits. Also, being mindful of the changing media landscape and the need to make NAEP relevant to diverse audiences, the Committee is exploring the development of additional focused reports on important aspects of NAEP data that can extend the impact of NAEP results beyond their initial release.

The Committee has expressed interest in ensuring that NAGB can impact NAEP reporting while preserving the legal responsibilities of the Governing Board, which sets policy for NAEP, and the National Center for Education Statistics (NCES), which assesses the students, analyzes the data, and uses the findings to prepare NAEP reports. The Governing Board’s NAEP reporting, release, and dissemination policy (in full below), adopted in 2006, was used as a starting point for this discussion.

After the Board meeting in August 2013, Committee Chair Andrés Alonso requested Board and NCES staff to begin collaboration on possible ideas to achieve the Committee’s goals for discussion at the December Board meeting. Important context for this discussion includes the fact that data from NAEP Report Cards are transitioning to being released mostly online through an interactive site as opposed to printed reports. In advance of report public releases, members and staff would see preliminary results through a preview of the interactive site and a brief printed summary. Additionally, the process for allowing earlier and higher-level input is complicated by the six-month reporting window for Report Cards in subjects like mathematics and reading. The ability to see data in a consumable form before a site preview would be a challenge. Preliminary discussion ideas and suggestions are listed below.

- **Pre-Data Discussions:** At the Board’s March meeting, Committee members can start discussion on assessments being undertaken for that calendar year. So for 2014, this would include NAEP Civics, U.S. History, Geography, and Technology and Engineering Literacy. Though testing is yet to be completed and there will be no data at that point, the Committee can discuss what types of data, trends, comparisons, etc., should be included and highlighted on the NAEP report site. These views can inform visioning meetings conducted by NCES and its NAEP contractors where data will be discussed and report structure determined.
• **Singling Out Topics:** Committee members can suggest topics within a subject they think the public might be especially interested in and then the website can highlight that in some way. In U.S. history, for example, if Committee members believe that topics such as the Civil War or the Civil Rights Movement would have a wide appeal, the Nation’s Report Card website can give prominence to the test items, response rates, subgroup performance, etc., for that topic.

• **Guiding Questions:** Committee members can suggest ideas for the main questions on the interactive NAEP website through which performance summaries and charts and tables are structured. The site for NAEP 2013 Reading and Mathematics Report Card, for example, has three big-picture questions such as “Are Students Making Progress?,” designed to tell a story through the findings.

• **Ideas for NAEP Website Graphics:** Committee members can suggest general ideas for potential trends, comparisons, etc., that would be make for a good chart or table. The purpose would be to highlight “hidden gems”, trends or patterns that normally are not covered in the media but to which NAEP should bring more attention.
National Assessment Governing Board
Reporting, Release, and Dissemination of NAEP Results
Policy Statement

Adopted: August 4, 2006

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; 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.
THE NATION’S REPORT CARD
2013 MATHEMATICS AND READING, GRADES 4 AND 8
RELEASE WEBINAR
EVENT DEBRIEF

OVERVIEW
On November 7, 2013, the National Assessment Governing Board coordinated a live webinar to release results of the 2013 National Assessment of Educational Progress (NAEP) Mathematics and Reading, Grades 4 and 8. Panelists included:

- Jack Buckley, Commissioner, National Center for Education Statistics
- Honorable Leticia Van de Putte, Texas State Senator; Member, National Assessment Governing Board
- William D. Waidelich, Executive Director, Association for Middle Level Education
- Cornelia Orr, Executive Director, National Assessment Governing Board (moderator)

The report card results were announced at the webinar, as was the new, interactive Nation’s Report Card website. As a demonstration showed, the new online format makes NAEP data more accessible and visually appealing in many ways, for example, by positioning mathematics and reading results side by side for the first time, and featuring videos that guide visitors on how to easily explore and interpret results. As was also explained, the site organizes important findings and trends through big-picture questions to reveal how students are performing by race and ethnicity, gender, income level, and additional contextual factors.

For the first time, the webinar extended the question and answer session, so the event lasted 90 minutes instead of 75 minutes. Thirty-six questions were submitted before the webinar.

WEBINAR ATTENDANCE
This release saw a record number of RSVPs, with 741 registrants.

- In attendance: 433
- Unique organizations represented: 298 (internal staff and contractors excluded)
- State departments of education had a strong turnout; more than one-third of attendees were from public schools or education agencies
- Education organizations attended in high numbers; 1 in 5 attendees were from such groups
- First-time attendees comprised 61 percent
A post-webinar survey drew responses from 118 attendees, in which they were asked questions about their satisfaction with the event and suggestions for future events. The responses were overwhelmingly positive:

- 95 percent found the information from the event “very relevant” or “relevant” to their work
- 83 percent of respondents said they would be willing to help spread the word for future release events
- Webinar event attendees commented favorably on both the webinar and the new online report format. Comments included:
  - “Very helpful - especially walk through how to use website resources.” — staff member from Massachusetts Business Alliance for Education
  - “Glad to see the Nation’s Report Card website updated. Looks like it will be aesthetic and easy to use!” — staff member from Mississippi State University
  - “This was a very helpful web event as it walked through the available information and means to access it, as well as offering a variety of perspectives and opportunity for Q&A.” — staff member from North Dakota Department of Public Instruction

For the first time, the Board is conducting user surveys with media and congressional staff to gather feedback on the usability and accessibility of the new online Nation’s Report Card format. Survey results will be presented to committee members at the Board meeting.

**Traditional Media Coverage**

A full overview of release media coverage will be presented to committee members at the Board meeting. Initial highlights include:

*Embargo Access Requests and Media Conference Call*

- A record-breaking 164 reporters registered for access to embargoed report card data.
- 55 reporters attended the embargoed media conference call on November 6.
- 42 of the news outlets on the call published articles.
- There was one media embargo break, by La Opinion reporter Maria Pena, who posted about the results to her personal Twitter account. She removed the post when asked to.

U.S. Secretary of Education Arne Duncan hosted a separate national conference call with journalists prior to the release date, which further promoted NAEP data.

**Media Coverage**

Within 24 hours of the release event, 15 national news outlets published 21 original stories about the math and reading report in print or online.

**National Coverage**

- Associated Press
- Atlantic
- Bloomberg
- Education Week (two stories, and two blog posts)
- The Educated Reporter (Education Writers Association blog)
- The Huffington Post
- Mother Jones
- The New York Times
- Politico Morning Education (two stories)
- Stateline (Pew Charitable Trusts news service)
- U.S. News & World Report (two stories)
- USA Today
- The Wall Street Journal
- The Washington Post (three stories)

**Local Coverage**
- 173 local news organizations published original stories about the report in print or online.
- There were 360 local broadcasts about the release on television and radio.

**Wire Distribution**
In addition to the original news stories developed by the Associated Press, three wire services distributed news of the results:
- AP Top Headline
- Federal News Service
- UWire, a university newswire

The news release was posted by 216 websites within 48 hours of the release event.
Eighteen news outlets used infographics citing NCES data, including the Huffington Post and Washington Post.

**SOCIAL MEDIA COVERAGE**
This release saw a burst of social media activity about NAEP and the Governing Board larger than any previous release, including:
- Approximately 1,300 conversations about the report during the November 7 webinar.
- More than 4,200 conversations on November 7, totaling more than 10,000,000 potential impressions.

Ninety percent of the online conversation took place on Twitter:
- NAEP was the top trending topic in the DC-area on Twitter during the webinar event
- Governing Board messages were retweeted 25 times on November 7, reaching a potential 333,426 people.
- In total, 18 stakeholders promoted the release event on Twitter
- The most influential tweets (based on reach and engagement) came from the Department of Education, NAEP/NCES, the Governing Board, Education Week, and Democrats for Education Reform.

The headlines below link directly to the respective articles online. Should an outlet change the article URL and the link become broken, we are happy to provide the full text of the article.

National Outlets

**Not good enough: Math, reading scores up slightly**
*Associated Press*—Kimberly Hefling
Published November 7, 2013

**American Math and Reading Skills Are Slowly Getting Better**
*The Atlantic*—Julia Ryan
Published November 7, 2013, 10:36 a.m.

**U.S. Schoolchildren Show Gains in Nation's Report Card Tests**
*Bloomberg*—Oliver Staley
Published November 7, 2013, 10 a.m.

**How States Present--and Spin--NAEP Scores for the Public**
*Education Week*—Andrew Ujifusa
Published November 7, 2013, 3:56 p.m.

**U.S. Math, Reading Achievement Edges Up, But Gaps Remain**
*Education Week*—Catherine Gewertz (subscription required; [full story can be read on Hechinger Report](#))
Published November 7, 2013

**Arne Duncan 'Encouraged' By NAEP Results**
*Politics K-12 (Education Week blog)*—Alyson Klein
Published November 7, 2013, 10 a.m.
NAEP Score Boosts: Was It the Teaching?
*Teacher Beat (Education Week blog)—Stephen Sawchuk*
Published November 7, 2013, 4:57 p.m.

The Nation's Report Card: A Slow Climb Up a Steep Hill
*The Educated Reporter (Education Writers Association blog)—Emily Richmond*
Published November 7, 2013

National Test Scores Show Slight Math, Reading Increases for American Students
*The Huffington Post—Joy Resmovits*
Published November 7, 2013, 10 a.m; updated 4:40 p.m.

New NAEP Scores Show Continued Improvement in American Schools
*Mother Jones—Kevin Drum*
Published November 7, 2013, 12:32 p.m.

U.S. Reading and Math Scores Show Slight Gains
*The New York Times—Motoko Rich*
Published November 7, 2013

It's time to fight - Education world locked in schoolyard brawl - Louisiana voucher program analysis out
*Politico Morning Education (10th item)—Caitlin Emma*
Published November 8, 2013, 10:01 a.m.

National Report Card Day - Watch for Louisiana voucher data today - Education Department gets an earful - Community colleges get a boost
*Politico Morning Education—Libby A. Nelson*
Published November 7, 2013, 10:02 a.m.

Reading, Math Scores Inch Up
*Stateline.org (Pew Charitable Trusts news service)—Adrienne Lu*
Published November 7, 2013

Education Reform May Leave High-Performing Students Behind
*U.S. News & World Report—Allie Bidwell*
Published November 8, 2013

Racial Achievement Gaps Remain Largely Unchanged, Despite Higher Test Scores
*U.S. News & World Report—Allie Bidwell*
Published November 7, 2013

Tennessee and D.C. lead education reform: Column
*USA Today—Richard Whitmire*
Published November 7, 2013, 4:52 p.m.
U.S. Students Make Slight Progress on Test Scores
The Wall Street Journal—Stephanie Banchero
Published November 7, 2013; updated 7:41 p.m.

U.S. students show incremental progress on national test
The Washington Post—Lindsey Layton
Published November 7, 2013

D.C. posts significant gains on national test, outpacing nearly every state
The Washington Post—Emma Brown
Published November 7, 2013

Amid testing gains, D.C. students exhibit achievement gaps
The Washington Post—Emma Brown
Published November 8, 2013

Local Outlets

Race gap expands on some Ohio academic tests; scores become fodder for position statements
Akron Beacon Journal—Doug Livingston
Published November 9, 2013, 12:58 a.m.

Report card: State fails to improve
Albuquerque Journal—Mike Bush
Published November 7, 2013; updated November 8, 2013, 12:05 a.m.

Michigan lagging behind on 'Nation's Report Card' in mathematics, elementary reading
All Michigan/MLive.com—Brian Smith
Published November 7, 2013, 12:48 p.m.; updated 1:07 p.m.

State 8th-graders gain on U.S. reading test
Arkansas Democrat-Gazette—Cynthia Howell
Published November 8, 2013

Arkansas 4th- and 8th-grade students score below national average
Arkansas Democrat-Gazette—Christina Huynh
Published November 7, 2013, 4:02 p.m.

New Jersey students outperform other states in NAEP results
Asbury Park Press—Gina Columbus
Published November 7, 2013
US report: NH students score well in math, reading
Associated Press—Rik Stevens
Published November 7, 2013

Tenn. students lead nation in improvement
Associated Press—Lucas L. Johnson II
Published November 7, 2013

Kansas NAEP scores in math, reading steady in 2013, still top national average
Associated Press—No Author Listed
Published November 7, 2013

NAEP results released; Wyoming beats national avg
Associated Press—No Author Listed
Published November 8, 2013

Wis. black students' reading scores rank low in US
Associated Press—No Author Listed
Published November 7, 2013, 8:36 p.m.

Report Card: Nevada students lag in math, reading
Associated Press—Sandra Chereb
Updated November 7, 2013, 3:46 p.m.

South Dakota Student Math, Reading Scores Steady
Associated Press—No Author Listed
Published November 7, 2013, 5:43 p.m.

Fla. math and reading scores rise slightly
Associated Press—No Author Listed
Published November 7, 2013

Glance: 4th and 8th grade math and reading scores
Associated Press—No Author Listed
Published November 7, 2013, 7:03 a.m.

Report card: Oregon's math, reading scores steady
Associated Press—Gosia Wozniacka
Published November 8, 2013, midnight

Duncan praises Hawaii's math, reading test scores
Associated Press—Jennifer Sinco Kelleher
Updated November 7, 2013, 1:50 p.m.
Report: Maine students score well in math, reading
Associated Press—No Author Listed
Published November 7, 2013, 11:45 a.m.

US report card: Missouri’s scores hold steady
Associated Press—No Author Listed
Published November 7, 2013, 11:54 a.m.

Kansas NAEP Scores Flat in 2013
Associated Press—No Author Listed
Published November 7, 2013, 11:47 a.m.

New North Carolina schools report card measures life skills
Associated Press—Emery P. Dalesio
Published November 7, 2013

Dayton, Minn. education commissioner hail progress of black students on national math tests
Associated Press—Patrick Condon
Published November 7, 2013, 2:45 p.m.

Pence: Data shows Indiana students second in gains
Associated Press—No Author Listed
Published November 7, 2013

NJ students score highly on national test
Associated Press—Geoff Mulvhill
Published November 7, 2013

Wash. scores up on national reading, math test
Associated Press—Donna Gordon Blankinship
Published November 7, 2013, 12:36 p.m.

Report: NY’s 4th-, 8th-graders scores up slightly
Associated Press—Carolyn Thompson
Published November 7, 2013; updated 3:21 p.m.

Arizona students still lag nation in math, reading
The Republic—Mary Beth Faller
Published November 7, 2013, 11:32 p.m.

New NAEP scores released: Georgia shows progress
The Atlanta Journal-Constitution—Maureen Downey
Published November 7, 2013, 10:14 a.m.
**Reading and math scores rise in Georgia**
*The Atlanta Journal-Constitution*—Wayne Washington
Published November 7, 2013, 9:59 a.m.

---

**Texas reading scores lag for all students**
*Austin American-Statesman*—Melissa B. Taboada
Published November 7, 2013, 6:09 p.m.

---

**Report shows Wisconsin black students scoring lowest in nation**
*The Badger Herald (Madison, Wis.)*—Joel Witt
Published November 7, 2013 11:30 a.m.

---

**Maryland students show no significant gains on national tests**
*The Baltimore Sun*—Liz Bowie
Published November 7, 2013, 6:32 p.m.

---

**Mass. students score well on national assessment exam**
*The Boston Globe*—Jasper Craven
Published November 7, 2013

---

**National report: Vermont test scores high, but flat**
*Brattleboro Reformer*—Howard Weiss-Tisman
Published November 8, 2013, 3 a.m.

---

**New York students hover near national average in math, reading**
*Capital (Albany, N.Y.)*—Jessica Bakeman
Published November 7, 2013, 11:26 a.m.

---

**In the News: Modest gains seen in NAEP results**
*Catalyst Chicago*—Cassandra West
Published November 8, 2013

---

**W.Va. students improve slightly in reading, math**
*The Charleston Gazette*—Mackenzie Mays
Published November 7, 2013

---

**New N.C. exams paint bleak picture of skills in state, CMS**
*The Charlotte Observer*—Ann Doss Helms
Published November 8, 2013

---

**US 'report card' for 2013: Student achievement creeps upward**
*The Christian Science Monitor*—Amanda Paulson
Published November 7, 2013
State math, reading scores flat; large percentage below basic achievement
Clarion-Ledger—Sam R. Hall
Published November 7, 2013, 12:32 p.m.

Nation's Report Card: Only a Third of 8th Graders Can Read, Compute At Grade Level
CNSNews.com (Cybercast News Service)—Barbara Hollingsworth
Published November 7, 2013, 1:04 p.m.

'Nation's Report Card' releases state data
Columbia Daily Tribune (Mo.)—Tribune Staff
Published November 7, 2013, 2 p.m.

Ohio students' reading, math scores not improving
The Columbus Dispatch—Catherine Candisky
Published November 8, 2013, 6:42 a.m.

2013 Nation's Report Card: Tennessee shows nation's best education gains
The Commercial Appeal—Richard Locker, Zack McMillin, Jane Roberts, Grant Smith
Published November 7, 2013, 9:37 a.m.; updated 7:24 p.m.

N.H. students outperform national average in reading, math
Concord Monitor—Kathleen Ronayne
Published November 7, 2013; published in print on November 8, 2013

Nation’s math and reading scores show progress
Education Matters (Connecticut Post blog)—Linda Conner Lambeck
Published November 7, 2013

Calif. student scores in reading, math climb, but mixed overall
Daily Democrat (Woodland, Calif.)—Barbara Jones
Published November 8, 2013, 12:13 a.m.

State test scores remain flat
Daily Journal (Tupelo, Miss.)—Chris Kieffer
Published November 8, 2013

Texas Hispanic students lag in “Nation’s Report Card”
The Dallas Morning News—Terrence Stutz
Published November 8, 2013

D.C.'s Fourth And Eighth-Graders Outpace Nearly Every State in Nation's Report Card
DCist—Matt Cohen
Published November 7, 2013, 12:15 p.m.
Delaware's scores on national student tests show progress
The News Journal—Matthew Albright
Published November 8, 2013

NAEP scores rise for Colorado 4th-graders, 8th graders lose ground
The Denver Post—No Author Listed
Published November 7, 2013, 4:29 p.m.

Utah schools improving in reading, achievement gap, new report shows
Deseret News—Benjamin Wood
Published November 7, 2013, 12:55 p.m.

U.S. students make small gains in reading and mathematics
Deseret News—Celia R. Baker
Published November 7, 2013, 8:35 a.m.; updated 2:11 p.m.

Michigan math exam scores trail U.S. average for 4th-, 8th-graders
The Detroit News—Jennifer Chambers
Published November 7, 2013, 11:56 p.m.

Nation's report card results show some Iowa education gains, needs
The Gazette—Rod Boshart
Published November 7, 2013, 12:30 p.m.

20 Years Later, U.S. Students Making Big Academic Gains
EdMedia Commons (National Education Writers Association)—Mikhail Zinshteyn
Published November 7, 2013, 11 a.m.

Colorado middle schoolers fall short on national report card
EdNews Colorado—Kate Schimel and Sarah Darville
Published November 7, 2013

Rise & Shine: Colorado reassesses school finance after Amendment 66's defeat
EdNews Colorado—Kate Schimel
Published November 7, 2013

Tennessee education gains top the list
Elizabethton Star—Kayla Carter
Published November 8, 2013, 11 a.m.

Tennessee students lead nation in academic growth
WHBQ—Matt Gerien
Published November 7, 2013, 4:19 p.m.; updated November 14, 2013, 5:55 p.m.
A few states, but not N.Y., see big gains on 'nation's report card'
Gotham Schools—Sarah Darville
Published November 7, 2013, 10 a.m.

'Nation's Report Card': Connecticut Students Strong in Reading, Less So In Math
The Hartford Courant—Kathleen Megan
Published November 7, 2013, 11:54 a.m.

Hawaii's math scores add up to first-ever ranking
HawaiiNewsNow (Kaimuki, Oahu)—Jim Mendoza
Published November 7, 2013, 10:49 p.m.; updated November 8, 2013, 12:14 a.m.

Students improve in math, reading
Hawaii Tribune-Herald—Erin Miller
Published November 8, 2013, 12:05 a.m.; updated 12:06 a.m.

Local schools rank 12th in new NC assessment
Hendersonville Lightning—Bill Moss
Published November 7, 2013

Ohio's Students Aren't Showing Much Progress, According to National Test Results
ideastream—No Author Listed
Published November 7, 2013, 6:06 p.m.

Indiana sees rise in test scores on 'nation's report card'
News and Tribune—Maureen Hayden
Published November 7, 2013

Idaho eighth-graders beat national average in reading proficiency
Idaho Statesman—Bill Roberts
Published November 7, 2013

NJ Students Score Well in 'Nation's Report Card' [AUDIO]
New Jersey 101.5—Dino Flammia
Published November 8, 2013, 5:48 a.m.

Dayton Discussing Student Test Scores
KNSI (Minn.)—No Author Listed
Published November 7, 2013, 4:53 a.m.

Report: Wash. Students Among Top 5 in U.S. Math, Reading
KPIU—Florangela Davis
Published November 7, 2013, 7:01 a.m.
Report: SD 4th graders below average in reading
KSFY—Denise DePaolo
Published November 7, 2013, 7:48 p.m.; updated November 14, 2013, 7:54 p.m.

How Texas Students Scored on the Nation's Report Card
KUHF—Laura Isensee
Published November 7, 2013, 10:11 a.m.

State test scores remain steady
KWSN (Pierre, S.D.)—Mark Brown
Published November 8, 2013, 7:02 a.m.

Nevada students score poorly on Nation's Report Card
Las Vegas Review-Journal—Trevon Milliard
Published November 7, 2013, 7 a.m.; updated 5:28 p.m.

Nevada students register gains on 'nation's report card,' but results remain well below average
Las Vegas Sun—Paul Takahashi
Published November 7, 2013, 7 a.m.

Latinos Show Modest Gains in 'National Report Card,' Performance Gap Remains
The Latino Post—No Author Listed
Published November 13, 2013, 2:09 p.m.

Kansas students score high on 2013 reading, math tests
Basehor Sentinel—Peter Hancock
Published November 7, 2013, 11:12 a.m.

Georgia's math, reading scores grow for 2013; Alabama far below national average
Ledger-Enquirer (Columbus, Ga.)—Adam Carlson
Published November 7, 2013

Kentucky fourth-, eighth-graders 'holding steady' on Nation's Report Card
Lexington Herald-Leader—Jim Warren
Published November 7, 2013

California student scores in reading, math climb, but news isn't all good
Los Angeles Daily News—Barbara Jones
Published November 7, 2013, 9:11 a.m.

Education official: Hawaii proved 'lot of skeptics wrong'
The Maui News—Jennifer Sinco Kelleher
Published November 8, 2013
'Nation's Report Card' shows some progress in reading, math  
McClatchyDC—Renee Schoof  
Published November 7, 2013

NY students not improving in reading, math  
Metro.us (New York edition)—Laura Shin  
Published November 8, 2013

Black students near bottom in nation on benchmark math, reading test  
Journal Sentinel (Milwaukee)—Lydia Mulvany  
Published November 8, 2013

Education commissioner points to achievement-gap progress  
Minnesota Public Radio—Tom Weber  
Published November 7, 2013, 10:48 a.m.

Achievement gap closing some, says Minnesota education commissioner  
MinnPost—Brian Lambert  
Published November 7, 2013

Tennessee Gets Most Improved In Education, Still Far From U.S. Leader  
WPLN (Nashville Public Radio)—Blake Farmer  
Published November 7, 2013

Is State Still Pedal to the Metal on Education Reform?  
Nashville Scene—Andrea Zelinski  
Published November 8, 2013, 8 a.m.

Nation's Report Card shows slight Hispanic gains  
NBC Latino—Suzanne Gamboa  
Published November 7, 2013, 12:15 p.m.

'Nation's Report Card: 'N.H. Students Rank High In Math, Reading Proficiency'  
New Hampshire Public Radio—Michael Brindley  
Published November 7, 2013, 12:45 p.m.

Louisiana ranks low on nation’s report card  
The Advocate (Baton Rouge, La.)—Will Sentell  
Published November 8, 2013

Report: L.a. ranks low in math, reading  
The Advocate (Baton Rouge, La.)—Will Sentell  
Published November 8, 2013
National reading, math scores shoot up
*Newsday (Long Island)*—John Hildebrand
Published November 7, 2013, 10:06 p.m.

NAEP shows Florida fourth-graders perform well in reading
*News-Press*—Ashley Smith
Published November 7, 2013

New York kids' statewide reading, math scores are middling
*New York Daily News*—Ben Chapman
Published November 8, 2013, 12:46 a.m.

60% of NY students not up to par
*New York Post*—Andy Soltis
Published November 7, 2013, 9:59 p.m.

State student performance on national tests remain unchanged
*The News & Observer*—Lynn Bonner
Published November 7, 2013

Scores On Respected National Exam Say NY Students Not Making Much Progress
*NY1*—Lindsey Christ
Published November 7, 2013, 9:52 p.m.

Nebraska, Iowa students among best on national tests, but there's room for improvement
*Omaha World-Herald*—Joe Dejka
Published November 7, 2013, 9:10 a.m.; updated 2:32 p.m.

CA students score among lowest in nation
*Orange County Register*—Elysse James
Published November 7, 2013, 6:40 p.m.

Oregon students' reading and math skill stuck mostly at average levels, national exam results say
*The Oregonian*—Betsy Hammond
Published November 7, 2013, 7:03 a.m.; updated 9:31 p.m.

Florida students improve on math, reading skills
*Orlando Sentinel*—Leslie Postal
Published November 7, 2013, 10:39 a.m.

Texas public school students' reading skills still lag behind national average
*Pegasus News (Medill News Service)*—Bryan Lowry
Published November 8, 2013
Eighth-grade reading scores improve
The Pine Bluff Commercial—John Lyon
Published November 7, 2013, 8:49 p.m.

Pa. 4th, 8th graders score higher than national average on math, reading exams
Pittsburgh Tribune-Review—Megan Harris
Published November 7, 2013, 10:09 a.m.

Maine students score high on national reading, math tests
Portland Press Herald—Noel K. Gallagher
Published November 8, 2013

South Carolina doesn't see significant increases on national exam while nation makes progress
The Post and Courier—Diette Courrege Casey
Published November 8, 2013, 12:01 a.m.

New York students make small gains on 'the nation's report card,' but most remain below standards
The Post-Standard—Paul Riede
Published November 7, 2013, 2:32 p.m.

RI students score above the national average on national test
Providence Journal—Linda Borg
Published November 7, 2013, 10:01 a.m.

R.I. students show improvement in national math, reading tests
Providence Journal—Linda Borg
Published November 7, 2013, 11:30 p.m.

Kostrzewa - RI test scores better, still not good enough to help economy
Providence Journal—John Kostrzewa
Published November 8, 2013, 3:59 p.m.

National test shows small improvement in state math and reading scores
Radio Iowa—Dar Danielson
Published November 7, 2013

N.J. students again rank near top on 'nation's report card'
The Record—Leslie Brody
Published November 7, 2013, 1:15 p.m.

RI Sees Little Improvement on National Testing
Rhode Island Public Radio—Elisabeth Harrison
Published November 7, 2013, 10 a.m.
Report card: Utah's minority students making gains in reading
*The Salt Lake Tribune*—Kristen Moulton
Published November 7, 2013, 9:22 a.m.; updated 9:59 p.m.

California students score at bottom of nation in reading, math
*San Jose Mercury News*—Sharon Noguchi
Published November 7, 2013 10:03 a.m.

California Students Achieve Higher Test Scores On Nation's Report Card
*Santa Clarita News*—Jeanina Joseph
Published November 7, 2013, 12:57 p.m.

'Nation's Report Card' Shows Students' Math, Reading Skills Slowly Improving
*School Library Journal*—Karyn M. Peterson
Published November 7, 2013

Schools Chief Tom Torlakson Announces California Students Make Major Gains in 2013 National Tests
*Sierra Sun Times*—No Author Listed
Published November 8, 2013

Minnesota school achievement gap narrows for some, test shows
*Pioneer Press (St. Paul, Minn.)*—Christopher Magan
Published November 7, 2013, 12:01 a.m.

Hawaii public school test scores above average
*Honolulu Star-Advertiser*—Nanea Kalani
Published November 7, 2013, 1:44 p.m.

New results from ‘Nation’s Report Card’ show slight improvement in reading and math
*The Star-Ledger*—Peggy McGlone
Published November 8, 2013, 6:30 a.m.; updated 6:32 a.m.

Minnesota 4th-graders tie for the top in national math tests
*Star Tribune*—Kim McGuire
Published November 7, 2013, 10:02 p.m.

Charting Florida's Progress on 'The Nation's Report Card'
*StateImpact Florida*—John O'Connor
Published November 7, 2013, 2:34 p.m.

How Indiana Students Fared On The Tests the Whole Country Cares About
*StateImpact Indiana*—Kyle Stokes
Published November 7, 2013, 10:01 a.m.
Florida students achieve small gains in national test
*Tampa Bay Times*—Jeff Solochek
Published November 7, 2013, 10:02 a.m.

Report: Tennessee shows greatest math, reading gains in nation
*The Tennessean*—Joey Garrison
Published November 7, 2013, 12:20 p.m.

Tennessee students show big gains on national exam
*Times Free Press (Chattanooga, Tenn.)*—Kevin Hardy
Published November 7, 2013

Reading, math scores climb in state, but still lag behind national average
*Tulsa World (Okla.)*—Kim Archer
Published November 7, 2013, 3:25 p.m.

D.C. Students Outpace Peers In Math, Reading Gains
*WAMU*—Kavitha Cardoza
Published November 7, 2013

Gov. Haslam lauds school improvements, says teachers are satisfied
*WATE (Knoxville, Tenn.)*—Gene Patterson
Published November 8, 2013, 5:39 p.m.

NAEP tests show Md. student scores still above average
*WBAL*—Katie Lange
Published November 7, 2013, 6:01 p.m.

TN is fastest improving state in reading and math
*WBIR*—WBIR Staff
Published November 7, 2013, 1:40 p.m.

Maryland Students Score Above Average
*WBOC*—Chris Messick
Published November 7, 2013, 7:25 p.m.

New standards bring lower test scores statewide
*WECT (Wilmington, N.C.)*—Kaitlin Stansell
Published November 7, 2013, 10:07 p.m.; updated November 11, 2013, 10:09 p.m.

School test scores released, show results of new curriculum
*WCBD (Chapel Hill, N.C.)*—Justin Quesinberry
Published November 7, 2013, 11:09 a.m.; updated 2:08 p.m.
Ohio Students Show Slight Improvement On National Test
WCBE—Jim Letizia and AP and Ohio Public Radio
Published November 8, 2013, 6:47 a.m.

PA Students Show Slight Progress in New Report Card
WESA—Kevin Gavin
Published November 7, 2013, 5:04 p.m.

Bennett Doesn't Take Credit, But Says Reforms Lead To Test Score Gains
WIBC (Ind.)—Ray Steele
Published November 7, 2013

Kansas test scores unchanged
Voice for Liberty in Wichita—Bob Weeks
Published November 7, 2013

TN leads nation in ed. gains
The Wilson Post—Sabrina Garrett
Published November 8, 2013

Education Expert Says Wisconsin Needs Wider Education Policy Discussion
Wisconsin Public Radio—Amanda Magnus
Published November 7, 2013, 3:06 p.m.

NY students show signs of improvement
WIVB—Colleen Hannon
Published November 8, 2013, 7:06 a.m.

D.C. reading and math scores make major gains versus national average
WJLA—Sam Ford
Published November 7, 2013, 11:05 a.m.

Report shows WI achievement gap worst in nation
WKOW—No Author Listed
Published November 8, 2013, 7:04 a.m.

New York State needs work on reading and math
WNYT—WNYT Staff
Published November 7, 2013, 5:28 p.m.; updated 8:35 p.m.

NAEP Test Score: NJ Near Top in Nation!
Woodbridge Patch—Tom Maras
Published November 8, 2013, 7:53 a.m.
Kentucky students scores on par for reading, slightly below for math
_WPSD Local_—Amanda Roberts
Published November 7, 2013, 9:38 a.m.

_Tennessee Schools See Historic Improvement_
_WREG_—Adam Hammond
Published November 7, 2013, 4:38 p.m.

_State Superintendent responds to 2013 NAEP report_
_WTVA (Jackson, Miss.)_—No Author Listed
Published November 7, 2013, 9:52 a.m.; updated 10:15 a.m.

_DC schools make the grade in test scores_
_WUSA_—Surae Chinn
Published November 7, 2013, 5:29 p.m.

"Nation's Report Card:" Wisconsin has the Widest Achievement Gap
_WUWM (Milwaukee)_—Ann-Elise Henzl
Published November 7, 2013, 4:40 p.m.

_State Officials Hope to Solve Achievement Gap Problem_
_WUWM (Milwaukee)_—Marti Mikkelson
Published November 7, 2013, 1 a.m.

_Virginia 4th graders among nation's best readers_
_WVVA_—Greg Carter
Published November 7, 2013, 4:14 p.m.

_New York Students Show Improvement_
_WNYF_—No Author Listed
Published November 7, 2013, 11:33 a.m.; updated 11:38 a.m.

2013 Nation's Report Card: Wyo. has "room to improve"
_Wyoming Tribune-Eagle_—Aerin Curtis
Published November 7, 2013

(Local Groups Work To Improve Columbus Reading Scores In Face Of Latest Report
_WBNS (Ohio)_—Staff
Published November 8, 2013, 4:50 p.m.; updated 5:53 p.m.

_Gov. Haslam: TN is fastest-improving state in education, according to NAEP results_
_WSMV_—Josh DeVine
Published November 7, 2013, 5:49 p.m.)
The Nation’s Report Card: Trial Urban District Assessment in Mathematics and Reading 2013

The 2013 NAEP Trial Urban District Assessment (TUDA) Mathematics and Reading Report Cards will be released together to the general public during December 2013 in one event, as approved by the Board at the December 2013 meeting. Following a review and approval of the report’s results, the release will be arranged as an online webinar. The release event will include a data presentation by the Commissioner of Education Statistics, with moderation and comments by at least one member of the National Assessment Governing Board and an additional panelist with expertise in education and assessment matters in large city school districts. Full accompanying data will be posted on the Internet at the scheduled time of release.


Results will be compared to those of the nation and to a large-city average that includes public schools located in the urbanized areas of cities with populations of 250,000 or more. Data will be presented for all students by such factors as race/ethnicity, achievement gaps, and eligibility for the National School Lunch Program. Contextual information (i.e., student, teacher, and school survey data) with findings of interest will also be reported. Main findings will be included in a brief report summary, with the majority of trends and findings posted in charts and graphs on the new Nation’s Report Card website.
DATE AND LOCATION

The release event for the media and the public will occur in December 2013. The release date will be determined by the Chair of the Reporting and Dissemination Committee, in accordance with Governing Board policy, following acceptance of the final report.

EVENT FORMAT

- Introductions and opening statement by a National Assessment Governing Board representative
- Data presentation by the Commissioner of Education Statistics
- Comments by at least one Governing Board member
- Comments by at least one expert in the field of education and assessment matters in large-city school districts
- Questions from the webinar audience
- Program will last approximately 75 minutes
- Event will be broadcast live over the Internet, and viewers will be able to submit questions electronically for panelists. An archived version of the webinar, with closed captioning, will be posted on the Governing Board website at www.nagb.org along with other materials such as the press release and panelist statements.

EMBARGOED ACTIVITIES BEFORE RELEASE

In the days preceding the release, the Governing Board and NCES will offer access to embargoed data via a special website to approved U.S. Congressional staff in Washington, DC; approved senior representatives of the National Governors Association and the Council of Chief State School Officers; and appropriate media as defined by the Governing Board’s Embargo Policy. A conference call for journalists who signed embargo agreements will be held to give a brief overview of findings and data and to answer questions from the media.

REPORT RELEASE

The Commissioner of Education Statistics will publicly release the report at the NAEP website—http://nationsreportcard.gov—at the scheduled time of the release event. An online copy of the report, along with data tools, questions, and other resources, will also be available at the time of release on the NAEP site. An interactive version of the release with panelists’ statements, a Governing Board press release, subject frameworks, and related materials will be posted on the Board’s web site at www.nagb.org. The site will also feature links to social networking sites and audio and/or video material related to the event.

ACTIVITIES AFTER THE RELEASE

The Governing Board’s communications contractor, Reingold, will work with Board staff to coordinate a post-event communications effort to extend the life of the results and provide value and relevance to stakeholders with an interest in student achievement and assessment in the nation’s large, urban school districts.
Introductory Note: Contextual Information Framework

The revisions proposed to the Background Information Framework for NAEP are intended to make it conform to the Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting, which the Board adopted in August 2012. The changes were recommended by the Ad Hoc Committee on NAEP Background Information that completed its work in August 2013. The Ad Hoc Committee proposed that the term contextual be used rather than background to avoid any misunderstanding that questionnaires were improperly intrusive. The original framework was adopted in 2003. The Ad Hoc Committee felt that although the framework needs updating, its approach and most details remain sound.

Also attached is an annotated version of the August 2012 resolution, showing the pages in the framework where changes are proposed.

A new foreword explains the key changes.
Contextual Information

Framework

for the

National Assessment of

Educational Progress

National Assessment Governing Board
Adopted August 1, 2003
Revised December 2013 -DRAFT
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
### Table of Contents

**ADD: Foreword – 2013**

Preface........................................................................................................5

Executive Summary....................................................................................7

**Chapter One: Introduction..................................................................12**

  - Nature and Purpose of NAEP..........................................................12
  - Nature and Purpose of Contextual Data.........................................12
  - Development of NAEP Contextual Information Framework........14
  - Requirements of NAEP Statute.......................................................15
  - Purpose and Rationale of Contextual Information Framework........16

**Chapter Two: Priorities and Criteria for Collecting and Reporting**

  **Non-Cognitive Data on NAEP.........................................................17**

  - Priorities for Non-Cognitive Information....................................17
  - Criteria for Selecting Topics and Questions...............................18
  - Criteria for Selecting Data Sources............................................20

**Chapter Three: Topics and Types of Contextual Data........................22**

  - Demographic Reporting Categories.........................................22
  - Socio-economic Status.............................................................23
  - Public Policy Contextual factors..............................................25
  - Subject-Specific Background Data.............................................26

**Chapter Four: Non-cognitive Data Sources and Collection.............28**

  - NAEP Student, Teacher, and School Samples..........................28
  - Other Data Sources.................................................................29
  - NAEP Contextual Information Matrix......................................31

**Chapter Five: Using Contextual Data to Report NAEP Results........33**

  - Use of Descriptive Information in NAEP..................................33
  - Implementing No Child Left Behind........................................33
Disaggregation of NAEP Data.................................................. 34
Minimizing Causal Interpretations........................................... 34

Chapter Six: Using NAEP in Educational Research.........................36
Based in part on papers by David Grissmer, RAND
NAEP’s Limitations and Strengths for Research..........................36
Previous Use of NAEP in Research........................................38
Secondary Analysis Grants and District Samples.........................39

Chapter Seven: Review and Improvement of Non-cognitive Questions......40
Independent Validity Studies..................................................40
Review of Contextual Information Framework..........................41

References..............................................................................42

Appendix A: Governing Board Policy on Collection and Reporting of Background Data by the National Assessment of Educational Progress (2002).........44
ADV: Appendix B: Governing Board Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting (2012)

Acknowledgments...................................................................51
Foreword – December 2013

In October 2011, eight years after adoption of the NAEP Background Information Framework, the National Assessment Governing Board convened an expert panel to study the NAEP contextual questions and recommend possible changes. The six-member group was chaired by Marshall S. Smith, former dean of the Graduate School of Education at Stanford University and a former U.S. Under Secretary of Education. The panel’s report, presented to the Board in March 2012, called the contextual questions “a potentially important but largely underused national resource.” (Smith, et al [2012]. *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*)

The report described the information gathered through background questionnaires as “a rich collection of student, teacher and school responses…that can help in understanding the context for NAEP achievement results and give insights into how to improve them.” But it said over the past decade the questionnaires had been cut back and little used in NAEP reports. It urged NAEP to “restore and improve upon” its practice of the early 1990s by “making much greater use of contextual data, but do so in a more sound and research-supported way.”

With “proper attention,” the expert panel declared, NAEP’s contextual data “could provide rich insights into a wide range of important issues about the nature and quality of American primary and secondary education.”

After gathering public comment, the Governing Board adopted a Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting. The policy, approved in August 2012, was based on recommendations by the expert panel and provided for an important change in emphasis:

- NAEP reporting should make greater use of contextual data in both regular Report Cards and special focused reports.
- The reporting of background data will describe patterns and trends, including the educational experiences of different groups of students. Such information will enrich NAEP reporting, but care should be taken not to suggest causation.
- Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions and their connection to student achievement.
- Modules will be prepared for special studies to provide descriptive information on issues of current policy interest.
• NAEP will include contextual questions from international assessments to obtain direct comparisons of states and TUDA districts with educational practices in other countries.

The Board resolution included a set of implementation guidelines. It also established an ad hoc committee, which reviewed the framework. The committee felt the approach adopted in 2003 and most of the details remain sound, but recommended some updating. The revisions are based largely on the resolution and are incorporated in the text that follows. As NAEP makes the transition from paper-and-pencil to a computer-delivered assessment, the Board hopes the new technology will help make possible the range of topics and flexibility in sampling envisioned a decade ago while limiting the burden on students and schools.

Note on Terminology

The document that follows has been renamed the NAEP Contextual Information Framework. The change was made, from background information framework—the title used in 2003—to avoid any misunderstanding that the information provided and the questionnaires from which it derives are overly intrusive or constitute a “background investigation.” Both the statute authorizing the National Assessment and the policies of the National Assessment Governing Board make it clear that this must not be the case.

By law, NAEP is authorized only to collect contextual information that is “directly related to the appraisal of academic achievement, and to the fair and accurate presentation” of assessment results. NAEP must not evaluate or assess “personal or family beliefs and attitudes.” The assessment may not disclose “personally identifiable information” and cannot report data on individual students or schools. Under Board policy, adopted in 2002 and retained in the 2003 framework and new update, any questions on student attitudes toward school or various academic subjects, such as reading or science, must be “non-intrusive and have a demonstrated relationship to academic achievement.”

In the text of the updated framework the terms contextual and background are used interchangeably, though contextual is the most common and preferred terminology.
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 contextual 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 contextual variables, the quality of the information obtained, and the validity of inferences drawn from it. There is also concern about the burden that collecting contextual 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 contextual 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 contextual 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. Used with other research, the contextual data collected by NAEP may give insights into how achievement can be improved as well report to the public on how school personnel and resources related to achievement are distributed.

This framework defines the purpose and scope of NAEP’s system of collecting contextual information, including background questionnaires and other sources of non-cognitive data. It establishes criteria for reporting contextual 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 contextual 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 contextual data shall be designed to obtain information that is objective, valid, reliable, and of consistently high quality.

• The system of contextual 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 contextual 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 contextual 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 contextual 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 Contextual 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, socioeconomic status, disability, and limited English proficiency. It requires fair and accurate presentation of achievement data and permits the collection of contextual 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 contextual questions, as well as the cognitive items over which it has had final authority since 1988. This Contextual Information Framework will guide the development and selection of non-cognitive topics and questions. It will fulfill the purposes of law and provide a clear statement of 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 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, but, used with other research, the contextual data collected by NAEP may give insights into how achievement can be improved as well report to the public on how school personnel and resources related to achievement are distributed.

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 May 18, 2002). That policy is incorporated into this framework. The framework also incorporates the Board’s more recent Policy Statement on NAEP Background Questions and the Use of
Contextual Data in NAEP Reporting (adopted August 4, 2012). Both policy statements are included 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. **Efforts should be made to develop a composite measure or index of SES.**

(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. **Modules should be prepared for special studies to provide descriptive information on issues of current policy interest.** Although many non-cognitive variables may be of interest, they must be limited to meet the needs of NAEP reporting. In all cases, they** non-cognitive variables 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 contextual 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?**

- **Will the information obtained be of value in understanding academic performance and taking steps to improve it?** This is a fundamental issue to be addressed in evaluating all background questions proposed for NAEP.

    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. NAEP should include
contextual questions from international assessments, such as PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study), to obtain direct comparisons of states and TUDA districts to educational practices in other countries.

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 contextual 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 the 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 information currently collected from students and schools. To the extent that the index can be sharpened by additional data from readily available sources, such as zip codes and the 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 accelerated to develop and use improved measures of socio-economic status, including an SES index, 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 contextual information should be focused, limited, and prioritized as part of the subject-matter frameworks adopted by the Board. For subjects tested regularly at two-year or four-year intervals, each subject 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.

Whenever feasible, student assessment samples should be divided (spiral sampling) and contextual questions rotated in different years in order to cover more topics without increasing respondent burden. These practices should be initiated in the assessments of reading and mathematics, which are conducted every two years, and considered for other subject areas if the frequency of testing permits.

Clusters of questions should 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. These clusters could be administered regularly or rotated across assessment cycles and may be used to construct indexes on topics of interest rather than reporting individual items alone.

Thorough reviews should be regularly conducted to eliminate duplicative or low-priority questions. Unproductive topics and questions should be dropped.

Detailed frameworks will be published with the theoretical rationale and research evidence that support the selection of topics and questions in contextual 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.

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 contextual 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 on paper-and-pencil tests, 15 minutes per student on computer-based assessments, 20 minutes for each teacher, and 30 minutes for each school. Consideration should be given to increasing student response time on paper-and-pencil questionnaires if deemed practical and productive.

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
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. These reports should feature significant contextual information as well as cognitive results. 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 contextual 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. 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.

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 contextual variables have added valuable information. 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 contextual 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.
Chapter One: Introduction

The National Assessment of Educational Progress is the only continuous long-term measure of student achievement in the United States in elementary and secondary schools. Its primary purpose is to report to the American public on academic achievement and its change over time.

Nature and Purpose of NAEP

The NAEP survey consists of two major components: academic assessments that measure the achievement of students on a broad range of content, and non-cognitive survey questions that collect descriptive information from students, teachers, and school administrators about demographic characteristics and the educational process. Since 1969 NAEP has measured achievement in most areas of the school curriculum, including mathematics, reading, writing, science, U.S. history, world geography, civics, economics, foreign language, computer science, and the arts. The content of NAEP assessments is determined through a framework development process that articulates the content parameters for each area and recommends subject-specific non-cognitive areas for data collection and reporting.

NAEP’s purpose is to report to the public on the status of academic achievement in America. The assessment does not report results for individual students, but only for groups of test-takers having large, representative samples, e.g., students from rural schools, from various ethnic groups, or from participating states, and, on a trial basis, large urban school districts. It must be able to provide data for fair and accurate comparisons between the states and subgroups on which it reports. The contextual data play a crucial role in ensuring the fair comparisons—over time and between student groups—that are at the heart of NAEP’s mission and value.

Nature and Purpose of Contextual Data

The most recent NAEP reauthorization (P.L. 107-110) gives the National Assessment Governing Board “final authority” to approve “all cognitive and non-cognitive assessment items.” This framework deals with the non-cognitive side of the Board’s responsibility, including the items that identify students in NAEP’s required reporting categories and the other information that provides a context for results and tracks factors associated with academic achievement.

The term “non-cognitive,” as used in the law, seems more inclusive than the phrase “background questions” by which the collection of non-academic information has been termed by NAEP in the past. However, non-cognitive is also less readily understandable than background or contextual information. In this document the terms will be used interchangeably to refer to all of the information beyond the academic assessment that NAEP uses to make its academic results more meaningful to the public.
When NAEP began, the collection of non-cognitive data was limited to the demographic categories of gender and race/ethnicity, and to two measures of home environment or socio-economic status—level of parents’ education and literacy materials in the home. In addition, an index was constructed, based on data from the U.S. Census and a brief school questionnaire, to report achievement results for schools in three types of communities—disadvantaged urban, advantaged urban, and rural.

During the 1980s the use of non-cognitive questions was greatly expanded to accommodate several functions within NAEP (Reckase, 2002). First, they were used to define a more extensive array of subgroups of the student population for reporting purposes. For example, NAEP results are now reported by gender, race/ethnicity, parents’ highest level of education, type of school, participation in Title I, and eligibility for free/reduced-price lunch.

A second reason for collecting non-cognitive information is to inform educational policy by describing the contexts for learning, sometimes called opportunities to learn (Mullis, 2002). Broadly, this involves the content specified in the curriculum, whether and how that content actually is taught, students’ propensity to learn, as well as home and school factors that can enhance learning.

In conjunction with the descriptions of students, contextual information about educational settings and experiences can reveal striking differences in how important aspects of education and educational resources are distributed among different groups. For example, do disadvantaged minority students have less access to science laboratory equipment than more advantaged groups? Do girls take less rigorous mathematics courses than boys? The data on course taking has been used widely to discuss the patterns and trends in mathematics achievement. Having this information as part of NAEP has added to the public impact of assessment results.

A third function of the non-cognitive questions has been to support research into factors that may be related to student achievement. The questions serving this function have sought information not only on curriculum, teaching methods, and discipline in the school, but also on educational activities at home. For example, The 1998 NAEP Reading Report Card (Donahue, Voelkl, Campbell, & Mazzeo, 1999) reports on television viewing, daily reading habits, classroom reading and writing assignments, and discussion of schoolwork at home. While secondary researchers have used NAEP to investigate relationships to student achievement, the basic design of the assessment as a cross-sectional survey without longitudinal data limits its usefulness. Research has been most productive when NAEP is combined with other data sources and in descriptive studies that track changes over time.

Non-cognitive data are also necessary to support certain technical functions of NAEP. For example, some non-cognitive information is used to evaluate the potential for bias resulting from non-participation. That is, did the students absent or refusing to participate in the assessment differ in such significant ways from those who did take part that results were changed? Non-cognitive variables also play an important role in
NAEP’s sampling and weighting procedures, and sometimes in checking the validity of results. Many of these variables are taken from other data sources, such as the Common Core of Data (CCD), but some come from the administration roster collected from schools prior to testing, the records kept by test administrators, and student questionnaires.

Finally, NAEP non-cognitive questions have been used in the technical process for preparing estimates of student proficiency distributions on the cognitive component of the assessment. But their role in this process is limited to facilitating data analysis. Only the student responses to cognitive questions are used to determine achievement results. Contextual variables are used to define the groups for which cognitive data are reported.

Once test results for a group are determined, the NAEP analytic process makes use of contextual data available to prepare a second data set—identical in its group scores to the first—that can be handled by much simpler computer programs to prepare other analyses and reports. However, only the contextual factors to be reported on are needed for this analytical work, called conditioning. The precision of NAEP results is not reduced if contextual items not used for reporting are eliminated.

This contextual information framework will focus the collection of non-cognitive information on NAEP’s primary mission: providing, as the law stipulates, “a fair and accurate measurement of student academic achievement and reporting trends in such achievement” over time. Thus, the framework is a guide for gathering important information that will assist in reporting and understanding NAEP results.

**Development of NAEP Contextual Information Framework**

In the Policy Statement on Redesigning the National Assessment of Educational Progress (adopted in August 1996), the Governing Board sought to improve the validity of contextual information on NAEP, increase the efficiency with which it is collected, and reduce the number of contextual questions in the assessment itself. The statement was based on the report of a Design/Feasibility Team (Forsyth et al, 1996), headed by Robert Forsyth, which recommended a design that would rotate the collection of non-cognitive data into distinct modules administered over several assessment cycles. NAGB endorsed implementing that recommendation through a system of comprehensive and standard NAEP assessments that would be administered on a cyclical basis (NAGB, 1996).

**Standard** assessments would ask a short, essential core of contextual questions associated with a content area. Periodically, a comprehensive assessment would employ a much fuller complement of such questions to probe that area more extensively. Although some efforts have been made to reduce the contextual questionnaires and streamline data collection, the full impact of the NAGB policy has not yet been realized.

In early 2002, the No Child Left Behind Act transferred final authority over the non-cognitive questions from the National Center for Education Statistics to the National
Assessment Governing Board. The Board adopted a new policy governing the development and selection of non-cognitive questions in May 2002, and initiated a process to prepare a general framework for non-cognitive data (NAGB, 2002). This framework would define the scope of NAEP contextual questionnaires, the priorities for collecting non-cognitive information, and the criteria for reporting non-cognitive data in NAEP. (See Appendix for full text of the policy.)

The Board created an Ad Hoc Committee on Background Questions and conducted an all-day workshop on the NAEP non-cognitive questions on September 24, 2002. Six consultants prepared and presented papers at the meeting that was attended by Board members, academic researchers, representatives of the national teacher organizations and other education groups, and NAEP contractors and staff. The six consultants are identified on the title page as contributors to this document.

In the months after the workshop, a draft framework was prepared. It was refined at several meetings of the Ad Hoc Committee, posted for public comment on the Internet, and was the subject of a public forum in Washington, D.C., on May 1, 2003. Altogether, oral comment and written testimony were received from 22 persons and organizations, many with differing perspectives and views. The Ad Hoc Committee and the Board carefully considered these comments, and the draft framework was revised at a Committee meeting on June 25. The Committee heard additional comment and made final revisions on July 31. The background information framework was reviewed by the full Governing Board several times during the course of its development. The Board adopted it unanimously on August 1, 2003.

While this framework is not a consensus document, it does encompass the thinking of a wide range of researchers, policy analysts, and users of NAEP data. It is the product of discussion and deliberation by the Governing Board, and incorporates Board decisions on the nature and focus of the contextual information to be included in NAEP.

**Requirements of NAEP Statute**

The No Child Left Behind Act of 2001 (P.L. 107-110) requires NAEP to collect information on gender, race/ethnicity, socio-economic status, disability, and limited English proficiency. It must report test data on these groups, whenever feasible, that is cross-tabulated, compared, and reported according to the categories required.

The law also requires NAEP to collect only information that is directly related to academic achievement and to the presentation of such information in a fair and accurate manner. This means that NAEP needs to concentrate on variables that are known to be related to achievement rather than on theoretical constructs. The statute requires the Governing Board to ensure that all NAEP questions are “free from racial, cultural, gender, or regional bias”—a provision from previous law. But it adds new language that questions must be “secular, neutral, and non-ideological” and must not “evaluate or assess personal or family beliefs and attitudes.”
In their report on the bill, the House-Senate conference committee that negotiated its final form says the law “does not preclude the use of non-intrusive, non-cognitive questions, approved by the National Assessment Governing Board, whose direct relationship to academic achievement has been demonstrated and is being studied as part of [NAEP] for the purposes of improving such achievement.” The report language is not binding, but is intended to guide implementation of the law. This framework emphasizes that the legal prohibitions must be followed in preparing contextual questions and collecting any other non-cognitive data for NAEP.

In addition, the law makes it clear that NAEP may not disclose any personally identifiable information or maintain any system of records that contains such data. These restrictions are not new. They have dictated careful procedures that must be continued.

**Purpose and Rationale of Contextual Information Framework**

The purpose of the framework for contextual information is similar to that of NAEP’s content area frameworks: to guide the development of the assessment. The content frameworks have described the topics to be tested by NAEP and provided an outline of the assessment for each subject area. Purposefully, the frameworks attempt to be independent of a particular pedagogy. They do not specify what educational resources or processes should be used, but rather describe important achievement results. They provide states, schools, policymakers, and the public with a logical outline of the approach used in constructing the assessment.

The framework for NAEP contextual data will specify the parameters of the assessment from a reporting perspective. The contextual information that NAEP uses in its reports helps to give context and meaning to the cognitive results. It must be collected in a systematic way from the NAEP testing samples either through questionnaires or from other reliable sources, such as school records and other federal surveys. Collecting descriptive information from a variety of sources can improve the quality of the data obtained and increase efficiency while reducing the burden on respondents.

The Governing Board adopted a Policy Statement on the Collection of Reporting of Background Data on May 18, 2002 (NAGB, 2002). The statement is incorporated into this framework and attached in the Appendix. A further statement, entitled Policy Statement on NAEP Background Questions and the Use of Contextual Data in NAEP Reporting, was adopted by the Board on August 4, 2012. It has been used in revising the framework text and has been included in the Appendix.
Chapter Two: Priorities and Criteria For Collecting and Reporting Non-cognitive Data on NAEP

This chapter presents priorities for collecting and reporting non-cognitive information on NAEP. It also includes the criteria for selecting particular topics and questions, and for determining the frequency with which various data elements are reported. A final section presents criteria for identifying and selecting contextual data sources.

Priorities for Non-Cognitive Information

The following priorities for collecting and reporting non-cognitive information are based on legal requirements, the purposes of NAEP, and the strengths and limitations of the assessment. They should be followed in planning contextual 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. Efforts should be made to develop a composite measure or index of SES.

(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 and course requirements, student mobility, school safety and discipline, teacher-related factors such as teacher demographics, preparation, credentials, and experience, and other factors related to students, schools, and educationally-relevant variables beyond the school. Modules should be prepared for special studies to provide descriptive information on issues of current policy interest. Although these types of non-cognitive variables are of interest, they must be limited so that they meet the needs of NAEP reporting. In all cases, non-cognitive variables must be clearly related to academic achievement or to the fair presentation of achievement results.
Subject-specific information may be gathered at the same time that academic achievement in a particular area is assessed. This may include relevant course content and requirements, teacher preparation, and other factors related to achievement in the subject assessed. Questions will not be designed to determine effective practices, but to show the patterns and trends of factors of interest, based on previous research. Like other 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.

With regard to the points above, Walberg (2002) makes a suggestion that might be a workable solution to consider. Just as students in the NAEP samples do not respond to all the questions, say, in reading, but only to a portion of those for any one grade-level, so too, the non-cognitive questions could be rotated through different (smaller) NAEP samples. These non-cognitive “testlets” could be rotated through the NAEP samples by class or school, with students receiving different, expanded “testlets” in addition to a core set of contextual questions.

Criteria for Selecting Non-cognitive Topics and Questions

The Advisory Council on Education Statistics (ACES), a technical panel that used to advise the National Center for Education Statistics, spent a considerable amount of effort on the issue of NAEP non-cognitive questions. Its guidelines, adopted in May 1997, include a set of key questions that should be utilized in selecting topics and questions for NAEP contextual data collection. The questions with commentary are summarized below:

- **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. It is already known that some non-cognitive variables in NAEP have weak reliability (e.g., data from 4th graders on their parents’ highest level of education and the self-reports of teachers on classroom practice). If more reliable sources of such data cannot be found, these variables should be deleted from the assessment.
• **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.

• **Is the proposed or current non-cognitive variable of timely interest?** The educational environment changes from time to time, and consequently public interest in particular variables will change as well. It would serve NAEP well to review the set of non-cognitive variables periodically with this criterion in mind, deleting those that do not meet the test of timeliness and substituting others of current interest.

• **If new questions are added, have others been deleted in order to limit the burden and expense of NAEP’s contextual questionnaires?** There will always be pressure to collect more information. Mechanisms must be developed to make sure the burden of contextual questionnaires does not expand over time.

• **Does a question address specific behavior rather than conclusions?** For example, a question that asks teachers whether they adhere to national standards in mathematics or another subject is conclusionary and hard to interpret, since many teachers are apt to say yes, regardless of what they do. It would be better to ask about specific behaviors, such as homework assignments or computer use. Caution is advisable in this area too because self-reports are often unreliable.

The Board believes three other important criteria must also be considered:

• **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 must readily be available to anyone requesting a copy. Under Board policy, all questions asked are to be posted on the Internet. Possible objections should be considered in deciding whether or not to ask them.

• **Does the topic or question deal with a factor in which trends over time are of importance?** If trends are deemed important and the factor is related to achievement, the topic or question should be included periodically on a four-year or eight-year cycle, rather than being part of the contextual questionnaire each year. For example, measuring television watching in every NAEP assessment is not necessary. But it can be valuable to measure TV-watching every four or eight years to find out whether or not it is increasing.

• **Will the information obtained be of value in understanding academic performance and taking steps to improve it?** This is a fundamental issue to be addressed in evaluating all contextual questions proposed for NAEP.
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. NAEP should include contextual questions from international assessments, such as PISA (Program for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study), to obtain direct comparisons of states and TUDA districts to educational practices in other countries.

Criteria for Selecting Data Sources

NAEP has collected non-cognitive information from students, teachers, and schools, using NAEP contextual questionnaires. There are also administration rosters, completed by test administrators at the school level in advance of testing to determine characteristics of the testing samples. The Common Core of Data (CCD) is used to identify characteristics of schools (e.g., Title I funding), and schools also complete a questionnaire on special needs students (e.g., students with disabilities and limited English proficiency).

However, the collection of non-cognitive data may be shifted among these sources or to new sources in order to improve reliability, increase efficiency, or reduce burden. State management information systems and data collected for school report cards, as required by the No Child Left Behind Act, may have become very increasingly useful for NAEP. Whenever possible, NAEP should use information from school records and other reliable data collections about students and schools in order to improve the validity of the information collected and limit the contextual questionnaires in NAEP itself.

In exploring the utility of different data sources, the following criteria should be considered:

- **Validity** – Is the data obtained from the new source a valid indicator of what it purports to measure?

- **Reliability** – Is the data from the new source at least as reliable and consistent as that from the source previously used?

- **Universality** – Can the required data be collected by this method for all (or almost all) of the students and schools participating in NAEP and will it support valid comparisons over time?

- **Currency** – Will data obtained from a new data source be current enough to relate clearly to the assessment being conducted? If data from the census or some other source is several years old it may not accurately describe school or neighborhood conditions at the time of testing.
• **Respondent Burden** – Will the new source(s) reduce the burden on students, teachers, and schools in filling out NAEP questionnaires? Will the total amount of respondent burden be decreased?

• **Logistics** – Will the alternative source(s) be logistically possible, or will there be more logistical problems than with the previous data source? Logistics includes such considerations as cost, time, administrative personnel resources, and steps needed to ensure accurate coding and data analysis.

• **Efficiency and cost-effectiveness** – How efficient will the new data source be in comparison to the previous one? For example, it may be more efficient to collect data from a state management information system about the state’s schools, teachers, or students, rather than obtaining it from the test samples directly, but efficiency and cost-effectiveness should be determined before a change is made.

• **Timeliness of NAEP reporting** – How will a change in data sources affect the speed with which NAEP can be reported? Some changes will speed operations, but those that slow down NAEP reporting are not desirable.
Chapter Three: Topics and Types of Contextual 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 socioeconomic status (SES), contextual factors of interest to public policy, and subject-specific variables.

Demographic Reporting Categories

The demographic variables currently collected by NAEP come from two sources. Information is obtained from school records on 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. The school records are also used to indicate whether a student is classified as disabled. In addition, information is obtained on disability status and on students who are classified as having limited English proficiency. 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 the highest level of each parent’s education and on several aspects of home environment, including number of books, internet access, and whether they have their own bedroom.

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 into different types of location, based on Census Bureau definitions, such as central city, suburban/large town, and rural/small town. The questions on newspapers and magazines were dropped in the mid-2000s as circulation dwindled, and were replaced by an item on internet access.

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 have been 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 used until the mid-2000s student responses are were 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, & 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, 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. The data on eligibility for the National School Lunch Program have also become increasingly problematic because of expansion of the program and administrative changes allowing whole-school or whole-district eligibility in high-poverty areas. 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 should be accelerated to develop and use improved measures of socio-economic status, including an SES index made to find an approximation that is more reliable than the current set of individual proxy variables.

In November 2012, an expert panel convened by the National Center for Education Statistics recommended prompt development of an SES composite measure. 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 (TUDA), 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 in 2000 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 item on number of parents in the home was restored in 2013. 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 Efforts should be conducted—made to develop into creating—a composite measure or index of SES based on school records and the 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.

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. Although steps must be taken to ensure that such data are reliable. An expanded set of additional SES variables may be included, such as number of siblings and parents at home, possession of computers, and parent occupation. Periodically, an expanded set may be administered.

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, TV-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. Modules should be prepared for special studies to provide descriptive information on issues of current policy interest. Information Data 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, & 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 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, & 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 Contextual Data

For each subject assessed by NAEP, additional subject-specific contextual 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 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.
Whenever feasible, student assessment samples should be divided (spiral sampling) and contextual questions rotated in different years in order to cover more topics without increasing respondent burden. These practices should be initiated in the assessments of reading and mathematics, which are conducted every two years, and considered for other subject areas if the frequency of testing permits.

The design for doing this was presented to the Board in the 1996 report of a Design/Feasibility Team of prominent researchers (Forsyth, R. 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 2007, the years in which revised frameworks are first used, and then should be repeated at an interval of 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.

In The NAEP assessments in other subjects given at intervals of four years or more, such as writing, science, history, geography, and civics, should have a core set of non-cognitive questions should be administered to the full sample, with different sets of 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 from what it was in the program’s first decades. This makes even more compelling the case for limiting the NAEP contextual questionnaires and rotating the background questions.

Clusters of questions should 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. These clusters could be administered regularly or rotated across assessment cycles and may be used to construct indexes on topics of interest rather than relying on stand-alone items only.

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

<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>TBD</td>
</tr>
<tr>
<td>World History (12)</td>
<td>2010</td>
<td>TBD</td>
</tr>
<tr>
<td>Economics (12)</td>
<td>2006</td>
<td>2006</td>
</tr>
<tr>
<td>Arts (8)</td>
<td>1997, 2008</td>
<td>2005</td>
</tr>
<tr>
<td>Science</td>
<td>2000, 2009</td>
<td></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.
Chapter Four: Non-cognitive Data Sources and Collection

This chapter discusses the sources of non-cognitive information for NAEP and the reporting categories that the information describes. It includes a NAEP Contextual Information Matrix, organized by priorities, which summarizes the types of descriptive information NAEP collects, reporting units, and data sources.

NAEP Student, Teacher, and School Samples

The NAEP student samples vary in size and purpose. Their overall total has become very large. Starting in 2003, national NAEP samples are specified at the state and jurisdictional levels, with approximately 3,000 students per subject and grade (4 and 8 only) for each of the 50 states, plus the District of Columbia, and Department of Defense domestic and overseas schools. Puerto Rico (in mathematics only) has a sample of about 3,000. In addition, the ten Trial Urban District Assessment (TUDA) districts have sample sizes of the order of 3,000 to 5,000 each. There also are a nationally-representative sample of charter schools, totaling about 3,000 students, and national private school samples totaling about 12,000 in each grade.

At grade four, therefore, the total NAEP sample approximates 436,000 students. The grade eight sample is about the same at 432,000 (excepting charter schools). The grade 12 sample is for a pilot test and includes only about 6,000 students (Rust, 2002). In most future years the twelfth grade samples are expected to have about 30,000-40,000 students assessed in national samples only for three subjects.

In addition to the nearly one million students tested, about 80,000 teachers of those students complete teacher questionnaires and some 13,000 schools complete school questionnaires. Several thousand school districts also supply data for the assessment. The sampling and weighting procedures in NAEP use data from the CCD files as well as census data and school-level achievement data from the states for improving NAEP stratification procedures. The NAEP non-cognitive data collection effort is enormous and challenging.
Other Data Sources

The Governing Board is strongly committed to improving the quality of contextual information while reducing respondent burden and the complexity of data collection and analysis. The self-report questionnaires given to students, teachers, and schools are sometimes burdensome to fill out, labor-intensive to collate and analyze, and subject to concerns about reliability. All questionnaires should be scrutinized to replace as many items as possible with data from centralized records, gathered by test administrators, or, ideally, from computerized data files.

The data available from federal, state, district, and school records should be carefully explored. With implementation of the school report card requirements of the No Child Left Behind law, in recent years much more information should become available soon in standardized computer formats. Barton (2002) has suggested some specific sources of data collected outside of NAEP that should be considered to improve NAEP reporting. These include the U.S. Census, Quality Education Data, Inc. (QED), and the Common Core of Data (CCD) and School and Staffing Survey (SASS), both compiled by the National Center for Education Statistics.

This approach of utilizing more data from outside specific NAEP data collections has been elaborated on extensively in the most recent evaluation of NAEP by the National Academy of Sciences (Pellegrino, J.W., Jones, L.R., & Mitchell, K.J., 1999). The panel proposed “a coordinated system of indicators for assessing educational progress, housed within NCES and including NAEP and other currently discrete, large-scale data collections (p. 34).” Figure 1 is reprinted from the NAS report to show the extent of these data collections on students, teachers, and schools, and to indicate what might be obtained from these other sources. To use them for NAEP would greatly lessen the burden on the assessment itself. Merged data sets could be made available, some to the general public, and more to researchers in restricted data files.

For many years state-level NAEP reports have included appropriate collateral data that provide a context for interpreting NAEP results; see for example the NAEP 1996 Mathematics: Report Card for the Nation and the States (Reese et al., 1997). These state contextual variables have included enrollment in elementary and secondary schools, poverty status of children from 5 to 17 years old, number of children receiving disability services, per-pupil expenditures, pupil-teacher ratios, and average teacher salaries. To the extent that these data are readily available and are helpful in setting a context for interpretation of NAEP results the practice ought to be continued. However, more effort should be made to ensure that such data are as up-to-date and easily-accessible as part of NAEP reporting on the Internet.
### Overview of Current NCES Data Collections

<table>
<thead>
<tr>
<th>Data and Design Elements</th>
<th>NAEP</th>
<th>NELS</th>
<th>ELS</th>
<th>ECLS</th>
<th>TIMSS</th>
<th>CCD</th>
<th>PSUS</th>
<th>SASS</th>
<th>NHES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Elements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student achievement</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Student background characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support for learning</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Standards and curricula</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Instructional practices and</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>learning resources</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School organization/governance</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Teacher education and</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>financial development</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial resources</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School climate</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Design Elements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of design (CS=cross-sectional; L=longitudinal)</td>
<td>CS,L</td>
<td>L</td>
<td>L</td>
<td>CS</td>
<td>L</td>
<td>L</td>
<td>CS</td>
<td>CS</td>
<td>CS</td>
</tr>
<tr>
<td>Periodicity (TBD=to be determined)</td>
<td>2,4, or 6 yrs</td>
<td>2-6 yrs</td>
<td>TBD</td>
<td>TBD</td>
<td>CS</td>
<td>Annual</td>
<td>L</td>
<td>Biennial</td>
<td>2-5 yrs</td>
</tr>
<tr>
<td>Unit of observation (S=student; T=teacher; A=administrators; P=parent; SC=schools; D=district; ST=states; H=households)</td>
<td>S,T,A</td>
<td>S,T,A</td>
<td>S,A,P</td>
<td>S,T,A,P</td>
<td>S,T,A,P</td>
<td>SC,D,ST</td>
<td>SC</td>
<td>SC</td>
<td>T,A,SC</td>
</tr>
<tr>
<td>Data collection method (S=survey; R=record analysis; I= interview; V=video; C=case study; O=other)</td>
<td>S</td>
<td>S,R</td>
<td>S,O</td>
<td>S,O</td>
<td>S,R,V,C</td>
<td>S,R</td>
<td>S</td>
<td>S</td>
<td>I</td>
</tr>
</tbody>
</table>

NELS: National Education Longitudinal Study of 1988  
ECLS: Early Childhood Longitudinal Study  
ELS: Educational Longitudinal Study of 2002  
PSUS: Private School Universe Survey  
SASS: Schools and Staffing Survey  
NHES: National Household Education Survey  

NAEP Contextual Information Matrix

The types of descriptive information NAEP collects, reporting units, and data sources are summarized in the NAEP Contextual Information Matrix, which is displayed as Figure 2. The matrix is intended to assist in conceptualizing NAEP contextual information collections. It is organized by priorities—both for types of information and for how data should be obtained. Note that in each case information is to be obtained from reliable official records before it is sought through questionnaires.

The entries in the cells are illustrative, showing the kinds of information that are currently collected by NAEP and the various data sources (records and questionnaires) that are used. As the principles of this framework are implemented, more information will come from records, less from questionnaires. The sources with higher reliability and less respondent burden should be utilized in priority order.

The Ad Hoc Committee on NAEP Background Questions considered a proposal by Paul Barton (2002) to permit states or groups of states to add customized sets of questions to the contextual questionnaires. Although these might track progress on topics of particular interest and increase support for NAEP, the Committee felt strongly that the proposal should not be pursued because any customization of NAEP questionnaires would create serious logistical and quality control problems.

In constructing questionnaires it is important to place strict limits on the respondent burden they impose. The average individual response time to answer contextual questionnaires for each assessment, as calculated in accordance with Office of Management and Budget (OMB) procedures, shall be limited as follows: ten–10 minutes for each student on paper-and-pencil tests, 15-minutes per student on computer-based assessments, 20 minutes for each teacher, and 30 minutes for each school. Consideration should be given to increasing student response time on paper-and-pencil questionnaires if deemed practical and productive.
Figure 2

NAEP Contextual Information Framework

<table>
<thead>
<tr>
<th>Reporting Unit and Data Sources</th>
<th>Type of Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STUDENT</strong></td>
<td><strong>Socio-Economic Status</strong></td>
</tr>
<tr>
<td>School Records</td>
<td><strong>Core Expanded</strong></td>
</tr>
<tr>
<td>Questionnaire</td>
<td><strong>Other Contextual Information</strong></td>
</tr>
<tr>
<td>Gender</td>
<td>Free/RP lunch participation</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>Title I</td>
</tr>
<tr>
<td>SD/LEP</td>
<td>Parent education</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>Reading materials and Internet access, in home</td>
</tr>
<tr>
<td><strong>SCHOOL</strong></td>
<td><strong>Subject-Specific Information</strong></td>
</tr>
<tr>
<td>Dist/State Reeds</td>
<td>Course taking in mathematics</td>
</tr>
<tr>
<td>School Records</td>
<td>Time spent on math homework</td>
</tr>
<tr>
<td>CCD/Census</td>
<td>Good in math?</td>
</tr>
<tr>
<td>Questionnaire</td>
<td><strong>Non-NAEP contextual variables</strong></td>
</tr>
<tr>
<td>School type</td>
<td>Graduation requirements in math and science</td>
</tr>
<tr>
<td>(public, private, charter, etc.)</td>
<td>Higher level math courses</td>
</tr>
<tr>
<td>School ach. data</td>
<td>Graduation testing</td>
</tr>
<tr>
<td>Community type</td>
<td>Extracurricular options in math and English</td>
</tr>
<tr>
<td><strong>TEACHER</strong></td>
<td><strong>Availability of computers for writing.</strong></td>
</tr>
<tr>
<td>School Records</td>
<td>Correct for spelling and grammar?</td>
</tr>
<tr>
<td>Dist/State Reeds</td>
<td>Frequency of science lab work</td>
</tr>
<tr>
<td>Questionnaire</td>
<td><strong>Non-NAEP contextual variables</strong></td>
</tr>
<tr>
<td><strong>STATE</strong></td>
<td><strong>Race and Gender Experience Credentials</strong></td>
</tr>
<tr>
<td>CCD/Census</td>
<td>Undergrad/Grad content training</td>
</tr>
<tr>
<td>State Records</td>
<td>Professional Devel</td>
</tr>
<tr>
<td>Questionnaire</td>
<td><strong>Communty type (urban, rural, etc.)</strong></td>
</tr>
<tr>
<td>Region</td>
<td><strong>Non-NAEP contextual variables</strong></td>
</tr>
<tr>
<td><strong>DISTRICT</strong></td>
<td><strong>Community type (urban, rural, etc.)</strong></td>
</tr>
<tr>
<td>CCD/Census</td>
<td><strong>Non-NAEP contextual variables</strong></td>
</tr>
<tr>
<td>State Records</td>
<td><strong>Community type (urban, rural, etc.)</strong></td>
</tr>
<tr>
<td>District Records</td>
<td><strong>Non-NAEP contextual variables</strong></td>
</tr>
<tr>
<td>Questionnaire</td>
<td><strong>Community type (urban, rural, etc.)</strong></td>
</tr>
</tbody>
</table>

NOTE: Information type and data sources are arranged in priority order.
Chapter Five: Using Contextual Data to Report NAEP Results

This chapter discusses the descriptive information that NAEP should provide, the levels of disaggregation now possible with merged national and state samples, and the importance of minimizing causal interpretations.

Use of Descriptive Information in NAEP

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 the patterns and trends of non-cognitive variables known to have a relationship to academic achievement.

In addition, special supplemental reports may be prepared that focus on particular aspects of the background data collected — topics of public interest and importance. Such reports should feature significant contextual information as well as cognitive results. Advisory committees, including a range of knowledgeable persons, may be appointed to provide input on reporting issues. 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 unsupported by research that may imply best practice.

All contextual questions and data collected by NAEP should be made available on the Internet at the time of the initial release of the principal academic results or soon afterwards so the public may be able to consider them in discussing results. Complete data files should be available to researchers for further analysis.

Implementing No Child Left Behind

The intent of the No Child Left Behind Act of 2001 (P.L. 107-110) is to hold public schools accountable for closing achievement gaps between different groups of students. NAEP is expected to contribute to this end by providing an accurate measure of current levels of student achievement and monitoring change over time.
Descriptive information about all students, but particularly on low-performing groups, would contribute powerfully to the dialogue on the challenges before American education. For example, the NAEP achievement levels focus on the segments of the performance distribution that are at or above Basic, Proficient, and Advanced. Information should also be provided about those Below Basic, who clearly have been “left behind”: e.g. the proportion having qualified teachers, receiving free or reduced-price lunch, or moving to different schools frequently, as measured by attending the same school for less than two years.

Such profiles of low-performing or high-performing students should not attempt to ascribe causation, but they would provide important information on the distribution of practices and resources that are of concern to the public and policymakers. Periodic collections of such contextual data could be used to track change in the distribution of these factors over time. Do the trends seem favorable or adverse to educational progress?

Disaggregation of NAEP Data

For more than three decades since it was established, NAEP has provided data disaggregated by race/ethnicity, gender, school type (e.g., public/private), and community type (e.g., urban/rural). The No Child Left Behind law calls for disaggregation by major subgroups (when feasible) of race, ethnicity, and gender, and also by socio-economic status, disability, and limited English proficiency.

Because of the large size of the recently combined national and state NAEP samples, NAEP reports should be able to provide information disaggregated at a much greater level of detail than was possible in the program’s first decades. Pooling the data from all states, which now are required to provide NAEP samples in fourth and eighth grade reading and mathematics, will produce a much enlarged national sample that will sharply reduce the number of empty cells in any cross-tabulations. Such disaggregation might add to the richness of NAEP reporting even for only a limited set of non-cognitive questions. Disaggregation is also very important for reporting on the distribution of student characteristics within the different achievement levels, as described above.

Minimizing Causal Interpretations

NAEP has often reported on the average performance of students by particular non-cognitive variables. One example, presented in many NAEP reports until the early 2000s, was the average scale score of students who watch different amounts of television each day, cf. The Nation’s Report Card: Reading, 2000 (Donahue et al., 2001). Another example is the average scale scores for 12th graders who report

While there may be a correlation between TV-watching and reading performance, or between hours working outside school and math results, NAEP is not designed to prove cause-and-effect relationships. As a cross-sectional survey, nearly all of its data is on current activities and practices—not on the complex chain of experience in school and outside, of prior learning and achievement that all contribute heavily to current academic performance. While the correlations may be of interest, they cannot be conclusive. But they may be cited to stimulate discussion or encourage further research. Yet, NAEP has encouraged simple causal inferences by reporting average scores for varying amounts of time spent on current activities.

There is one important exception to the absence of data on learning-related activity over time. This is the information NAEP collects on the transcripts of high school seniors and its questionnaires on courses that students have taken and schools provide. These do show prior instruction before current exams. The trends in course taking have been of great public interest and it is reasonable to relate them to student achievement.

NAEP reports should present information on the patterns and trends of non-cognitive variables known from other sound research to have a relationship to academic achievement. These presentations should be straightforward and impartial, and care must be taken to avoid stating conclusions as to cause and effect relationships. Further analysis of any relationships should be left to researchers.
Chapter Six: Using NAEP in Educational Research

As a cross-sectional survey without longitudinal data, the National Assessment of Educational Progress 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-four decades, NAEP has been part of two-three 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; and evaluating the stringency of state standards enacted under No Child Left Behind.

By doing its main task of monitoring achievement well, NAEP has provided sound data for researchers to use. NAEP results have been critical in identifying hypotheses for other research to pursue. Its large data sets, including contextual variables, 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 the National Assessment can be of value to educational research and the interest of researchers in the assessment should be encouraged. The NCES program of secondary analysis grants for researchers to use NAEP data should continue. Educational researchers should be involved, under the auspices of NCES and its contractors, in developing NAEP contextual questionnaires and other data collection efforts to carry out the provisions of this framework.

This chapter considers the limitations and strengths of NAEP for educational research and discusses research that has made use of NAEP data. The chapter draws on papers by David Grissmer, senior research scientist at RAND, who has used NAEP extensively in analyzing educational factors and trends.

NAEP’s Limitations and Strengths for Research

The primary purpose of NAEP is to accurately and fairly monitor achievement over time and accurately and fairly compare achievement across states and important sub-groups of students. Beyond providing such data, any research with NAEP, particularly into the causes of academic achievement, is severely limited by its design.

As a representative sample survey, in which no individual student takes more than a small part of the full exam, NAEP has shortcomings in most of the elements commonly used to evaluate academic achievement (Podgursky, 2002):

- It provides no prior data on student achievement, and can’t be made longitudinal to do so.
• It can only collect contemporaneous information on school practices and resources, and has no way of ascertaining how students were taught or what school experiences they may have had in previous years.

• There is considerable measurement error in survey responses obtained from teachers and schools because they may well give the expected “right” answers rather than report accurately what they do.

• The current classroom practices that teachers report may be a response to student achievement levels, not the cause of such achievement, and it is difficult to disentangle causation.

• It is difficult for NAEP to get good information on socio-economic status or family background factors, but these are powerfully correlated with academic achievement, and must be controlled for in any analysis of school effects.

On the other hand, NAEP does have unique strengths and comparative advantages (Grissmer, 2003), and thus has the potential to address some important research and public policy questions with its cognitive data and contextual information:

• NAEP is the only data set on student achievement that has collected data from nationally representative samples of students continuously from 1969-70 to the present.

• It is the only data set that has collected academic achievement data simultaneously, repeatedly, and consistently from three separate age groups.

• It is the only data set that collects from statistically reliable samples at the state level, and within states for different types of communities (central city, suburban and rural) and for racial/ethnic groups within most states.

• NAEP has far larger sample sizes than any other nationally representative survey of student achievement, such as the National Education Longitudinal Study (NELS) and the Early Childhood Longitudinal Study (ECLS). These surveys are only approximately 10 to 20 percent as large as NAEP in any single application, and 1 to 5 percent as large as NAEP for any repeated data collection.

• NAEP is the only survey that tests a wide range of academic subjects.

• NAEP achievement measures at fourth and eighth grade fill an important void in measuring the well-being of children during this developmental period.

• NAEP generally incorporates a higher quality and unique design of test instruments, administrative procedures, and scoring methodology, compared to other data sets.
Previous Use of NAEP in Research

As a result of its strengths, NAEP has been used in important educational research by authors such as David Grissmer, Alan Krueger, David Armor, and Christopher Jencks. These studies point to an important comparative advantage of NAEP, namely, that it is the only representative sample data in existence on student achievement in the United States from 1969 to 2002. Thus, research into important historical questions about the effects of changing families, communities, and schools on achievement almost require NAEP data. Without NAEP, it is unlikely that the significant narrowing of the black-white score gap would be known and its possible causes the subject of research.

Similarly, NAEP data have been used to help analyze the effects of differences in resources, systemic reform initiatives, differential opportunity for learning, and other educational policies on state-level academic achievement. Such research has concluded that the rates of improvement in achievement varied markedly across states in the 1990s, and that changing resources or demographics cannot account for the gains in the states with most rapid improvement. This research points to another strong comparative advantage of NAEP. State NAEP is the only survey that includes representative samples of students in many different states, and thus plays a central role in monitoring and explaining the differences in academic achievement and achievement trends across the states. NAEP can identify where positive trends are occurring so researchers can puzzle out causation.

A review of research studies using NAEP (Grissmer, 2003) suggests that only a small proportion of the non-cognitive items collected by the assessment have been utilized in productive research. Also, such research has often supplemented NAEP with data from other sources, such as the U.S. Census and the Common Core of Data (CCD) and Schools and Staffing Survey (SASS), both conducted by the National Center for Education Statistics. However, the National Assessment played such a crucial role in these studies that they could not have been conducted without NAEP data, including some of its non-cognitive variables, principally those on socio-economic status, family structure, and school resources.

On the other hand, NAEP data have also been misused for simplistic and weak research. Many contextual data items on school practice and student behavior have been used in a simplistic way to imply a direct, causal relationship to achievement while ignoring the complex mix of other, more fundamental factors that may well have a stronger impact. NAEP has encouraged such associations by presenting one-way tabulations in its reports, e.g. average scale score by hours of television watched, type of reading instruction, or books read per week, and these have been disseminated widely to support particular beliefs or public policy positions. Simple, single variable linkages can often be misleading because of the strong correlations between many contextual variables, particularly with socio-economic status, prior academic achievement, or family background. They should only be included in NAEP reports when there is strong justification based on previous research.
Also, most of the hundreds of contextual questions in NAEP have never been used for either public reporting or research. Many come from the early 1980s, and would be difficult to justify in a sound research design today.

Secondary Analysis Grants and District Samples

For many years almost two decades NCES has been making awards to education researchers for secondary analyses of NAEP data. These explored a range of topics, often in combination with other data sets. Many of the studies have focused on state-to-state differences in student achievement and the impact of state-level policies, relying on NAEP academic data, a few contextual questions for SES controls, and much additional information from other sources. The program has been valuable as a means of encouraging the use of NAEP for research, and, in a few cases, notably the Grissmer studies, has had considerable impact. As in any grant program, all findings are the responsibility of the individual researchers, not of the agency making the grant.

The program should continue, and now that NCES has become part of the Institute for Education Sciences in 2003, the leadership of the new agency should ensure that the separate NAEP analysis grants were absorbed in a more general research program, are aligned with the research priorities of the Institute. We believe this program should increase awards that make use of NAEP data. Efforts should be made through training and other small-scale grants to expand capabilities for using NAEP in productive education research.

In addition, data from the school district NAEP samples in the Trial Urban District Assessment, which started in 2002, will provide important opportunities for research. NAEP results for school districts can readily be combined with Census data, which include pertinent information on family background and socio-economic status. The school district samples can also be tied to important education policy variables, such as per pupil spending, for which information is available at this level but not for schools.

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.
Chapter Seven: Review and Improvement of Non-cognitive Questions

This chapter discusses several mechanisms for the review and improvement of NAEP’s non-cognitive questions and for implementation of the NAEP Contextual Information Framework.

Independent Validity Studies

Since the early 1990s NAEP has had the benefit of independent outside advice on topics of urgency or interest. These studies have been very helpful to the Governing Board and NCES as they made decisions about the future of the NAEP program. For example, several years ago some research was conducted to examine the possibility of combining the NAEP national and state samples to achieve more efficiency and cost-savings. Starting in 2003 NAEP moved in that direction. The decisions surrounding the change, however, were only as good as the research that bolsters it. The work of the current NAEP Validity Panel, in conjunction with the current NAEP operations contractors, has contributed significantly to making the change possible.

The value of this kind of applied research cannot be overestimated. Neither can the value of the independent nature of this work. The NAEP program is very large and complex and demands a commitment of many resources from the NAEP contractors. NAEP contractors should not be burdened with conducting simultaneous research studies while carrying out the requirements of the operations contracts. There is a precedent for this approach in the current separation of responsibilities for operations and research in separate NAEP contracts. There are two reasons why independent validity studies on topics associated with the non-cognitive framework are recommended. First, there are some non-cognitive variables that will need validation, particularly if those variables are new or are new composite indexes of existing variables. Second, following the approach already established for the NAEP cognitive components, recommendations from such research studies must be truly independent and free from any conflict of interest.

Review of the Contextual Information Framework

This contextual information framework should be reviewed on a periodic basis. The NAEP cognitive frameworks are reviewed every ten years. This policy was adopted at the time of the NAEP redesign in 1996. Reviewing a NAEP framework can result in major revision, minor revision, or even no revision and re-adoption. The framework may be updated as needed. A thorough review of the Contextual Information Framework should be undertaken since the background framework is a new undertaking.
a required review after five years is appropriate with additional reviews every ten years thereafter.
References


Armor, D.J. (December 18, 2002). *Comments on NAEP non-cognitive questions.* Available from the National Assessment Governing Board, 800 N. Capitol Street, NW, Washington, DC 20002.


Appendix A

Adopted May 18, 2002

National Assessment Governing Board

Policy Statement on Collection and Reporting of Background Data by the National Assessment of Educational Progress

INTRODUCTION

As the Nation’s Report Card, the National Assessment of Educational Progress (NAEP) is an on-going, Congressionally-authorized program to collect data through surveys on the academic knowledge and skills of American students. Its primary goal is to report fair and accurate information on student achievement in reading, mathematics, and other subjects taught in elementary and secondary schools. This information is to be made available in a clear and timely manner to members of the public, policymakers, and educators throughout the country.

Since it began in 1969-70, NAEP has administered, in addition to cognitive questions, background questionnaires that provide information for reporting categories and collect non-cognitive data on students, their family background, teachers, and schools. These have enriched reporting of the National Assessment and increased the precision of NAEP results. The background data have also been used in secondary analyses. However, because NAEP tests a cross-section of students at a particular time with no follow-up of the students tested, the assessment can only show correlations or associations rather than causal relationships between background factors and achievement.

By statute (P.L. 107-110), the National Assessment shall include, “whenever feasible, information collected, cross-tabulated, compared, and reported by race, ethnicity, socioeconomic status, gender, disability, and limited English proficiency.” The statute provides that NAEP may “not evaluate or assess personal or family beliefs and attitudes” and may “only collect information that is directly related to the appraisal of academic achievement and to the fair and accurate presentation of such information.” These provisions are intended to prevent intrusive, inappropriate, or unnecessary questions being asked about students and their families.
The law requires that the Governing Board take steps to ensure that all NAEP questions are “free from racial, cultural, gender, or regional bias, and are secular, neutral, and non-ideological.” However, a House-Senate Conference report, accompanying the legislation, says the law does not preclude the use of “non-intrusive, non-cognitive questions,” with a direct relationship to academic achievement.

The National Assessment is conducted by the Commissioner of Education Statistics under the policy guidance of the National Assessment Governing Board. The Board’s specific areas of responsibility include: (1) assessment objectives and test specifications; (2) the methodology of the assessment; (3) guidelines for reporting and disseminating results; and (4) “appropriate actions needed to improve the form, content, use, and reporting” of the National Assessment. Under the statute, the Board has “final authority” on the appropriateness of all NAEP items—both cognitive and non-cognitive.

To carry out these responsibilities, the National Assessment Governing Board hereby adopts guiding principles, policies, and procedures for the collection and reporting of background data by the National Assessment of Educational Progress.

GUIDING PRINCIPLES

1. Background data on students, teachers, and schools is needed to fulfill the statutory requirement that NAEP include information, whenever feasible, disaggregated by race or ethnicity, socioeconomic status, gender, disability, and limited English proficiency. In addition, background data is collected to enrich the reporting of NAEP results by examining factors related to academic achievement. However, the collection of such data should be limited, and the burden on respondents kept to a minimum. It must always be considered in light of NAEP’s primary purpose: providing sound, timely information on the academic achievement of American students.

2. All background questions must be directly related to academic achievement or to the fair and accurate presentation of achievement results.

3. Issues of cost, benefit, appropriateness, and burden shall be carefully considered in determining the background questions to be asked and the samples to which they shall be administered.

4. In accordance with law, questions shall be non-intrusive and free from bias, and must be secular, neutral, and non-ideological.

5. No personally identifiable information shall be included in NAEP reports or data releases.

6. Decisions on the retention or addition of background items shall draw on technical studies on the reliability and validity of current and proposed
questions and on the contribution such items make to the precision of NAEP results.

7. Consideration should be given to obtaining background information from non-NAEP sources and to avoiding duplication with other federal surveys.

8. Questionnaires should be revised to keep background questions timely and related to academic achievement. Those questions showing little change over time and/or a stable relationship to achievement should be deleted or asked less frequently and to limited samples, unless required to assure the precision of NAEP results.

9. Questions should not address personal feelings and attitudes.

10. Since security considerations do not apply, background questionnaires shall be readily available to the public.

11. Interpretation of results shall be limited in official NAEP reports and must be strongly supported by NAEP data. Because of the survey nature of the assessment, reports may show correlations and generate hypotheses, but may not state conclusions as to cause and effect relationships.

12. Background questions for NAEP assessments shall be prepared in accordance with frameworks and specifications adopted by the Governing Board.

13. The Governing Board shall review and approve all background items before they are administered in NAEP surveys or pilot and field tests.

POLICIES AND PROCEDURES

1. Framework and Specifications

The Governing Board will adopt a general framework for background questionnaires and specifications for the questions on selected topics and in specific subject areas.

Since this is a new area of responsibility for the Board, the process of developing a framework for background questions and specifications will begin with commissioned papers on relevant issues, such as the reliability and validity of current background questions, their contribution to improving the precision of NAEP results, their value and limitations for educational research, and changes that may be needed in response to the No Child Left Behind legislation. Following consideration of these issues, the Board will define the scope of background questionnaires and adopt a process for preparing a framework and specifications. This work will include the active participation of teachers,
education researchers, state and local school administrators, assessment specialists, parents of children in elementary and secondary schools, and interested members of the public.

2. Background Question Development

In preparing background questions, the National Center for Education Statistics shall follow adopted frameworks and specifications, and consider the review criteria adopted by the Governing Board. NCES may use cognitive laboratories of students, teachers, and school officials to help determine the clarity and burden of proposed questions. Ad hoc advisory committees may also be established, comprised of teachers, parents, technical experts, and others interested in NAEP. Steps shall be taken to determine the reliability of questions used.

3. Governing Board Review and Approval of Background Questions

Background questions for all NAEP pilot tests, field tests, and operational use shall be reviewed and approved by the Governing Board. The category of respondents, e.g. students, schools, and grade level, shall clearly be designated, as will the NAEP samples, e.g. national, state, or district, in which the questions will be asked.

For each questionnaire there shall be an explanation of its intended use in NAEP reporting and analysis and of the hypothesized relationships between the background items and student achievement that demonstrates the need to know such information. Technical data shall be presented on the reliability and validity of questions and, if applicable, on their contribution to improving the precision of NAEP results. The Board will use the explanations and data presented along with the review criteria in this policy statement in determining the appropriateness of background questions.

The Reporting and Dissemination Committee will have primary responsibility for the review and approval of background questions. The Assessment Development Committee will participate in the approval of questions relating to specific subject-matter assessments. Ad hoc committees of Board members may be established by the Board Chairman for background question review. Questions may also be reviewed by external advisors, including teachers, parents, and technical experts. Recommendations on background questionnaires shall be subject to final approval by the full Governing Board.
4. Criteria for Governing Board Review

The following criteria for review and approval of background questions are based on the most recent revision of the authorizing statute of the National Assessment of Educational Progress (P.L. 107-110) and the Guiding Principles of this policy statement:

A. Background information is needed to fulfill the statutory requirement that NAEP report and analyze achievement data, whenever feasible, disaggregated by race or ethnicity, gender, socio-economic status, disability, and limited English proficiency. Non-cognitive data may enrich the reporting and analysis of academic results, but the collection of such data should be limited and the burden on respondents kept to a minimum.

B. All background questions must be related to the primary purpose of NAEP: the fair and accurate presentation of academic achievement results.

C. Any questions on conditions beyond the school must be non-intrusive and focused on academic achievement and related factors.

D. Questions shall be free from racial, cultural, gender, or regional bias.

E. All questions must be secular, neutral, and non-ideological. Definitions of these terms, accompanied by clarifying examples, are presented in Appendix A, as adopted in the Governing Board Policy on NAEP Item Development and Review.

F. NAEP must not evaluate or assess personal feelings or family beliefs and attitudes unless such questions are non-intrusive and have a demonstrated relationship to academic achievement.

G. Issues of cost, benefit, appropriateness, and burden shall be carefully considered in determining which questions to include in background questionnaires. These factors must also be considered in determining the frequency with which various questions shall be administered and whether they shall be included in both national and state samples.

H. Background questions that do not differentiate between students or have shown little change over time should be deleted or asked less frequently and to limited samples.

5. Public Access to Background Questions

Since security considerations do not apply, all background questionnaires shall be readily available to parents, teachers, state and local officials, and interested members of the public. Such questionnaires shall be available before
field tests and operational assessments or at any other time members of the public wish to obtain them. Background questions in operational use shall be posted on the Internet prior to each assessment, accompanied by explanations and rationales.

6. Reporting of Background Information

The presentation of background data in official NAEP reports shall be straightforward and impartial. Because of the survey nature of the assessment, reports may show correlations and generate hypotheses, but may not state conclusions as to cause and effect relationships. Any composite indices including demographic and socioeconomic factors shall be presented to the Board for approval before use as reporting categories in NAEP data releases and reports.

Background data should be available for extensive secondary analyses by scholars and researchers, who are responsible for conclusions reached. Responses to background questions shall be presented and tabulated on the Internet, although, if necessary, posting may be delayed for a brief period after release of the principal NAEP results.

Definitions of Secular, Neutral, and Non-ideological Item Review Criteria

From Governing Board Policy on NAEP Item Development and Review—5/18/02

Items shall be secular, neutral, and non-ideological. Neither NAEP nor its questions shall advocate a particular religious belief or political stance. Where appropriate, NAEP questions may deal with religious and political issues in a fair and objective way. The following definitions shall apply to the review of all NAEP test questions, reading passages, and supplementary materials used in the assessment:

**Secular** — NAEP questions will not contain language that advocates or opposes any particular religious views or beliefs, nor will items compare one religion unfavorably to another. However, items may contain references to religions, religious symbolism, or members of religious groups where appropriate.

Examples: The following phrases would be acceptable: “shaped like a Christmas tree,” “religious tolerance is one of the key aspects of a free society,” “Dr. Martin Luther King, Jr. was a Baptist minister,” or “Hinduism is the predominant religion in India.”

**Neutral and Non-ideological** — Items will not advocate for a particular political party or partisan issue, for any specific legislative or electoral result, or for a
single perspective on a controversial issue. An item may ask students to explain both sides of a debate, or it may ask them to analyze an issue, or to explain the arguments of proponents or opponents, without requiring students to endorse personally the position they are describing. Item writers should have the flexibility to develop questions that measure important knowledge and skills without requiring both pro and con responses to every item.

Examples: Students may be asked to compare and contrast positions on states rights, based on excerpts from speeches by X and Y; to analyze the themes of Franklin D. Roosevelt’s first and second inaugural addresses; to identify the purpose of the Monroe Doctrine; or to select a position on the issue of suburban growth and cite evidence to support this position. Or, students may be asked to provide arguments either for or against Woodrow Wilson’s decision to enter World War I. A NAEP question could ask students to summarize the dissenting opinion in a landmark Supreme Court case.

The criteria of neutral and non-ideological also pertain to decisions about the pool of test questions in a subject area, taken as a whole. The Board shall review the entire item pool for a subject area to ensure that it is balanced in terms of the perspectives and issues presented.
Appendix B

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.
APPENDIX A
Acknowledgements

The Ad Hoc Committee on NAEP Background Questions, convened in 2002, was chaired by Governing Board member John H. Stevens. He also served as chairman of the Board’s standing Committee on Reporting and Dissemination, which has responsibility for reviewing all core NAEP background questionnaires and making recommendations on them to the full Board.

The Ad Hoc Committee also included members of the Board’s two other standing Committees—Assessment Development and Standards, Design, and Methodology—with a wide range of backgrounds and perspectives: Amanda Avallone, Dwight Evans, Thomas Fisher, Sheila Ford, Jo Ann Pottorff, and Sister Lourdes Sheehan. The Board Chairman, Darvin Winick, participated in many of the lively discussions that marked the Committee’s deliberations.

Among the many discussants and presenters at the workshop and public forum, we wish to recognize the care and thoughtfulness of Robert Mislevy, of the University of Maryland, and Harold Wenglinsky, of Baruch College of the City University of New York. The comments submitted by Sandra Feldman, president of the American Federation of Teachers, were particularly perceptive and helpful. This project also benefited greatly from the continuing advice and insight of Paul Barton and David Grissmer, both of whom have used NAEP data for many years to understand and explain American education to its public.

The Ad Hoc Committee convened in 2012 was chaired by Terry Holliday and also included seven other Governing Board members: Doris Hicks, Andrew Ho, Brent Houston, Dale Nowlin, Joseph O’Keefe, S.J., Susan Pimentel, and Leticia Van de Putte. Lawrence Feinberg, Assistant Director for Reporting and Analysis, provided staff support to both the 2002 and 2012 committees.
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. [New Foreword, pp. 7 and 9]

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. [Chapter 5, pp. 37-39; also pp. 7 and 9]

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. [p. 13]

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. [pp. 13 and 30]

2. Modules will be prepared for special one-time studies to provide descriptive information on issues of current policy interest. [p.29]
3. A thorough review will be conducted to eliminate duplicative or low-priority questions. Unproductive topics and questions will be dropped. [p. 13]

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. [pp. 12 and 23]

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. [pp. 11 and 23]

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. [pp. 12 and 27]

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. [pp. 13 and 35]

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. [pp. 13 and 30]

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. [p. 37]

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. [p. 39]
Using NAEP Data for Key Education Indicators

As authorized by the Governing Board Policy Statement on NAEP Background Data adopted in 2012, consultants have been preparing an exploratory analysis on using NAEP data for key education indicators. The purpose of this project is to illustrate the usefulness of NAEP in developing a limited number of indicators to represent crucial components of the education system and their interrelationships. The key idea is that instead of starting with background variables and looking for education issues they might address, there should first be a framework of important education policy issues and objectives that can be used to identify relevant background variables.

The work is being undertaken by Alan Ginsburg, former Director of Policy and Program Evaluation at the U.S. Department of Education, and Marshall (Mike) S. Smith, former U.S. Under Secretary of Education and former Dean of the Stanford University Graduate School of Education. Smith chaired the Board’s Expert Panel on Strengthening NAEP Background Questions, which presented its report in February 2012. Ginsburg served as a panel member and executive secretary, and has prepared several other exploratory analyses for the Board.

As explained in the statement of work for the project, an education indicator is an individual or composite statistic that measures progress toward an educational objective and is useful in a policy context. Such objectives are concerned not only with student performance but with the quality, equity, and efficiency of the education system in supporting academic achievement. One possible indicator might be the percentage of 8th grade science students with a teacher who majored or minored in science in college. Others might be the extent of severe absenteeism or the use of technology in science instruction.

At the joint meeting Alan Ginsburg and Mike Smith will discuss the purposes and outcomes of the indicator project and how organizing contextual questionnaires to produce data for education indicators might increase the usefulness and impact of NAEP.
DEVELOPING A NAEP INDICATORS FRAMEWORK:
LESSONS FROM MAJOR INTERNATIONAL AND
DOMESTIC EDUCATION INDICATOR REPORTS

EXECUTIVE SUMMARY

By Alan Ginsburg and Marshall S. Smith

Introduction

This is the first of two reports exploring the use of background data collected by the National Assessment of Educational Progress (NAEP) to develop key education indicators at national, state, and urban district levels. Key indicators are statistics that regularly measure an important condition of education. For example, NAEP can tie to its achievement results the reporting of background conditions on: student attitudes toward learning, motivation and excessive absenteeism; measures of teacher quality; and indicators of the nature of reading and math instruction (e.g., instructional time).

The Government Accountability Office identified three broad purposes of indicators:

• Increase transparency and public awareness.
• Foster civic engagement and collaboration.
• Monitor progress, establish accountability for results, and aid decision-making.

In a NAEP context, indicators would also serve to:

• Identify for each subject assessed (e.g. reading) a set of key indicators, which are derived from the background variables and are continuously monitored.

Specifically, this first report is intended to develop a general indicators framework specifying an organizing structure, potential indicators, measurement criteria and reporting design. The report is based on a review of several major international and domestic data collections and reports produced by organizations other than NAEP:

International

• Organization for Economic Cooperation and Development, Education At a Glance
• International Education Association’s 2011 TIMSS Mathematics Assessment covering grade 4 and 8.
• OECD’s 2009 PISA Report

Domestic

• National Center for Education Statistics, Condition of Education
• Education Week’s Quality Counts
• U.S. Department of Education’s Annual Priority Performance Goals
• National Academy of Sciences' Key National Education Indicators
### Potential NAEP Education Indicators From Which To Select Key Indicators For K-12

<table>
<thead>
<tr>
<th>Locus of Education Activity</th>
<th>Key Drivers</th>
<th>Enablers</th>
<th>Context/Constraints</th>
</tr>
</thead>
</table>
| **Student**                 | - Command of core content, using NAEP scores  
- College readiness levels by age and grade  
- Career readiness (21st century skills)  | - Attended preprimary education  
- Chronic absenteeism  
- Student motivation and belief that hard work is more important than luck  
- Student positive attitudes toward subject  
- Student uses research-based approaches to learning subject  
- Student respect for teacher and visa versa  
- Participation in extra-curricular activities including community service  | - Home learning environment  
- Formal and informal learning outside school – nature of the their neighborhood  |
| **Teacher**                 | - Proportion of teacher evaluations that distinguish them from a basic standard  
- Quality of work that the students have  
- Teachers spend time supporting other teachers  | - Teachers with less than 3-years experience  
- Teachers with mastery-level and current knowledge of content they are teaching  
- Teachers with mastery-level and contemporary knowledge of child and adolescent development  
- Teacher-student interactions that demonstrate high levels and qualities of involvement, language, stimulation, and expansion of thinking and cognition, and sensitivity to students’ perspectives, individual experiences, and backgrounds  
- Teacher student interactions that indicate that teachers respect students  | - Teacher working conditions  
- Average district teacher salary  
- Time teachers spend teaching  
- Teacher has high quality professional development and comprehensive induction programs Quality of the principal  
- Teachers belong to professional learning communities  |
| **School/Classroom**        | - School subject area assessment outcomes  
- School performance rating/ranking within their state  
- Parent satisfaction (on surveys)  
- Completion rates from each kind of school – elementary to middle, middle to high, high to graduate, graduate to college or job?  | - Content of instruction aligned with standards  
- Effective use of technology to support instruction  
- School Climate – whether the school is a learning organization – do teachers work together?  
- Instructional time per subject  
- Engaged instruction in subject  
- Emphasis on continuous improvement on outcomes through both formative and summative assessments aligned with standards  
- Emphasis on continuous improvement of practices of teaching  | - School SES Composition  
- Safe & orderly school climate  
- Teacher-student ratio  
- School resource shortages  
- School lacks key characteristics, coaches for teachers, support systems for students, technology, books  |
| **System (district, state or nation)** | - System core content outcomes  | - Support for implementation of new content standards  
- Alignment of assessment with content standards  
- Accountability with emphasis on continuous improvement  | - K-12 education spending as a share of gross domestic product  
- K-12 spending per student  
- Disparity in resources across districts within states  |
The indicator structure in Exhibit EX-1 is focused primarily around variables at student, teacher, school/classroom and system levels that support learning outcomes across three aspects of education conditions:

- **Results** indicators include student assessment outcomes (such as from NAEP), but also teacher evaluations that reflect student outcomes, and other outcomes such as secondary school completion and parent satisfaction with the school.

- **Enabler indicators** reflect formal learning at different levels of education. These include student exposure to preschool, teachers' knowledge and skills and their ability to apply them to create a challenging and supportive classroom learning environment; and school instructional time and student engagement in the content areas. Enablers also include system policies and regulations at district, state and national levels regarding teacher certification, standards, assessment, and accountability.

- **Context/constraint indicators** reflect factors not readily manipulable by educators but may be changed by policy and funding shifts or proper interventions in the home learning environment. These factors include: learning at home and outside the school in formal and informal settings; factors influencing teacher quality, such as salaries and working conditions; and factors affecting the school learning environment including school safety, climate and class size.

**Indicator Measurement**

A sound measure for an indicator should meet criteria of validity, reliability, and consistency overtime.

**Validity of Indicators.** A valid measure is one that adequately captures the underlying education condition of interest. Combining responses from a number of questions around a topic into a larger comprehensive indicator scale produces richer indicator measures than reporting on a single question, but this approach currently is not used in NAEP background factor analyses. Exhibit EX-2 illustrates a scale developed from TIMSS at grade 4 measuring students’ early numeracy activities before beginning primary school.
Reliability of Indicators. A reliable indicator produces consistent results when repeatedly measuring the same underlying condition. Qualitative responses may be unreliable when sensitive to the position of the respondent. For example, Exhibit EX-3, taken from the NAEP background paper on science (by Alan Friedman and Alan Ginsburg), shows that teachers were more likely to indicate that resources within a school are “not at all available” than were principals in the same school. This is not surprising as it is principals who are responsible for school resource availability.

### Exhibit EX-3 Differences between teacher and school reported responses about science resource availability raise issues of response reliability

<table>
<thead>
<tr>
<th>Science Kits are provided (teacher reported)</th>
<th>Not at all</th>
<th>Small Extent</th>
<th>Moderate extent</th>
<th>Large extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science Kits are provided (school reported)</td>
<td>7</td>
<td>24</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>Science magazines and books are provided (teacher reported)</td>
<td>22</td>
<td>35</td>
<td>33</td>
<td>11</td>
</tr>
<tr>
<td>Science magazines and books are provided (school reported)</td>
<td>2</td>
<td>19</td>
<td>35</td>
<td>44</td>
</tr>
</tbody>
</table>

Source: IEA, TIMSS, 2011
**Consistency over time.** A consistent measure requires using the same measure for an indicator over time. When measures are changed from time period to time period it is unclear whether a change comes about because of a real change in the underlying condition or changes in the measure. The report by the Expert Panel on Strengthening NAEP Background Questions (2012) addressed this issue in its 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.

For example, the Expert Panel found that only one-third of the 2011 questions asking about course offerings provided at least a 6-year trend. No 2011 questions about curriculum or school resources were found on the 2005 or earlier questionnaires.

**Sources of Indicator Data.** The reports that were studied use two ways to obtain indicator data, which differentiate them from NAEP.

First, TIMSS and PISA both conduct a household survey to obtain information directly from parents or guardians about socio-economic status and the home learning environment. TIMSS innovatively combined with PIRLS to develop a joint sample household survey for grade 4 students. The household survey included questions about:

- Early numeracy activities in the home before beginning primary school (See Exhibit EX-2)
- Early literacy activities in the home before beginning primary school
- Amount of exposure to preschool
- Family perception about child’s literacy and numeracy skills before entering primary school
- Family interaction with the child about school work
- Family perceptions about school
- Family literacy environment
- Family SES

A second source of data that is different from NAEP is the pooling of information across different surveys. The Condition of Education and Education at a Glance are drawn almost entirely from data series generated by other surveys. Quality Counts is a state-level amalgam of direct analyses of state policies by Education Week combined with data from other surveys, which prominently features NAEP assessment results.

A form of pooling could be the aligning of NAEP survey questions with international assessment items as illustrated in Exhibit EX-4. The exhibit suggests that at least for U.S. middle schools, only about 12 percent of U.S. principals are having at least some difficulty filling vacancies for mathematics teachers. This compares with other Western English-speaking countries of 41 percent of the principals having difficulty hiring math teachers in Australia, 37 percent in England, and 44 percent in New Zealand.
the same question about vacancies to the NAEP principal survey for mathematics would yield U.S. state-by-state comparisons.

Next Steps: Using the International and Domestic Indicator Framework as a Guide, Develop a NAEP Education Indicators Framework and Provide Examples with Current Data

A second report will be prepared for the Governing Board with a recommended set of Key Indicators and recommended improvements in NAEP data to strengthen indicator measurement or fill indicator gaps. This report will:

- Specify a NAEP Education Indicators Framework for the background variables applicable across cognitive assessments.
- Propose indicators that are research-based and estimable by:
  - offering examples using current NAEP data.
  - suggesting changes to the current NAEP questionnaires.
  - introducing a fundamentally new NAEP questionnaire or drawing data from education surveys other than NAEP.
- Explore opportunities for combining NAEP with international or other NCES indicator-supporting data.
- Explore how NAEP reports could best display a pyramid information approach along the lines of an indicator dashboard.
Marshall (Mike) S. Smith is retired and a Senior Fellow in Education Policy at the Carnegie Foundation for the Advancement of Teaching. He is a board member of a number of non-profit organizations in the San Francisco Bay Area. During the first two years of the Obama administration he served as Senior Counselor to Secretary of Education Arne Duncan and as Director of International Affairs. From 2001-2009 he directed the Education Program at the William and Flora Hewlett Foundation where he focused on developing the Open Education Resources movement, improving instruction, and reforming California’s educational system.

Prior to that, in the Clinton Administration, he was the Undersecretary of Education for seven years responsible for all policy and budget matters. For the last four of those years he also served as the acting deputy secretary, the Education Department’s second-ranked official under Secretary Richard Riley. During the Carter administration, he served as chief of staff to the first secretary for education, Shirley Hufstedler, and assistant commissioner for policy studies in the Office of Education. In the Ford administration he was the director of policy and budget for the National Institute of Education, the education research arm of the U.S. Government. While not in government, he was at different times an associate professor at Harvard and a professor at the University of Wisconsin (at Madison) and at Stanford University. At Stanford, he was also the dean of the School of Education.

Smith has authored a large number of publications on topics varying from computer content analysis, evaluation and research methodology, social and educational inequality, early childhood education, open educational resources, federal policy, standards-based reforms and the use of technology in education in the developed and developing worlds. He is a member of the National Academy of Education and a fellow of the American Academy of Arts and Sciences. He holds bachelor’s, master’s, and doctorate degrees from Harvard.
ALAN GINSBURG

Alan Ginsburg was Director of Policy and Program Studies (retired) within the Office of Planning, Evaluation, and Policy Development at the U.S. Department of Education. He coordinated the Education Department’s Government and Performance Results Act indicators and annual reports to Congress. Ginsburg’s international work includes: Lead Shepherd (chair) of the Human Resources Development Working Group; and chair of the APEC Education Network (EDNET). His international mathematics work about Singapore and other Asian countries is extensively cited by the Common Core State Standards Initiative.

Ginsburg received his Ph.D. in economics from the University of Michigan. He received the Distinguished Presidential Rank Service Award, the federal government’s highest award given to its civil service employees. He also received the American Evaluation Association’s Gunnar Myrdal award for his contributions to the field of evaluation. He has been advisor to Education Week on their annual reports.
Upcoming NAEP Reports as of December 2013

<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>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>2013 Reading and Mathematics, Grade 12</td>
<td>April 2014</td>
</tr>
<tr>
<td><strong>Other NAEP Reports</strong></td>
<td></td>
</tr>
<tr>
<td>Grade 8 Black Male Students: Through the lens of the National Assessment of Educational Progress</td>
<td>February 2014</td>
</tr>
<tr>
<td>Linking NAEP and TIMSS 2011 Mathematics and Science Results for the 8th Grade- (Technical Report)</td>
<td>June 2014</td>
</tr>
<tr>
<td><strong>Other Related Reports from NCES</strong></td>
<td></td>
</tr>
</tbody>
</table>
2014 NCES Assessment Data Release Timeline

LEGEND
- NAEP Report Cards
- NAEP Studies

Grade 8 Black Male Students Report
2013 Reading and Math 12
2011 Linking NAEP and TIMSS (Technical)
Releases in 2013

- 2012 Program for the International Assessment of Adult Competencies (PIAAC)
- 2013 Reading Report Card: Grades 4 and 8
- 2013 Mathematics Report Card: Grades 4 and 8
- 2012 Programme for International Student Assessment (PISA)
- 2013 Reading Report Card: Trial Urban Districts (TUDA): Grades 4 and 8
- 2013 Mathematics Report Card: Trial Urban Districts (TUDA): Grades 4 and 8
Releases in 2014

- Grade 8 Black Male Students: Through the Lens of the National Assessment of Educational Progress
- 2013 Reading and Mathematics Report Card: Grade 12
- Linking NAEP and TIMSS 2011 Mathematics and Science Results for the 8th Grade (Technical Report)

Assessment Data Collection Schedule 2014

- U.S. History: Grade 8
- Civics: Grade 8
- Geography: Grade 8
- Technology and Engineering Literacy: Grade 8
Education Summit for Parent Leaders

**Date:** January 13, 2014

**Location:** Crystal Gateway Marriott Hotel, Arlington, VA

**Summit Planning Group:** Rebecca Gagnon, Terry Mazany, Tonya Miles, Eileen Weiser

**Summit Workshop Advisors:** Shannon Garrison, Doris Hicks, Dale Nowlin

**Summit Workshop Consultants:** Amanda Avallone, Kathi King (former Board members)

**Summit Workshop Facilitators:** Lou Fabrizio, Rebecca Gagnon, Shannon Garrison, Doris Hicks, Andrew Ho, Brent Houston, Hector Ibarra, Father Joseph O’Keefe, Terry Mazany, Tonya Miles, Dale Nowlin, and former Board members Robin Hall, Kim Kozbial-Hess, and Sister Mary Frances Taymans

**Summit Goal:** 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 Summit will enable attendees to use NAEP data and resources to ask the right questions of their education leaders about the status of student achievement and gaps in achievement locally, and to promote productive conversations about what is being done to improve achievement and close achievement gaps.

**General Description:** The day-long event has two major components. The morning plenary will consist of individual speakers and panel presentations from a range of perspectives on the need to raise student achievement and close achievement gaps. In the afternoon, hands-on workshops will familiarize participants with NAEP data and resources and illustrate their potential in framing questions relevant to educational improvement.

**Audience:** 150 parent leaders concerned about education from across the nation, with special attention to the 21 districts participating in the NAEP Trial Urban District Assessment Program. Sessions will be available across the nation via live-streaming internet video and/or live TV and radio coverage, if feasible.

**Confirmed Speakers:** Kati Haycock (the Education Trust); Marc Morial (National Urban League); Janet Murguia (National Council of La Raza); Steve Murdock (Rice University); Charles Payne (University of Chicago); Otha Thornton (National PTA)

**Invited Speakers:** First Lady Michelle Obama; Secretary of Education Arne Duncan.
Using NAEP Data for Key Education Indicators

As authorized by the Governing Board Policy Statement on NAEP Background Data adopted in 2012, consultants have been preparing an exploratory analysis on using NAEP data for key education indicators. The purpose of this project is to illustrate the usefulness of NAEP in developing a limited number of indicators to represent crucial components of the education system and their interrelationships. The key idea is that instead of starting with contextual variables and looking for education issues they might address, there should first be a framework of important education policy issues and objectives that can be used to identify relevant contextual variables.

The work is being undertaken Marshall (Mike) S. Smith, former U.S. Under Secretary of Education and former Dean of the Stanford University Graduate School of Education, and Alan Ginsburg, former Director of Policy and Program Evaluation at the U.S. Department of Education. Smith chaired the Board's Expert Panel on Strengthening NAEP Background Questions, which presented its report in February 2012. Ginsburg served as a panel member and executive secretary, and has prepared several other exploratory analyses for the Board.

As explained in the statement of work for the project, an education indicator is an individual or composite statistic that measures progress toward an educational objective and is useful in a policy context. Such objectives are concerned not only with student performance but with the quality, equity, and efficiency of the education system in supporting academic achievement. One possible indicator might be the percentage of 8th grade science students with a teacher who majored or minored in science in college. Others might be the extent of severe absenteeism or the use of technology in science instruction.

At the working lunch Mike Smith will discuss the indicator project and the implications of using NAEP for this purpose. Board members will have a chance to ask questions and discuss the important issues involved.
Marshall S. (Mike) Smith

Marshall S. (Mike) Smith is retired and a Senior Fellow in Education Policy at the Carnegie Foundation for the Advancement of Teaching. He is a board member of a number of non-profit organizations in the San Francisco Bay Area. During the first two years of the Obama administration he served as Senior Counselor to Secretary of Education Arne Duncan and as Director of International Affairs. From 2001-2009 he directed the Education Program at the William and Flora Hewlett Foundation where he focused on developing the Open Education Resources movement, improving instruction, and reforming California’s educational system.

Prior to that, in the Clinton Administration, he was the Undersecretary of Education for seven years responsible for all policy and budget matters. For the last four of those years he also served as the acting deputy secretary, the Education Department’s second-ranked official under Secretary Richard Riley. During the Carter administration, he served as chief of staff to the first secretary for education, Shirley Hufstedler, and assistant commissioner for policy studies in the Office of Education. In the Ford administration he was the director of policy and budget for the National Institute of Education, the education research arm of the U.S. Government. While not in government, he was at different times an associate professor at Harvard University and a professor at the University of Wisconsin (at Madison) and at Stanford University. At Stanford, he was also the dean of the School of Education.

Smith has authored a large number of publications on topics varying from computer content analysis, evaluation and research methodology, social and educational inequality, early childhood education, open educational resources, federal policy, standards-based reforms and the use of technology in education in the developed and developing worlds. He is a member of the National Academy of Education and a fellow of the American Academy of Arts and Sciences. He holds bachelor’s, master’s, and doctorate degrees from Harvard.
ALAN GINSBURG

Alan Ginsburg was Director of Policy and Program Studies (retired) within the Office of Planning, Evaluation, and Policy Development at the U.S. Department of Education. He coordinated the Education Department’s Government and Performance Results Act indicators and annual reports to Congress. Ginsburg’s international work includes: Lead Shepherd (chair) of the Human Resources Development Working Group; and chair of the APEC Education Network (EDNET). His international mathematics work about Singapore and other Asian countries is extensively cited by the Common Core State Standards Initiative.

Ginsburg received his Ph.D. in economics from the University of Michigan. He received the Distinguished Presidential Rank Service Award, the federal government’s highest award given to its civil service employees. He also received the American Evaluation Association’s Gunnar Myrdal award for his contributions to the field of evaluation. He has been advisor to Education Week on their annual reports.
EXECUTIVE SUMMARY

By Alan Ginsburg and Marshall S. Smith

Introduction

This is the first of two reports exploring the use of background data collected by the National Assessment of Educational Progress (NAEP) to develop key education indicators at national, state, and urban district levels. Key indicators are statistics that regularly measure an important condition of education. For example, NAEP can tie to its achievement results the reporting of background conditions on: student attitudes toward learning, motivation and excessive absenteeism; measures of teacher quality; and indicators of the nature of reading and math instruction (e.g., instructional time).

The Government Accountability Office identified three broad purposes of indicators:
- Increase transparency and public awareness.
- Foster civic engagement and collaboration.
- Monitor progress, establish accountability for results, and aid decision-making.

In a NAEP context, indicators would also serve to:
- Identify for each subject assessed (e.g. reading) a set of key indicators, which are derived from the background variables and are continuously monitored.

Specifically, this first report is intended to develop a general indicators framework specifying an organizing structure, potential indicators, measurement criteria and reporting design. The report is based on a review of several major international and domestic data collections and reports produced by organizations other than NAEP:

International
- Organization for Economic Cooperation and Development, Education At a Glance
- International Education Association’s 2011 TIMSS Mathematics Assessment covering grade 4 and 8.
- OECD’s 2009 PISA Report

Domestic
- National Center for Education Statistics, Condition of Education
- Education Week’s Quality Counts
- U.S. Department of Education’s Annual Priority Performance Goals
- National Academy of Sciences’ Key National Education Indicators
### Potential Indicators by Organizing Structure

#### Exhibit EX-1

<table>
<thead>
<tr>
<th>Locus of Education Activity</th>
<th>Potential NAEP Education Indicators From Which To Select Key Indicators For K-12</th>
<th>Key Drivers</th>
<th>Context/Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
<td>• Command of core content, using NAEP scores&lt;br&gt;• College readiness levels by age and grade&lt;br&gt;• Career readiness (21st century skills)</td>
<td>• Attended preprimary education&lt;br&gt;• Chronic absenteeism&lt;br&gt;• Student motivation and belief that hard work is more important than luck&lt;br&gt;• Student positive attitudes toward subject&lt;br&gt;• Student uses research-based approaches to learning subject&lt;br&gt;• Student respect for teacher and visa versa&lt;br&gt;• Participation in extra-curricular activities including community service</td>
<td>• Home learning environment&lt;br&gt;• Formal and informal learning outside school – nature of the their neighborhood</td>
</tr>
<tr>
<td><strong>Teacher</strong></td>
<td>• Proportion of teacher evaluations that distinguish them from a basic standard&lt;br&gt;• Quality of work that the students have&lt;br&gt;• Teachers spend time supporting other teachers</td>
<td>• Teachers with less than 3-years experience&lt;br&gt;• Teachers with mastery-level and current knowledge of content they are teaching,&lt;br&gt;• Teachers with mastery-level and contemporary knowledge of child and adolescent development&lt;br&gt;• Teacher-student interactions that demonstrate high levels and qualities of involvement, language, stimulation, and expansion of thinking and cognition, and sensitivity to students’ perspectives, individual experiences, and backgrounds&lt;br&gt;• Teacher student interactions that indicate that teachers respect students.</td>
<td>• Teacher working conditions&lt;br&gt;• Average district teacher salary&lt;br&gt;• Time teachers spend teaching&lt;br&gt;• Teacher has high quality professional development and comprehensive induction programs Quality of the principal&lt;br&gt;• Teachers belong to professional learning communities</td>
</tr>
<tr>
<td><strong>School/Classroom</strong></td>
<td>• School subject area assessment outcomes&lt;br&gt;• School performance rating/ranking within their state&lt;br&gt;• Parent satisfaction (on surveys)&lt;br&gt;• Completion rates from each kind of school – elementary to middle, middle to high, high to graduate, graduate to college or job?</td>
<td>• Content of instruction aligned with standards&lt;br&gt;• Effective use of technology to support instruction&lt;br&gt;• School Climate – whether the school is a learning organization – do teachers work together?&lt;br&gt;• Instructional time per subject&lt;br&gt;• Engaged instruction in subject&lt;br&gt;• Emphasis on continuous improvement on outcomes through both formative and summative assessments aligned with standards&lt;br&gt;• Emphasis on continuous improvement of practices of teaching</td>
<td>• School SES Composition&lt;br&gt;• Safe &amp; orderly school climate&lt;br&gt;• Teacher-student ratio&lt;br&gt;• School resource shortages&lt;br&gt;• School lacks key characteristics, coaches for teachers, support systems for students, technology, books</td>
</tr>
<tr>
<td><strong>System district, state or nation</strong></td>
<td>• System core content outcomes</td>
<td>• Support for implementation of new content standards&lt;br&gt;• Alignment of assessment with content standards&lt;br&gt;• Accountability with emphasis on continuous improvement</td>
<td>• K-12 education spending as a share of gross domestic product&lt;br&gt;• K-12 spending per student&lt;br&gt;• Disparity in resources across districts within states</td>
</tr>
</tbody>
</table>
The indicator structure in Exhibit EX-1 is focused primarily around variables at student, teacher, school/classroom and system levels that support learning outcomes across three aspects of education conditions:

- **Results** indicators include student assessment outcomes (such as from NAEP), but also teacher evaluations that reflect student outcomes, and other outcomes such as secondary school completion and parent satisfaction with the school.

- **Enabler indicators** reflect formal learning at different levels of education. These include student exposure to preschool, teachers' knowledge and skills and their ability to apply them to create a challenging and supportive classroom learning environment; and school instructional time and student engagement in the content areas. Enablers also include system policies and regulations at district, state and national levels regarding teacher certification, standards, assessment, and accountability.

- **Context/constraint indicators** reflect factors not readily manipulable by educators but may be changed by policy and funding shifts or proper interventions in the home learning environment. These factors include: learning at home and outside the school in formal and informal settings; factors influencing teacher quality, such as salaries and working conditions; and factors affecting the school learning environment including school safety, climate and class size.

**Indicator Measurement**

A sound measure for an indicator should meet criteria of validity, reliability, and consistency overtime.

**Validity of Indicators.** A valid measure is one that adequately captures the underlying education condition of interest. Combining responses from a number of questions around a topic into a larger comprehensive indicator scale produces richer indicator measures than reporting on a single question, but this approach currently is not used in NAEP background factor analyses. Exhibit EX-2 illustrates a scale developed from TIMSS at grade 4 measuring students' early numeracy activities before beginning primary school.
Reliability of Indicators. A reliable indicator produces consistent results when repeatedly measuring the same underlying condition. Qualitative responses may be unreliable when sensitive to the position of the respondent. For example, Exhibit EX-3, taken from the NAEP background paper on science (by Alan Friedman and Alan Ginsburg), shows that teachers were more likely to indicate that resources within a school are “not at all available” than were principals in the same school. This is not surprising as it is principals who are responsible for school resource availability.

### Exhibit EX-3 Differences between teacher and school reported responses about science resource availability raise issues of response reliability

<table>
<thead>
<tr>
<th>Science Kits are provided (teacher reported)</th>
<th>Not at all</th>
<th>Small Extent</th>
<th>Moderate extent</th>
<th>Large extent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science Kits are provided (school reported)</td>
<td>26</td>
<td>30</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td>Science magazines and books are provided (teacher reported)</td>
<td>7</td>
<td>24</td>
<td>32</td>
<td>37</td>
</tr>
<tr>
<td>Science magazines and books are provided (school reported)</td>
<td>22</td>
<td>35</td>
<td>33</td>
<td>11</td>
</tr>
<tr>
<td>NAEP Data Explorer</td>
<td>2</td>
<td>19</td>
<td>35</td>
<td>44</td>
</tr>
</tbody>
</table>
**Consistency over time.** A consistent measure requires using the same measure for an indicator over time. When measures are changed from time period to time period it is unclear whether a change comes about because of a real change in the underlying condition or changes in the measure. The report by the Expert Panel on Strengthening NAEP Background Questions (2012) addressed this issue in its 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.

For example, the Expert Panel found that only one-third of the 2011 questions asking about course offerings provided at least a 6-year trend. No 2011 questions about curriculum or school resources were found on the 2005 or earlier questionnaires.

**Sources of Indicator Data.** The reports that were studied use two ways to obtain indicator data, which differentiate them from NAEP.

First, TIMSS and PISA both conduct a household survey to obtain information directly from parents or guardians about socio-economic status and the home learning environment. TIMSS innovatively combined with PIRLS to develop a joint sample household survey for grade 4 students. The household survey included questions about:

- Early numeracy activities in the home before beginning primary school (See Exhibit EX-2)
- Early literacy activities in the home before beginning primary school
- Amount of exposure to preschool
- Family perception about child’s literacy and numeracy skills before entering primary school
- Family interaction with the child about school work
- Family perceptions about school
- Family literacy environment
- Family SES

A second source of data that is different from NAEP is the pooling of information across different surveys. The Condition of Education and Education at a Glance are drawn almost entirely from data series generated by other surveys. Quality Counts is a state-level amalgam of direct analyses of state policies by Education Week combined with data from other surveys, which prominently features NAEP assessment results.

A form of pooling could be the aligning of NAEP survey questions with international assessment items as illustrated in Exhibit EX-4. The exhibit suggests that at least for U.S. middle schools, only about 12 percent of U.S. principals are having at least some difficulty filling vacancies for mathematics teachers. This compares with other Western English-speaking countries of 41 percent of the principals having difficulty hiring math teachers in Australia, 37 percent in England, and 44 percent in New Zealand. Adding
the same question about vacancies to the NAEP principal survey for mathematics would yield U.S. state-by-state comparisons.

### Exhibit EX-4

<p>| Schools Having Difficulties Filling Vacancies With Mathematics Teachers, Grade 8 |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Country</th>
<th>No Vacancies</th>
<th>Vacancies Are Easy To Fill</th>
<th>Vacancies Are Somewhat Difficult To Fill</th>
<th>Vacancies Are Very Difficult To Fill</th>
<th>Total of Vacancies Somewhat or Very Difficult To Fill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent of students</td>
<td>Percent of students</td>
<td>Percent of students</td>
<td>Percent of students</td>
<td>Percent of students</td>
</tr>
<tr>
<td>Australia</td>
<td>25</td>
<td>34</td>
<td>31</td>
<td>10</td>
<td>41</td>
</tr>
<tr>
<td>Chinese Taipei</td>
<td>46</td>
<td>44</td>
<td>10</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>England</td>
<td>28</td>
<td>35</td>
<td>27</td>
<td>10</td>
<td>37</td>
</tr>
<tr>
<td>Finland</td>
<td>42</td>
<td>46</td>
<td>10</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Hong Kong SAR</td>
<td>48</td>
<td>44</td>
<td>8</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Japan</td>
<td>82</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Korea, Rep. of</td>
<td>67</td>
<td>16</td>
<td>15</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>New Zealand</td>
<td>30</td>
<td>27</td>
<td>38</td>
<td>6</td>
<td>44</td>
</tr>
<tr>
<td>Norway</td>
<td>38</td>
<td>40</td>
<td>20</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>81</td>
<td>11</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Singapore</td>
<td>59</td>
<td>38</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>United States</td>
<td>63</td>
<td>25</td>
<td>9</td>
<td>3</td>
<td>12</td>
</tr>
</tbody>
</table>

Source: 2011 TIMSS, Mathematics

### Next Steps: Using the International and Domestic Indicator Framework as a Guide, Develop a NAEP Education Indicators Framework and Provide Examples with Current Data

A second report will be prepared for the Governing Board with a recommended set of Key Indicators and recommended improvements in NAEP data to strengthen indicator measurement or fill indicator gaps. This report will:

- Specify a NAEP Education Indicators Framework for the background variables applicable across cognitive assessments.
- Propose indicators that are research-based and estimable by:
  - offering examples using current NAEP data.
  - suggesting changes to the current NAEP questionnaires.
  - introducing a fundamentally new NAEP questionnaire or drawing data from education surveys other than NAEP.
- Explore opportunities for combining NAEP with international or other NCES indicator-supporting data.
- Explore how NAEP reports could best display a pyramid information approach along the lines of an indicator dashboard.
Biography: CCSSO Executive Director Chris Minnich

Chris Minnich assumed his role as Executive Director of the Council of Chief State School Officers (CCSSO) in December, 2012. Chris has worked at CCSSO since 2008, when he was hired to direct the Council's standards program—the work that would later become the Common Core State Standards (Common Core). In 2009 he assumed the role of Strategic Initiative Director of Standards, Assessment and Accountability, where he led the development and adoption of the Common Core in 45 states. In 2010, Minnich became the Senior Membership Director, where he has served as the lead contact for all 57 of CCSSO's members. In the last two years, Minnich has led the Council's advocacy and communications teams in rallying states to collaboratively reform their standards and accountability systems. His efforts have been invaluable in advancing the national implementation of the Common Core, and cultivating essential cooperation between states and the federal government to improve education for all students.

Minnich has an extensive background in assessment and accountability work. From 2005-2008, he held multiple positions at Harcourt (now Pearson), all focused on the advancement and improvement of assessments. Minnich led the development and deployment of a teacher-centered online portal focused on assessment education. As the Director of Test Design and Implementation at the Oregon Department of Education from 2003-2005, Minnich led the statewide, online assessment implementation for the Oregon Department of Education.

Minnich holds a Bachelor of Science in Political Science from the University of Washington, Seattle, as well as a Masters of Public Policy from the University of Maryland, College Park. Minnich lives in Alexandria, VA with his wife Whitney and his son Carson.
October 1, 2013

Dear Colleague,

Chief state school officers have demonstrated strong leadership in adopting new standards that are designed to make sure all students graduate from high school with the knowledge and skills necessary for success in college and careers. The new standards challenge our students to develop a deeper understanding of subject matter, learn how to think critically, and apply what they are learning to the real world. States are working hard to implement these new standards and raise the bar for students across the country.

To realize the full potential of these standards, states are designing new assessments to match the rigor of their college- and career-ready standards and to accurately measure student progress toward readiness. The underlying purpose behind new assessments is to help to inform better teaching and learning. These new assessments will not be limited to surface level knowledge — they will better assess the deep knowledge students need to succeed post-graduation.

It is the right and responsibility of each state to determine the appropriate path for ensuring its students are taking high-quality assessments. Many states have joined an assessment consortium, either the Partnership for Assessment of Readiness of College and Careers (PARCC) or Smarter Balanced, to work together to build new, high-quality assessments. The states working together on these assessments have demonstrated significant promise in creating new assessments aligned to the new standards and will field test these assessments starting in the spring of 2014. We applaud these efforts and will continue to work to strongly support these states as they refine and finalize the new assessments. Other states will select a different path as they transition to new assessments aligned to their new standards, and we remain supportive of these states and are working with them to ensure all students have an opportunity to know if they are on track for success in college and careers.

Regardless of the approach, all states must remain committed to ensuring that their students are taking high-quality assessments. Each consortium or state will go through a rigorous process to ensure the assessments are truly measuring student progress. States will adhere to a set of principles to ensure the assessments they select are meeting the high-bar they expect. The principles included herein are intended to be used as a tool to help states hold themselves and their assessments accountable for high quality.

States took a significant step in adopting college- and career-ready standards that raised the bar for students. Moving to new assessments that accurately measure the standards is the next step. States are
further exerting their leadership by committing to this transition to new, high-quality assessments to ensure every student graduates from high school prepared for their future.

Sincerely,

Chris Minnich
Executive Director, Council of Chief State School Officers
States’ Commitment to High-Quality Assessments Aligned to College- and Career-Readiness

Introduction

CCSSO, on behalf of the states, hereby commits to further states’ proactive leadership in promoting college and career readiness for all students by establishing or adopting high quality systems of assessments, including both formative and summative assessments, based on college- and career-ready (CCR) standards. These assessments will align to CCR standards, report annually on each student’s progress toward readiness, adhere to best test administration practices, and be accessible to all students. Many states are already demonstrating leadership in moving in this direction by developing higher-quality college- and career-ready assessments through participation in the Partnership for Assessment of Readiness for College and Careers (PARCC) or Smarter Balanced assessment consortia; other states are also developing new CCR assessments. Our intent is to ensure that CCR-aligned assessments support an education that prepares all students for success in college and careers.

As part of states’ comprehensive assessment systems, states are moving toward systems that improve upon traditional assessments so that students’ in-depth knowledge can be measured. These assessment systems will: assess higher-order cognitive skills; assess critical abilities with high-fidelity; be internationally benchmarked; be instructionally sensitive and educationally valuable; and be valid, reliable, and fair.1 States’ ability to implement systems of assessment that meet these criteria is evolving, along with the technology that is used as part of instructional practice to teach and assess these types of college and career skills and knowledge.

High-Quality Summative Assessment Principles for ELA/Literacy and Mathematics Assessments Aligned to College- and Career-Readiness Standards2

To ensure that all students have access to an education that prepares them for college and careers, summative assessments in grades 3 – 8 and high school should:

1. **ALIGN to CCR standards, by**

   —in **ENGLISH LANGUAGE ARTS / LITERACY ASSESSMENTS**:

   A. **ASSESSING STUDENT READING AND WRITING ACHIEVEMENT IN BOTH ELA AND LITERACY:**
   
   The assessments are English language arts and literacy tests that are based on an aligned balance of literacy and informational texts.

---

1 These criteria are taken from a June 2013 report “Criteria for High-Quality Assessment,” written by Linda Darling-Hammond, Joan Herman, James Pellegrino, Jamal Abedi, J. Lawrence Aber, Eva Baker, Randy Bennett, Edmund Gordon, Edward Haertel, Kenji Hakuta, Andrew Ho, Robert Lee Linn, P. David Pearson, James Popham, Lauren Resnick, Alan H. Schoenfeld, and Richard Shavelson. Click here to access the report.

2 Federal review of assessments should remain limited solely to summative assessments. While the principles set forth herein should apply to all assessments, including formative, the principles are intended to provide guidance to the US Department of Education (USED) on its peer review process for summative assessments only.
B. **FOCUSING ON COMPLEXITY OF TEXTS:** The assessments require appropriate levels of text complexity; they raise the bar for text complexity each year so students are ready for the demands of college- and career-level reading no later than the end of high school. Multiple forms of text are assessed, including written, audio, visual, and graphic as technology permits.

C. **REQUIRING STUDENTS TO READ CLOSELY AND USE EVIDENCE FROM TEXTS:** The assessments consist of reading and writing test questions, tasks, and/or prompts, as appropriate, that demand that students read carefully and deeply and use specific evidence from increasingly complex texts to obtain and defend correct responses.

D. **REQUIRING A RANGE OF COGNITIVE DEMAND:** The assessments require students to demonstrate a range of higher-order, analytical thinking and performance skills in reading, writing, and research based on the depth and complexity of CCR standards, allowing robust information to be gathered for students with varied levels of achievement. Assessments should have a significant portion of total score points come from items that demonstrate a deeper level of knowledge (i.e., represent the high complexity levels designated by taxonomies of cognitive demand).

E. **EMPHASIZING WRITING THAT DEMONSTRATES PROFICIENCY IN THE USE OF LANGUAGE, INCLUDING VOCABULARY AND CONVENTIONS:** The assessments require students to demonstrate college- and career-ready abilities in writing, vocabulary knowledge and tools, and the use of language and its conventions.

F. **ASSESSING RESEARCH AND INQUIRY:** The assessments require students to demonstrate research and inquiry skills, demonstrated by the ability to find, process, synthesize, organize, and use information from sources.

G. **ASSESSING SPEAKING AND LISTENING:** Over time, and as assessment advances allow, the assessments measure the speaking and listening communication skills students need for college and career readiness.

— *In MATHEMATICS ASSESSMENTS:*

H. **FOCUSING STRONGLY ON THE CONTENT MOST NEEDED FOR SUCCESS IN LATER MATHEMATICS:** The assessments help educators keep students on track to readiness by focusing strongly on the content most needed in each grade or course to pave the way for later mathematics. In a CCR-aligned assessment system, the elementary grades focus strongly on arithmetic; the middle grades focus strongly on ratio, proportional relationships, pre-algebra, and algebra; and high school focuses on widely applicable prerequisites for careers and postsecondary education.

I. **ASSESSING A BALANCE OF CONCEPTS, PROCEDURES, AND APPLICATIONS:** The assessments measure conceptual understanding, fluency and procedural skill, and application of mathematics, as set out in CCR standards.

J. **CONNECTING PRACTICES TO CONTENT:** The assessments include brief questions and also longer questions that connect the most important mathematical content of the grade or course to mathematical practices, such as reasoning and modeling.
K. REQUIRING A RANGE OF COGNITIVE DEMAND: The assessments require students to demonstrate a range of performance based on the depth and complexity of CCR standards, allowing robust information to be gathered for students with varied levels of achievement. Assessments include questions, tasks, and/or prompts, as appropriate, about the basic content of the grade or course as well as questions that reflect the complex challenge of CCR standards. Assessments should have a significant portion of total score points come from items that demonstrate a deeper level of knowledge (i.e., represent the high complexity levels designated by taxonomies of cognitive demand).

2. Yield valuable REPORTS ON STUDENT PROGRESS, by:
   A. FOCUSING ON PROGRESS TO READINESS: Score reports illustrate a student’s progress on the continuum toward college and career readiness, grade by grade, and course by course. Reports stress the most important content, skills, and processes and show how the assessment focuses on them.
   B. PROVIDING TIMELY DATA THAT INFORMS INSTRUCTION: Reports are instructionally valuable, are easy to understand by all audiences, and are delivered in time to provide useful, actionable data to students, parents, and teachers.

3. Adhere to best practices in TEST ADMINISTRATION, by:
   A. MAINTAINING NECESSARY STANDARDIZATION AND ENSURING TEST SECURITY: In order to ensure the validity, fairness, and integrity of state test results, the assessment systems maintain the security of the items and tests as well as the answer documents and related ancillary materials that result from test administrations.

4. Provide ACCESSIBILITY to all students, by:
   A. FOLLOWING THE PRINCIPLES OF UNIVERSAL DESIGN: The assessments are developed in accordance with the principles of universal design and sound testing practice, so that the testing interface, whether paper- or technology-based, does not impede student performance.
   B. OFFERING APPROPRIATE ACCOMMODATIONS AND MODIFICATIONS: Allowable accommodations that maintain the constructs being assessed are offered where feasible and appropriate. Decisions about accessibility are based on individual student needs.
National Assessment Governing Board  
Council of Chief State School Officers  
Policy Task Force

**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 35 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 December 6, 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 (2013-2014)**

**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

**Vincent Dean**  
Director, Office of Standards and Assessment  
Michigan Department of Education  
Lansing, MI

**Pete Goldschmidt**  
Director of Assessment  
New Mexico Public Education Department  
Santa Fe, NM

**Kelli Gauthier**  
Director of Communications  
Tennessee Department of Education  
Nashville, TN

**Susie Morrison**  
Deputy Superintendent/Chief of Staff  
Illinois State Board of Education  
Springfield, IL

**Michael Muenks**  
Coordinator of Assessment, Office of College and Career Readiness  
Missouri Department of Elementary and Secondary Education  
Jefferson City, MO

**Nate Olson**  
Communications Manager  
Washington Office of Superintendent of Public Instruction  
Olympia, WA

**Joyce Zurkowski**  
Director of Student Assessment  
Colorado Department of Education  
Denver, CO

*The Task Force is currently recruiting to fill one deputy slot vacancy.*
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 on-going 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 2012, are denoted in yellow highlighting. Topics with new Task Force discussion points are:

- Topic #3: NAEP Schedule of Assessments
- Topic #4: Grade 12 NAEP
- Topic #6: NAEP Background Questions
- Topic #7: Common Core State Standards

* 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>
</table>
| **Executive summary of NAEP Report Cards**  
Increase audience attention to the Executive Summary by conveying more clearly the key findings of each Report Card. | 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  
▫ New online Report Card addresses the Task Force recommendations |
| **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. | *  
The Board is monitoring NCES-led efforts to improve socioeconomic status measures. Some of these efforts were piloted in the 2010 assessment administrations. The 2013 recommendations from the Ad Hoc Committee on Background Information also addresses prospective improvements. |
| **Shaping development of Report Cards**  
Identify questions the data should attempt to answer. Use these questions and question types to shape Report Cards. | *  
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*.  
Starting in 2013, Report Cards have been redesigned in order to take advantage of web-based delivery and interactivity. |
### 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>General NAEP talking points as a resource</strong>&lt;br&gt;Develop 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>*&lt;br&gt;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. The Board’s communications contractor Reingold has conducted an audit of Board communications. The Board adopted a communications plan related to the audit’s findings in November 2010.</td>
</tr>
<tr>
<td><strong>Report-Card-specific talking points</strong>&lt;br&gt;Develop talking points to convey the “common message” in release of specific NAEP Report Cards.</td>
<td>*&lt;br&gt;This resource is provided to NAEP State Coordinators; and is also to be included in executive summaries of Report Cards. 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><strong>Anti-testing sentiment</strong>&lt;br&gt;Address anti-testing sentiment, highlighting how districts and schools benefit from NAEP.</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><strong>Social media</strong>&lt;br&gt;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>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>
</tbody>
</table>
## 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>&lt;br&gt;Prioritize target audiences, highlighting policy and business audiences.</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. 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>
<tr>
<td><strong>Accessible informational resources</strong>&lt;br&gt;Consider a more accessible format for the NAEP Data Explorer, such as an iPhone app.</td>
<td>*&lt;br&gt;NCES and Board staffs are working to develop ideas for apps using NAEP data and items.</td>
</tr>
<tr>
<td><strong>Improving outreach to schools</strong>&lt;br&gt;  • 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.&lt;br&gt;  • Carefully craft messages about the intended use and purpose of NAEP.&lt;br&gt;  • Emphasize information beyond NAEP scores (e.g., profiles on what students are able to do and should be able to do).&lt;br&gt;  • Recognize that as more data are made available, there is a greater possibility for misinterpretation and inappropriate use of these data.&lt;br&gt;  • Maintain a balance between making NAEP meaningful as a significant driver of policy, and making NAEP meaningful at the school level.&lt;br&gt;  • Clarify the benefits of NAEP from the school perspective.&lt;br&gt;  • 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>*&lt;br&gt;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. &lt;br&gt;The Board’s Information for Educators initiative will be implemented beginning in 2014.</td>
</tr>
<tr>
<td><strong>Enhancing outreach to states</strong>&lt;br&gt;The Task Force provided the following suggestions on state use of NAEP information beyond score reports:&lt;br&gt;  • Strategically engage state or consortia assessment programs to create a shared knowledge base and collaboration around lessons learned (e.g., webinars, users’ groups).&lt;br&gt;  • 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).&lt;br&gt;  • Introduce NAEP data tools to educators during pre-service as a resource for teaching data use.</td>
<td>*&lt;br&gt;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. &lt;br&gt;Senior Board staff met with consortia leaders in January 2011 to explore concrete areas for collaboration between NAEP and the consortia. The group meets for periodic follow-up regularly.</td>
</tr>
</tbody>
</table>
## Topic#1: NAEP Reporting Process

<table>
<thead>
<tr>
<th>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
</table>
| **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. | For 2011, chiefs and governors in-person briefings were changed to notification of the embargoed release website with question-and-answer follow-up on request.  
In 2013, the Board implemented online briefings for chiefs. |
| **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. | ✭  
See summary of Board action on this topic below. |
| **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.  
The Reporting and Dissemination Committee will be reviewing the Board’s embargo policy guidelines in light of the changing media landscape. |
# 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. 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>
### Task Force Discussion and Input

**Within-year and across-year variations**  
Ensure that the Ad Hoc Committee addresses the variations in exclusion rates within states over time, as well as variations among states in a given year.

**Determining inclusion for each student**  
- 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.
  - Create a guidance document to support state development of IEP templates.

**The current decision trees**  
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.
  - Start with a least invasive approach, and then scale up if needed.
  - Examine how current policy can align with Title I.
  - Examine how prospective changes will be communicated to IEP teams.

In considering the possibility of uniform national rules:
  - Uniform national rules may have unintended negative consequences if more students opt out. Participation and inclusion rates may decrease.
  - Consistency in inclusion is important to maximize comparability. Using the language screener would allow case-by-case determinations for each student.
  - Policy guidance in defining ELL subgroups could provide comparability.

### Follow-up Activities

In the policy adopted by the Board on March 6, 2010:

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.

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.

In the policy adopted by the Board:

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.
### 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>Feasibility of new accommodations</strong>&lt;br&gt;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>*&lt;br&gt;In the policy adopted by the Board:&lt;br&gt;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>
<tr>
<td><strong>Prominence of exclusion rate data</strong>&lt;br&gt;- 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.&lt;br&gt;- 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.&lt;br&gt;- Identify states’ distance from the 85% inclusion goal to motivate improvement.</td>
<td>In the policy adopted by the Board:&lt;br&gt;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><strong>Guidance for states</strong>&lt;br&gt;Provide compelling guidance on this issue, similar to how NCLB regulations prompted states to align with 1% and 2% waivers for students with disabilities.</td>
<td>In the policy adopted by the Board:&lt;br&gt;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.</td>
</tr>
</tbody>
</table>
### 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>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>
</tbody>
</table>

| Policy suggestions for the Ad Hoc Committee |  |
|---------------------------------------------|  |
| - Adopt a guiding principle that includes language about maximizing meaningful participation in order to provide the most accurate assessment of student performance. | In the policy adopted by the Board: |
| - Focus on maintaining the purpose of NAEP. | 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. |
| - Encourage NAEP to accommodate to students. |  |
| - Distinguish modifications, which change what is being measured, from accommodations. |  |

| Feedback on preliminary recommendations of SD Expert Panel |  |
|-----------------------------------------------------------|  |
| - Clarify whether the 95% participation rate mentioned in the fourth recommendation includes or excludes students with significant cognitive disabilities. | In the policy adopted by the Board: |
| - 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. | 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. |
| - 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. |  |

| 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. | In the policy adopted by the Board: |
| - 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. | Support is provided for data displays of ELLs and former ELLs. |
## Topic#2: NAEP Inclusion and Accommodations

### Task Force Discussion and Input

**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 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.

### 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.

### Follow-up Activities

- 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.

* 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.

In December 2013, the Committee on Standards, Design and Methodology will meet jointly with the Reporting and Dissemination Committee to address implementation issues related to the Board’s policy on students with disabilities and English language learners.
**TOPIC #3: NAEP SCHEDULE OF ASSESSMENTS**


<table>
<thead>
<tr>
<th>TASK FORCE DISCUSSION AND INPUT</th>
<th>FOLLOW-UP ACTIVITIES</th>
</tr>
</thead>
</table>
| **Key subjects and grades for states**  
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. | 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. |
| **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.  
In August 2013, the Board acted to reduce the frequency of grade 12 NAEP. |
| **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. | *  
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. |
**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><strong>Considerations on Computer-Based Grade 4 Writing</strong></td>
<td></td>
</tr>
<tr>
<td>- 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:</td>
<td></td>
</tr>
<tr>
<td>- 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.</td>
<td></td>
</tr>
<tr>
<td>- Developmental concerns. Some teachers are concerned that prior to grade 5, students may lack the developmental ability to compose writing on a computer.</td>
<td></td>
</tr>
<tr>
<td>- 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.</td>
<td></td>
</tr>
<tr>
<td><strong>Planning the NAEP Schedule through 2022</strong></td>
<td></td>
</tr>
<tr>
<td>- <strong>ESEA subjects.</strong> 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.</td>
<td></td>
</tr>
<tr>
<td>- <strong>NAEP reading and mathematics.</strong> 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.</td>
<td></td>
</tr>
<tr>
<td>- <strong>Subjects other than NAEP reading and math.</strong> 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.</td>
<td></td>
</tr>
<tr>
<td>- <strong>Common Core-NAEP alignment.</strong> 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.</td>
<td></td>
</tr>
<tr>
<td>- <strong>State readiness for computer-based delivery.</strong> 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.</td>
<td></td>
</tr>
</tbody>
</table>
**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><em>Priorities for the schedule given the changing funding landscape</em></td>
<td><em>The Board is considering the Task Force’s feedback in ongoing discussions of the NAEP budget situation.</em></td>
</tr>
<tr>
<td>• The Task Force reiterated the value of having NAEP assessments that the state is not able to do on its own, including the computer-based assessments. This is an important benefit from NAEP.</td>
<td></td>
</tr>
<tr>
<td>• If it is necessary to prioritize due to limited resources, 12th grade assessments are less of a priority. For 12th grade assessment, there are questions about the motivations of the students and how the information is used. In addition, the 12th grade assessment will be less of a priority as states implement the new college and career-ready assessments. There is potential that states could be sending mixed messages to the public with both assessments.</td>
<td></td>
</tr>
<tr>
<td>• Other priorities are the new NAEP Technology and Engineering Literacy (TEL) assessment. This is an important issue and area to emphasize. Ideally, TEL could be expanded and states could have state-level results.</td>
<td></td>
</tr>
<tr>
<td>• Information from NAEP in new areas, such as the grade 4 computer based writing pilot, is especially useful in informing related state initiatives.</td>
<td></td>
</tr>
<tr>
<td>• A foreign language assessment will likely be more appealing to some states than others.</td>
<td></td>
</tr>
<tr>
<td>□ For some states a foreign language assessment is very appealing. If done, it would need to be highly targeted to specific students.</td>
<td></td>
</tr>
<tr>
<td>□ If this is identified as an area of expansion, however, NAGB may want to consider whether there is anything more generic that could be assessed as opposed to a specific language.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Considering the overall subject coverage of the NAEP Schedule</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Reducing the sample size may be preferable to completely eliminating an assessment in a subject area. Focal vs. non-focal reporting seems to be a promising strategy.</td>
<td><em>The Board is considering the Task Force’s feedback in ongoing discussions of the NAEP budget situation.</em></td>
</tr>
<tr>
<td>• The Task Force encouraged NAGB to consider the new landscape with the assessment consortia and what will be covered by the assessment consortia.</td>
<td></td>
</tr>
<tr>
<td>□ Given that the assessments in math and English language arts will now be much more comparable across states, the Governing Board should consider how to avoid providing redundant information.</td>
<td></td>
</tr>
<tr>
<td>□ There is value in keeping the NAEP Reading and Mathematics assessments for several years as states make this transition to help states make determinations about comparability. NAEP serves as an important independent indicator. It will be particularly useful to have a fixed assessment for some period of time (perhaps through 2020).</td>
<td></td>
</tr>
<tr>
<td>• Maintain current frequency of NAEP reading and mathematics during at least the transition period toward college and career ready assessments – in light of changing membership in consortia, NAEP’s role as an independent monitor is important.</td>
<td></td>
</tr>
<tr>
<td>• Long-term trend NAEP’s focus solely on basic skills is discordant with the current national context of education—this is subsumed in main NAEP assessments. This justifies reducing the frequency of this assessment, and possibly discontinuing it entirely.</td>
<td></td>
</tr>
</tbody>
</table>
### 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><strong>Low priorities from the state perspective</strong></td>
<td>*</td>
</tr>
<tr>
<td>- National only level reporting has little utility at the state level.</td>
<td>The Board is considering the Task Force’s feedback in ongoing discussions of the NAEP budget situation.</td>
</tr>
<tr>
<td>- From a state perspective, TUDA is less of a priority.</td>
<td>*</td>
</tr>
</tbody>
</table>

**Considerations in selecting the mode of assessment**

- The Governing Board should consider concerns about time and burden of online testing in the coming years given the increased backlash against assessments.
### 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. |
| **Reporting results from the state-level grade 12 NAEP pilot**  
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. | 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. |
<table>
<thead>
<tr>
<th>TASK FORCE DISCUSSION AND INPUT</th>
<th>FOLLOW-UP ACTIVITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pilot status of the grade 12 NAEP state-level assessment</strong>&lt;br&gt;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.</td>
<td>*</td>
</tr>
<tr>
<td><strong>Benchmarking performance at grade 12</strong>&lt;br&gt;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.</td>
<td>*</td>
</tr>
<tr>
<td><strong>Prospective information to collect from examinees</strong>&lt;br&gt;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.</td>
<td>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 just completed its work in exploring enhancements to NAEP contextual information.</td>
</tr>
<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>*</td>
</tr>
<tr>
<td></td>
<td>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. 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>
</tbody>
</table>
## 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>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. 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>
<tr>
<td><strong>Implications for state accountability systems</strong>&lt;br&gt;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.</td>
<td><strong>ɿ</strong>&lt;br&gt;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.</td>
</tr>
<tr>
<td><strong>T</strong>opic#4: Grade 12 NAEP</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Task Force Discussion and Input</strong> &amp; <strong>Follow-up Activities</strong></td>
<td></td>
</tr>
</tbody>
</table>
| *Distinguishing 12th grade preparedness from 12th grade proficient*  
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. This issue has also been revisited in 2013 discussions on prospective NAEP preparedness reporting.* |
| **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.  

**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.  

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.  

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. |
**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>Implications of changing NAEP to a measure of preparedness</strong>&lt;br&gt;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.</td>
<td>*&lt;br&gt;The Board has engaged a commission comprised of Board members and individuals with experience in government, industry, and education. The commission is focused on communication efforts and strategies for the Board’s plans to report grade 12 NAEP results in terms of preparedness.</td>
</tr>
<tr>
<td><strong>Contextual information to consider with preparedness indicators</strong>&lt;br&gt;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.</td>
<td>*</td>
</tr>
<tr>
<td><strong>Clarifying the definition of a 12th grade student</strong>&lt;br&gt;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.</td>
<td>*</td>
</tr>
</tbody>
</table>
| **Considering the Board’s working definition of preparedness**<br>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. | 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. |
**Topic#4: Grade 12 NAEP**

### Task Force Discussion and Input

**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 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.

- **∗**
  - 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.

### 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.

### Suggestions regarding grade 12 NAEP preparedness reporting

- Hold pre-release data-free briefings to facilitate better understanding of the results.
- Emphasize that the first report is just one piece of information, and that additional findings will be released as part of an iterative process.
- Incorporate the use of other measures to determine preparedness, e.g., transcript data.

### Key considerations regarding defining grade 12 NAEP preparedness

- Consider whether differences exist between preparedness and proficiency at grade 12.
- 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—its definition of “readiness” seems to be headed in a different direction compared to the Governing Board’s preparedness initiative.
- Hold joint conversations between NAEP, consortia, and vendors to support comparability. There should be efforts to align definitions of readiness and preparedness.

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.
**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 and readiness terminology</strong>&lt;br&gt;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.</td>
<td>*&lt;br&gt;As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed.</td>
</tr>
</tbody>
</table>

| Strategies for using NAEP academic preparedness data<br>▪ 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.<br▪ Frame the results in terms of establishing a baseline of performance.<br▪ Tie the results to the common goal of Common Core readiness.<br▪ Emphasize the definition of preparedness being used.<br▪ Distinguish between the needs of colleges and careers on a continuum, e.g., two-year, four-year, types of careers.<br▪ Include comparisons with tests administered by the business community to strengthen the validation, e.g., industry certification. | *<br>As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. |

| Task Force feedback on reporting recommendations<br▪ 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.<br▪ Encouraging student tracking. An unintended consequence of readiness and preparedness reporting may be: reinforcing the tracking of students. | *<br>As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. |

| Task force suggestions for next steps in preparedness research and reporting<br▪ Use nuance to avoid making statements that appear to relate to individual students.<br▪ Continue to collaborate on aligning consortia readiness and NAEP preparedness to avoid confusion and validity concerns.<br▪ Consider segmenting the types of colleges and careers that are represented by NAEP performance labels to provide more fine-grained information.<br▪ Continue studying readiness for entry into job training programs in order to make statements about academic preparedness for this area.<br▪ Ensure that the career preparedness discussions are focusing on what current employers expect. | As the Board moves forward with its preparedness research and reporting initiative, several of these ideas have been discussed. |
### Topic#4: Grade 12 NAEP

#### Task Force Discussion and Input

**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?

**Feedback: Communication strategies to avoid misinterpretation of the progress report**

- 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.

- It is important to consider how college readiness is discussed in relation to career readiness.

- 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.

- Discussion among NAEP and the two main assessment consortia (PARCC and Smarter Balanced) should be held soon to address this issue.

#### Follow-up Activities

* 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.

* 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.
**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>Suggested communication strategies</em> (formulated in conjunction with the NAGB Business Policy Task Force in January 2013 in-person meeting)</td>
<td></td>
</tr>
<tr>
<td>▪ Participate in State symposia. Representatives from NAEP could come to the state and present to key stakeholders the story about 12th grade preparedness, for example.</td>
<td></td>
</tr>
<tr>
<td>▪ Engage the statewide superintendents associations. There are many innovative groups of superintendents who are trying to determine how to prepare students for careers even if students do not want to go to college.</td>
<td></td>
</tr>
<tr>
<td>▪ Take advantage of social media and pre-existing networks. One example is the National Parent Teachers Association.</td>
<td></td>
</tr>
<tr>
<td>▪ Engage statewide business groups. Representatives from NAEP could meet with roundtable groups about targeted issues relevant to the groups.</td>
<td></td>
</tr>
<tr>
<td>▪ Highlight the NAEP test questions. When people see the actual questions, it improves their understanding of NAEP.</td>
<td></td>
</tr>
<tr>
<td>▪ Share information through national public television.</td>
<td></td>
</tr>
<tr>
<td>▪ Define the purpose of sharing information. It would be helpful for the Governing Board to articulate a few core points about the purpose of the communications. For example, when someone sees communication from NAEP, what is it NAEP wants them to do (be better informed, to advocate for something specific, etc.)?</td>
<td></td>
</tr>
</tbody>
</table>

*Communication strategies for the 2013 grade 12 NAEP Report Card*

- A clear distillation of the research findings (as well as an executive summary with key takeaways) is needed in the first Report Card with preparedness results.
- Explain how and why the preparedness indicator score for math falls below the proficiency cut score.
- Prepare information packages for state communication directors early so they can start spreading the message in their states—tailoring for states with state pilot results. This should be part of a larger effort in supporting assessment literacy, e.g., clearly distinguishing the role of NAEP versus the role of state assessments.
- Conduct informational briefings with the media early to prepare them directly.
- A video would be especially effective. See the recent NAEP Technology and Engineering Literacy video for a good model.

*Future considerations for the Board’s grade 12 NAEP preparedness reporting and program of research*

- The Board’s definition of preparedness appears to be vague, especially relative to other sources such as ACT. Job training also needs more detail to avoid being interpreted as basic literacy.

*The Board is preparing for an early 2014 release of the first grade 12 NAEP Report Card that includes preparedness information.*
**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>Technical decision and process</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.

**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.

*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. In 2013, the Task Force discussion focused on NAEP’s efforts to refine measures of socioeconomic status and Board initiatives to improve NAEP questionnaires.

### Task Force Discussion and Input

<table>
<thead>
<tr>
<th>Strategic considerations for the transition process</th>
</tr>
</thead>
<tbody>
<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>

<table>
<thead>
<tr>
<th>Suggestions for NAEP implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
</tr>
<tr>
<td>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.</td>
</tr>
<tr>
<td>Distinguish data collection and reporting issues.</td>
</tr>
<tr>
<td>Involve policy and assessment staff in these ongoing conversations to assess full implications beyond data.</td>
</tr>
<tr>
<td>Poll states using the CCSSO network to determine how varied states’ plans are.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting race/ethnicity data under the new OMB guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preserve trend lines by maintaining the old categories in the body of the report and reporting the new categories in the appendices.</td>
</tr>
<tr>
<td>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.</td>
</tr>
</tbody>
</table>

### Follow-up Activities

- *

The Board is monitoring changes in race/ethnicity categories and their potential impact on trends.

- *

The 2011 NAEP Reading and Mathematics Report Cards were released on November 1, 2011. The 2011 reports include data on the new race/ethnicity categories for NAEP reporting.
## Task Force Discussion and Input

### 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.

### 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.

## Follow-up Activities

* The Board convened an expert panel on background questions, and an Ad Hoc Committee has completed its work in exploring enhancements to NAEP contextual information. The Board is considering recommendations from the Ad Hoc Committee on NAEP Background Information. The Board is planning to adopt responsive changes to the framework in December 2013.
## Task Force Discussion and Input

**Task Force feedback on the 2012 expert panel recommendations report**

- **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.

- **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.

- **Purpose of the questionnaires**
  - The purpose of the background questions and analyses should be to illuminate correlations between NAEP student performance and background questions.
  - 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.

**Task Force suggestions for future questionnaires**

- Make better use of existing NAEP background data in an accessible and useful format.
- Improve online NAEP data analysis tools to facilitate use of data by various audiences.
- Consider removing some existing questions to create space for new questions in order to maintain the existing time allotment for background questions.
- Focus some new questions on “career skills” and post-secondary plans. This information would be particularly useful to states (as previously noted).
- 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.
- Explore different disaggregations. Aggregated state-level data may obscure meaningful differences in student performance.

## Follow-up Activities

**∗**

The Board convened an expert panel on background questions, and an Ad Hoc Committee has completed its work in exploring enhancements to NAEP contextual information. The Board is considering recommendations from the Ad Hoc Committee on NAEP Background Information. The Board is planning to adopt responsive changes to the framework in December 2013.
**Task Force Discussion and Input**

*Determining how to measure socioeconomic status (SES)*
- To inform which data are collected, consider: what is the desired decision to improve learning and teaching that will result by changing the definition of SES or by collecting various data related to SES?
- It is important to develop a definition for SES that is strong yet also one for which it is possible to collect valid and reliable data.
- What difference does the SES indicator make? This is an important question to drive NAEP’s questionnaire.
- Avoid confounding the individual (or home) notion of socioeconomic status and community socioeconomic status. If the two components are put into the same variable, analysts lose the ability to disentangle where policy can impact improving schooling and where there are factors upon which a policy impact is not possible.
- Keep community and school indicators separate to disentangle issues such as high SES students in schools with mostly low SES students.
- What composite picture can be developed about whether the data to be collected for SES represent the things that have an influence on student performance? In considering which variables to include in the U.S. background questions, focus on the factors that prepare students or help them to be better prepared (access to a computer, for example).
- Keep NAEP’s SES indicators educationally relevant, e.g., WiFi access and dedicated space for homework, but sensitivity of questions should also be carefully reviewed.

*Aggregating data to guide useful interpretations*
- The high-density poverty measure (at the community or school level) has been used as one way to define high-risk students in some states.
- Comparability should not be assumed between what a low-income person has in the U.S. versus what a low-income person has in other countries, for example.

*Utility of background information for states*
- The background information that is collected is still useful in states. It allows states to make important comparisons that go beyond overall scores to better understand the performance of different subgroups.
- In states with high proportions of low SES students, low SES indicators are not currently providing actionable information — additional specificity is needed.
- Avoid an SES indicator that is so complex that it is not comparable at the state-level.

**Follow-up Activities**

* The Board is considering recommendations from the Ad Hoc Committee on NAEP Background Information. The Board is planning to adopt responsive changes to the framework in December 2013.

* The Board is considering recommendations from the Ad Hoc Committee on NAEP Background Information. The Board is planning to adopt responsive changes to the framework in December 2013.
**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>Streamlining data elements collected in the questionnaire</strong></td>
<td></td>
</tr>
<tr>
<td>• The Task Force recommends determining if there is more information that can be obtained from</td>
<td>The Board is considering recommendations from the Ad Hoc Committee on NAEP Background</td>
</tr>
<tr>
<td>other information systems (as opposed to collecting the information during the assessment)</td>
<td>Information. The Board is planning to adopt responsive changes to the framework in</td>
</tr>
<tr>
<td>to reduce the amount of time this adds to the time for the assessment.</td>
<td>December 2013.</td>
</tr>
<tr>
<td>• The Task Force recommends that the Governing Board be especially thoughtful in discussing</td>
<td></td>
</tr>
<tr>
<td>the collection of background information given the current sensitivities on data collection.</td>
<td></td>
</tr>
<tr>
<td>• The Task Force encouraged NAGB to consider how to streamline the information they collect</td>
<td></td>
</tr>
<tr>
<td>directly from students.</td>
<td></td>
</tr>
<tr>
<td>• Coordinate with other federal programs to minimize the burden on students and schools.</td>
<td></td>
</tr>
<tr>
<td>• Align with the “economically disadvantaged” indicator to avoid data mismatches that</td>
<td></td>
</tr>
<tr>
<td>complicate state comparisons between NAEP and other information sources.</td>
<td></td>
</tr>
<tr>
<td>• Questions on out-of-school topics are acceptable, but should be carefully pilot tested.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Improving dissemination of information from background data</strong></td>
<td></td>
</tr>
<tr>
<td>• Disseminate daily or weekly emails with expert summaries, blurbs, or snippets on the</td>
<td></td>
</tr>
<tr>
<td>contextual data that can help principals and test coordinators support student success—</td>
<td></td>
</tr>
<tr>
<td>more attention is needed in helping states make use of this information.</td>
<td></td>
</tr>
<tr>
<td>• The Task Force advocates for shorter briefs on the data so chiefs and others are more</td>
<td></td>
</tr>
<tr>
<td>likely to read them.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>!</td>
</tr>
<tr>
<td></td>
<td>The Board is considering recommendations from the Ad Hoc Committee on NAEP Background</td>
</tr>
<tr>
<td></td>
<td>Information. The Board is planning to adopt responsive changes to the framework in</td>
</tr>
<tr>
<td></td>
<td>December 2013.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>!</td>
</tr>
<tr>
<td></td>
<td>The Board is also planning to prepare focus reports to further disseminate information</td>
</tr>
</tbody>
</table>
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. The most recent annual briefing for the Board from state assessment consortia representatives was at the August 2013 Board meeting. The session included an extended discussion of the accommodations that are being permitted for assessments of reading. In 2013, the Task Force shared its perspectives on the current policies being considered for inclusion, particularly with respect to accommodations.

### Task Force Discussion and Input

**Role of NAEP**
- 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.
- 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.
- International measures could be the key contribution from NAEP that complements the Common Core.

**NAEP communications and informational resources**
- 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.
- Efforts should be pursued to avoid confusion between Report Card releases and releases related to the Common Core Standards initiative.
- Create a compare/contrast document that clarifies similarities and intended purposes of all assessments. Focus on comparison as opposed to “alignment.”

### Follow-up Activities

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.

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.

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 Board meetings.

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 Board meetings.
### Task Force Discussion and Input

*Accommodations being considered by the assessment consortia*
- Differences in accommodation policies could generate data and information that would inform the debate on valid accommodations.
- Student IEPs drive accommodation decisions for many states, within the allowable parameters specified by the U.S. Department of Education (ED).
- Alignment with the construct (defined in advance) should inform decisions on accommodations; if an accommodation redefines a construct for NAEP, it should not be classified as an allowable accommodation for NAEP.
- NAEP should be consistent with ED-supported state assessment accommodations.

### Follow-up Activities

*
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**.

### **TOPIC #8: MISUSE AND MISINTERPRETATION OF NAEP DATA**

#### **Task Force Discussion and Input**

*Resources for countering instances of misuse and misinterpretation*

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.

#### **Definitions of proficient**

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.

- NAEP performance levels are aspirational goals, developed in the early 1990s.
- State performance levels are accountability determinations developed in terms of grade-level performance as part of No Child Left Behind.
- Options:
  - 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.
  - Expand the interpretation of results section in Report Cards to include more explanation regarding differences in proficiency definitions.
- 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.

### **Follow-up Activities**

* At the August 2009 Board meeting, COSDAM received a briefing from NCES on related efforts they are spearheading.

At the August 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.

The NAEP Validity Studies (NVS) panel, an expert advisory group to NCES, has initiated efforts to develop materials that address interpretation issues.

* 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.
### Topic#8: Misuse and Misinterpretation of NAEP Data

#### Task Force Discussion and Input

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.

#### Follow-up Activities

* As a new type of report, the Board has outlined a mega-states special analytical report that would include states with the largest public school enrollment. The primarily internet-based report would feature new data displays that can be considered for future NAEP Report Cards.

The Mega-States report was released in early 2013.

Starting in 2013, Report Cards have been redesigned in order to take advantage of web-based delivery and interactivity.
## 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.

### 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 adopted 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 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.

### 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.
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>Task Force Discussion and Input</th>
<th>Follow-up Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Considerations for possible future roles of NAEP</strong></td>
<td>In November 2009, the Board adopted a resolution supporting international linking projects for NAEP.</td>
</tr>
<tr>
<td>▪ NAEP can be used as the international benchmark.</td>
<td></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>
<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>▪ Embed non-secured NAEP items on state assessments to be used as a set of anchor data to determine alignment capabilities.</td>
<td>Results from the 2011 NAEP/TIMSS Linking Study was released in October 2013.</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>
<td></td>
</tr>
</tbody>
</table>
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

**Task Force Discussion and Input**

- Considerations and suggestions for future initiatives in development
  - Involve states instead of reaching schools directly without burdening state budgets.
  - Identify stakeholder groups that are already engaging parents in community-based efforts to raise awareness of public education issues.
  - Emphasize comparative international standing and achievement gaps to change expectations.
  - 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.
  - 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.
  - Feature sample items to demonstrate test rigor, or make sample tests available.
  - Extract and share lessons about teachers or use of time from the NAEP background questionnaire.
  - 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.
  - Foster collaboration between NCES and the consortia on AI scoring.

**Follow-up Activities**

- 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 Board meetings.

The Board is organizing a Parent Summit for January 2014, which will highlight ways to better engage with and understand NAEP data.

**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).

- 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 Board meetings.

In March 2010 and May 2010, the Board adopted changes to the NAEP Schedule of Assessments to further support international linking projects.
### Considerations for the role of NAEP
- **Stretching the intended role of NAEP.** The Task Force is concerned with the Board’s desire to make NAEP more relevant, given the intended purpose of NAEP. Moving away from NAEP’s purpose may result in complicated messaging and negative media attention for NAEP.
- **Describing best practices at the state-level.** Promoting “best practices” for states is problematic and may be used against some jurisdictions. States are presented in NAEP reporting as homogeneous jurisdictions without sensitivity to differences among states.
- **Highlighting best practices without influencing curricular decisions.** Legislation precludes the use of NAEP to influence curricular decisions; there is a thin line between influencing curriculum and sharing best practices.
- **Policy context shifts affecting NAEP’s role.** The reauthorization of the Elementary and Secondary Education Act could change the role of NAEP in the context of Common Core and consortia assessments.
- **Losing impact in the context of over-testing sentiment.** There is a sentiment of over-testing, and there are several negative reports on student performance, which may be overwhelming for the public.
- **Identifying NAEP’s relevance to parents.** Research findings indicate that international comparisons are not resonating with parents.

### Suggestions for using NAEP data
- Use background information to contextualize what students are learning, as done with the Programme for International Student Assessment (PISA).
- Use NAEP results to identify high-performing student groups and report the results in terms of what is working for groups of students.
- Triangulate NAEP results with results from other large-scale assessment programs external to states (e.g., SAT, ACT, TIMSS, PISA, PIRLS) to answer the questions: What are the data telling us? How can we inform expectations about rigor?

### Suggestions for sharing NAEP data with new audiences
- Enhance pre-service teachers' understanding of NAEP by working with national teacher preparation organizations to promote NAEP’s value early in teachers' careers and leverage the university research-based perspective.
- Build researcher capacity to appropriately use, interpret, and report on best practices.
- Focus the parent initiative on the college preparedness discussion, which is relevant to parents and connects to the Common Core State Standards.

### Follow-up Activities
- The Board developed a list of priority activities and action plans at the December 2011 Board meeting.

- The Board convened an expert panel on contextual questions in November 2011 with a report to the Board in March 2012.

- The Board is organizing a Summit for Parent Leaders in January 2014, which will highlight ways to better engage with and understand NAEP data.
**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>
</table>
| **Suggested improvements for planning the parent initiative**  
  ▪ 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?  
  ▪ 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.  
  ▪ 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. | *  
  The Board used the Task Force’s input to refine the audience and specificity of its final recommendations for implementing the parent initiative. |
| **Task Force suggestions for focus report development**  
  In developing focus reports, the Task Force provided the following suggestions to the Board:  
  ▪ Ensure a direct relationship between the topic and NAEP achievement.  
  ▪ Ensure objectivity of analysis.  
  ▪ 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. | *  
  NCES has several focus reports in the development stage. The Board is considering additional topics for focus reports. |
| **Task Force priority topics for future focus reports**  
  ▪ Charter schools: A 10-year report  
  ▪ Opportunity-to-Learn  
  ▪ Education policies and instructional practices of high-performing or high-growth states and districts  
  ▪ Learning in the South  
  ▪ Other regional reports  
  ▪ Eighth-grade algebra (access to algebra rigor) | * |
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

<table>
<thead>
<tr>
<th>New constructs</th>
<th>Investigate research-based questions and measure new constructs. Students need non-academic constructs to be competitive globally.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reporting relative to state/consortia assessments</td>
<td>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.</td>
</tr>
<tr>
<td>Engaging higher education</td>
<td>Use linking of 12th grade NAEP preparedness with SAT, ACT, state assessments, and state longitudinal databases to initiate conversations with higher education policymakers.</td>
</tr>
<tr>
<td>Career readiness</td>
<td>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.</td>
</tr>
<tr>
<td>Possible ways to use NAEP in state accountability</td>
<td>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.</td>
</tr>
</tbody>
</table>
Panel Presentation: Assessing Learning and Innovation Skills

At the May 2013 quarterly Governing Board meeting, members engaged in a “blue sky” brainstorming session to explore topics the Board and NAEP might pursue. Among the ideas presented was one that focused on whether NAEP should examine how to measure 21st Century Skills, which are sometimes referred to as learning and innovation skills, work readiness skills, and other titles.

In August 2013, Board members discussed several of the “blue sky” ideas in more depth. To provide additional background information on measuring 21st century skills, it was suggested that a panel of experts present information on the latest research and work in this area.

On Friday December 6, Chairman Driscoll will moderate a panel discussion on the assessment of learning and innovation skills (a.k.a. 21st Century Skills). The panel members are listed below. Biographical information and background materials are included on the following pages.

- Steven Paine, Partnership for 21st Century Skills
- Martin West, Harvard Graduate School of Education
- Deirdre Knapp, Human Resources Research Organization (HumRRO)
- James Pellegrino, University of Illinois at Chicago
Dr. Steven Paine is President of the Partnership for 21st Century Skills. A consummate life-long educator, Dr. Paine has held numerous positions of leadership in the private and public sectors. Prior to joining P21, he served as Senior Advisor to the McGraw-Hill Education Research Foundation and as Senior level Vice President for CTB/McGraw-Hill, the assessment company within McGraw-Hill Education.

From July, 2005 to January, 2011, he served as West Virginia’s 25th state superintendent of schools. Under his leadership, West Virginia was internationally and nationally recognized for its 21st century learning program entitled Global21: Students deserve it. The world demands it. Led by Paine, West Virginia transformed the rigor and relevance of its public school instructional program with the goal of providing all West Virginia children the skills that would enable them to excel in a fiercely competitive global world. West Virginia's Global21 program specifically focused on the development of internationally rigorous and relevant curriculum standards; a balanced assessment strategy; research and performance based instructional practices; an accountability system based on multiple measures of student performance; aligned teacher preparation and professional development programs; establishment of a 21st century leadership development continuum; emphasis on high quality pre-K programs; and integration of technology tools and skills in every classroom.

While state superintendent, Dr. Paine was active in national education policy discussions as past president and board member of the Council of Chief State School Officers, as a member of the National Commission on Teaching and America's Future (NCTAF) Board of Directors and as a member of the National Assessment Governing Board and High School Readiness Commission.

He joined the West Virginia Department of Education in 2003 as the Deputy State Superintendent of Schools after serving as Superintendent of Morgan County Schools in West Virginia. He has also served as principal, assistant principal, teacher, and curriculum director in Upshur and Harrison County School Systems. As a result of his work as principal, he was named a recipient of the prestigious Milken Family Foundation National Educator Award. Dr. Paine is concurrently serving as the Chief Academic Officer for Engrade, an education technology company based in Santa Monica, California.

Dr. Paine earned his undergraduate degree from Fairmont State University, in Fairmont, West Virginia. He furthered his education by attending West Virginia University in Morgantown, West Virginia, where he received his master’s degree in educational administration and his doctorate in educational leadership and curriculum and instruction.
Key P21 Resources

- P21 Common Core Toolkit
  www.P21.org/P21Toolkit
- Assessment of 21st Century Skills
- P21 Mile Guide: Milestones for Improving Learning & Education

Executive Summaries of P21 Surveys

- AMA 2012 Critical Skills Survey
- Key Findings: Are They Really Ready To Work? 2006 Survey
- Voter Attitudes on 21st Century Skills
Martin West is Associate Professor of Education at the Harvard Graduate School of Education, Deputy Director of the Harvard Kennedy School’s Program on Education Policy and Governance, and Executive Editor of Education Next, a journal of opinion and research on education policy.

His research examines the effects of education policy choices on student achievement and non-cognitive skills, as well as the politics of American education. His current projects include a federally-funded randomized trial of the use of interim assessment data to improve instruction and studies of the causal effect of grade retention on educational attainment, charter school impacts on cognitive and non-cognitive skills, and the views of teachers and the general public on education policy.

West is currently on leave to work as Senior Education Policy Advisor to the ranking member of the Senate Committee on Health, Education, Labor, and Pensions. He has also taught at Brown University and served as a research fellow in Governance Studies at the Brookings Institution, where he is now a Non-resident Senior Fellow. A 1998 graduate of Williams College, he received his M.Phil. in Economic and Social History from Oxford University in 2000 and his Ph.D. in Government and Social Policy from Harvard in 2006.
Promise and Paradox:
Measuring Non-Cognitive Traits of Students and the Impact of Schooling

Martin R. West, Harvard Graduate School of Education
Matthew A. Kraft, Brown University
Chris Gabrieli, Mass2020
Angela L. Duckworth, University of Pennsylvania

November 2013
Draft – Please do not cite or circulate without permission.

Abstract
We used surveys to gather information on a broad set of non-cognitive traits from 1,368 8th-grade students attending Boston public schools and linked this information to administrative data on their demographics and test scores. Scales measuring students’ Conscientiousness, Self-control and Grit are positively correlated with test-score growth between 4th- and 8th-grade. Yet students who attend over-subscribed charter schools with higher test-score growth score lower, on average, on these scales than students attending district schools. Exploiting admissions lotteries, we replicate previous findings indicating positive impacts of charter school attendance on math achievement but find negative impacts on these non-cognitive traits. We provide suggestive evidence that this paradoxical result is an artifact of reference bias, or the tendency for survey responses to be influenced by social context. Our results therefore highlight the importance of improved measurement of non-cognitive traits in order to capitalize on their promise as a tool for informing education practice and policy.
Introduction

Recent evidence from economics and psychology highlights the importance of traits other than general intelligence for success in school and in life (Almlund et al. 2011; Borghans et al. 2008; Moffitt et al. 2011). Disparities in so-called non-cognitive skills appear to contribute to the academic achievement gap separating wealthy from disadvantaged students (Evans and Rosenbaum, 2008). Further, non-cognitive skills may be more malleable than cognitive ability, particularly beyond infancy and early childhood (Borghans et al. 2008; Cunha and Heckman 2009). Understandably, popular interest in measuring and developing students’ non-cognitive skills has escalated (see, e.g., Tough 2012).

Non-cognitive is, of course, a misnomer. Every psychological process is cognitive in the sense of relying on the processing of information of some kind. Characteristic patterns of attending to and interpreting information underlie many if not most personality traits (Bandura 1999; Mischel and Shoda 1999). Moreover, emotion and personality certainly influence the quality of one’s thinking (Baron 1982) and how much a child learns in school (Duckworth and Seligman 2005).

Why, then, does the term non-cognitive persist? Cognitive in this context is shorthand for cognitive ability and knowledge, constructs that can be validly measured by standardized intelligence and achievement tests. Non-cognitive, therefore, has become a catchall term for skills and traits not captured by assessments of cognitive ability and knowledge. Many educators prefer the umbrella term “social and emotional learning,” whereas some psychologists and philosophers embrace the moral connotations of “character” and “virtue.”

Educators are increasingly interested in developing students’ non-cognitive skills in support of academic success and long-term life outcomes. For example, several high-performing
charter management organizations have implemented comprehensive discipline systems aimed at molding student behavior in school and beyond (e.g. homework completion) in pro-social and pro-academic directions (Lake et al. 2012). KIPP Academies goes so far as to issue a regular “Character Report Card” for each student that tracks the development of various non-cognitive skills. Related efforts include the movement to address social and emotional learning needs of students alongside traditional academic goals (Durlak et al. 2011). One indication of this movement’s growing policy influence is the U.S. Department of Education’s August 2013 approval of waiver of federal accountability requirements requested by a consortium of eight school districts known as the California Office to Reform Education (CORE). The CORE districts, which collectively serve more than one million students, proposed a new school accountability metric that weights achievement test outcomes as only 60 percent of overall performance, with the balance assigned evenly to measures of school climate and student social-emotional development.

As practice and policy race forward, however, research on the measurement of non-cognitive traits remains in its infancy. There is little agreement on which non-cognitive traits are most important and limited evidence on their relative malleability. There are neither widely accepted standards for the application of extant measures nor evidence on their susceptibility to gaming if used for high-stakes purposes. Absent consensus on these points, educators cannot rely on available measures of non-cognitive traits or their underlying theories of personal development to assess and support individual students or to evaluate the success of schools, teachers, or interventions. As if to illustrate this dilemma, the CORE waiver request noted only that the specific social-emotional measures to be incorporated into school evaluations would be determined at a later time.
In this paper, we draw on cross-sectional data from an unusually large sample of students in the city of Boston to examine the strengths and limitations of extant survey-based measures of four prominent non-cognitive traits as tools for practice and policy. We used survey instruments to gather self-reported information on non-cognitive traits from a sample of more than 1,300 8th-grade students across a wide range of the city’s public schools and linked this information to administrative data on the students’ demographics and test score performance. The schools attended by students in our sample included both open-enrollment public schools operated by the local school district and over-subscribed charter schools that have been shown to have large positive impacts on student achievement as measured by state math and English language arts tests (Abdulkadiroglu et al. 2011; Angrist et al. 2013).

The non-cognitive traits we measured and focus on in this paper include Conscientiousness, Self-Control, Grit, and Implicit Theory of Intelligence (ITI). Of the many non-cognitive traits that psychologists have studied, Conscientiousness and Self-Control have arguably the strongest evidence of predictive power over long-term outcomes even when controlling for cognitive skills and demographics (Almlund et al. 2011). We also consider two newer measures, Grit and ITI, because of their current salience among educators seeking to influence non-cognitive traits to support immediate academic success and long-term life outcomes. Grit refers to the tendency to sustain interest in and effort toward very long-term goals (Duckworth et al. 2007), while ITI is a measure of students’ academic mindset – in particular the extent to which they believe that their academic ability can improve with effort, rather than being fixed by factors outside of their control (Blackwell et al. 2007).

Our results highlight both the potential value of these measures in explaining the proximate outcome of academic success and a less discussed paradox that may be inherent to
many available measures of non-cognitive traits. The promise is illustrated by the fact that the non-cognitive traits we measure through student self-reports are generally correlated with both the level at which students perform on standardized tests and the growth in their test scores over the previous four years. The paradox is illustrated by the fact that differences in the mean levels of three of the four non-cognitive traits between district and charter schools are in the opposite direction of what would be expected based on these student-level correlations. Students who attend over-subscribed charter schools score lower, on average, on measures of Conscientiousness, Self-Control, and Grit than students attending open-enrollment district schools. Exploiting data from the admissions lotteries for these schools, we replicate previous findings indicating positive impacts of charter school attendance on math achievement within the students in our sample but find large and statistically significant negative impacts on these non-cognitive traits. This pattern is especially puzzling in light of the emphasis the over-subscribed charter schools in our study place on behavior management and character development as a means to foster academic success (Angrist et al. 2013; Seider 2012).1

Two competing hypotheses could explain this paradox. One is that the measures of non-cognitive traits are accurate and that the charter schools, contrary to their goals, and despite their success in raising test scores, reduce students’ non-cognitive abilities along crucial dimensions such as Conscientiousness and Self-Control. An alternative hypothesis is that the measures, all self-reported by students, are misleading because they are prone to reference bias – the tendency for survey responses to be influenced by the context in which the survey is administered. We find suggestive evidence supporting this alternative hypothesis, highlighting the importance of

1 In contrast with these outcomes, we find that students in over-subscribed charter schools score higher on ITI. Our lottery-based analysis, however, shows no effect of charter school attendance on ITI.
improved measurement of non-cognitive traits in order to capitalize on their promise as a tool for informing education practice and policy.

**Measurement of Non-Cognitive Traits and the Perils of Reference Bias**

Recognition of the importance of non-cognitive traits has, with few exceptions, preceded the development of valid and reliable measures thereof. Whereas performance tasks to assess how well children can read, write, and cipher are widely available, non-cognitive skills are typically assessed using self-report and, less frequently, informant-report questionnaires. Like standardized achievement tests, questionnaires have the advantage of quick, cheap, and easy administration.

Like all measures, questionnaires have limitations. Most obviously, questionnaires are subject to social desirability bias (to seem more attractive to observers or to oneself) and faking. When endorsing a survey item such as “I am a hard worker” a child (or her teacher or parent) might be inclined to choose higher ratings. To the extent that social desirability bias is uniform within a population under study, it can alter the absolute level of individual responses but not their rank order. If some individuals are more influenced by social pressure than others, however, their relative placement within the overall distribution of responses can change.

Less obvious but possibly more pernicious is reference bias, or the extent to which responses are influenced by implicitly held standards of comparison. When considering whether “I am a hard worker” should be marked “very much like me,” a child must conjure up mental image of “a hard worker” to which she can then compare her own habits. A child with very high standards might consider a hard worker to be someone who does all of her homework, well before bedtime and, in addition, organizes and reviews all of her notes from the day’s classes.
Another child may consider a hard worker to be someone who attempts to bring home her assignments, even if most of them remain unfinished by the next day.

Reference bias was first documented in cross-cultural psychology, and, indeed, culturally shared standards are a primary influence on implicit standards of comparison. In studies of distinct ethnic groups, cultural experts have often drawn conclusions which were not borne out by self-report questionnaires. For instance, self-report questionnaires of values (e.g., living an exciting life, honoring parents and elders, respect for tradition) by Chinese and Americans hardly converge with judgments by cultural experts (Peng, Nisbett, & Wong, 1997). Likewise, among 56 nations in a cross-cultural study of self-reported personality, East Asian countries including Japan, China, and Korea rated far lower in conscientiousness than any other region (Schmitt et al., 2007). In a separate cross-cultural study by Heine et al. (2008), self-reported conscientiousness at the country level (i.e., the average self-reported conscientiousness rating for citizens of a particular country) was inversely correlated with several objective proxies for conscientiousness, including postal workers’ speed, accuracy of clocks, walking speed, and longevity.

The cultural heterogeneity of the United States allows similar reference bias effects to operate within the same country. In fact, some evidence for reference group bias among American students already exists. Naumann and John (2013) found that European-American undergraduates at UC Berkeley rated themselves higher in conscientiousness than did their Asian-American classmates, despite earning lower GPAs. The paradoxical finding disappeared when both groups are asked to complete the same questions with an explicit referent group of “typical Asian-American Berkeley student.”
Other than ethnicity, what are likely influences on implicit frames of reference? Because children (like adults) are far from omniscient, unable to see the full distribution of human functioning, their peer groups and other aspects of their social context are likely to shape how they evaluate themselves when considering statements such as “I am a hard worker.” It follows that the environment of the school in which they spend much of their waking lives could exert a powerful influence on students’ perspectives on their own attributes. At a low-performing school, where the implicit standard for “hard worker” may be considerably more lax than at a high-performing school, reference bias might even be so severe as to reverse the expected pattern of student responses.

**Data and Measures**

*Sample*

To shed light on the extent to which survey-based measures of key non-cognitive traits are subject to reference bias in a policy-relevant setting, we collected data from a large sample of 8th-grade students attending 32 of the 49 public schools in the city of Boston with an 8th-grade cohort in the spring semester of the 2010-11 school year. The schools that agreed to participate in the study included 22 open-enrollment district schools, 5 over-subscribed charter schools, 2 test-in exam schools, and 3 charter schools which were not over-subscribed at the time the students in our study entered middle school. Within those schools, we sampled all students for whom we were able to obtain parental consent to participate in the study and who were in attendance on the day we collected our data.

We obtained school enrollment and demographic information, data on attendance and suspensions, and math and ELA test scores on the Massachusetts Comprehensive Assessment
System (MCAS) for the students in our sample from databases maintained by the Massachusetts Department of Elementary and Secondary Education. MCAS scores were standardized to have mean zero and unit variance by grade, subject, and year across all tested students in Massachusetts. We limit our analysis to the 1,368 of a total of 1,852 students who participated in the non-cognitive data collection for whom a MCAS math and ELA scores were available in 2007 (when most students were in 4th grade) and 2011, making it possible to track their academic progress and school enrollment since they entered middle school.

Table 1 compares the demographic characteristics and academic indicators of students in our analytic sample to those of all 8th-grade students attending public schools in Boston, as well as to those of 8th-graders attending schools participating in the study. The demographic characteristics of sampled students are quite similar to those of all 8th-grade students attending public schools in the city and to those of 8th-graders attending the same schools. However, the 8th-grade test scores of sampled students are 0.26 standard deviations and 0.19 standard deviations higher in math and ELA, respectively. Comparing the test scores of the sampled students to those of 8th-graders attending the same schools reveals that more than half of this difference reflects positive selection into the study sample within participating schools.

Much of our analysis focuses on comparisons between sampled students attending open-enrollment district and over-subscribed charter schools. Looking separately at these two groups of schools, we see that this positive selection with respect to academic indicators is somewhat more pronounced within the district schools. Specifically, the 8th-grade test scores of sampled students in district schools exceeded those of all students by 0.15 (math) and 0.11 (ELA) standard deviations, while the analogous differences in the over-subscribed charter schools were 0.05 (math) and 0.01 (ELA). This difference does not appear to stem from substantially higher
rates of participation in the study within the charter schools: the share of all 8th-graders participating in the study was 63 percent in the over-subscribed charter schools, as compared with 61 percent in the open-enrollment district schools.

Table 1 also reveals that the 8th-grade test scores of sampled students are considerably higher in the over-subscribed charter schools. The test scores of students in these schools exceeded the statewide mean by 0.27 (math) and 0.37 (ELA) standard deviations, while students in open-enrollment district schools trailed the statewide average by -0.69 (math) and -0.33 (ELA) standard deviations. The students in our sample attending over-subscribed charter schools also experienced larger gains in test scores (relative to the statewide average) between 4th and 8th grade, especially in math. Charter students gained 0.72 standard deviations relative to the state average over those four years, while their district peers lost 0.07 standard deviations. The difference in ELA gains was less pronounced, with charter and district students making positive gains of 0.92 and 0.72 standard deviations, respectively. Sampled students in over-subscribed charter schools were also 10 percentage points more likely to be white, 16 percentage points less likely to be Hispanic, and 21 percentage points less likely to be eligible for a free or reduced price lunch than their counterparts in open-enrollment district schools.

Non-cognitive Measures

All students participating in our study completed a battery of surveys designed to measure their cognitive and non-cognitive abilities along various dimensions. These surveys, which were administered in the students’ regular classrooms, included questionnaires probing students’ Conscientiousness, Self-Control, Grit, and ITI that have been validated for adolescents. After scoring student responses to these questionnaires based on the appropriate rubrics, we then standardized the scores to have a zero mean and unit variance within our analytic sample.
To assess students’ Conscientiousness, we administered the Big Five Inventory (John and Srivastava 1999), a well-established 44-item questionnaire measuring the “Big Five” personality traits: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Students endorsed items (e.g., “I think I am someone who is a reliable worker”) using a 5-category Likert scale, where 1 = *strongly disagree* and 5 = *strongly agree*. Each student’s Conscientiousness score is calculated as the average of their response the 9 items that comprise the Conscientiousness scale. Among the students in our data, this Conscientious scale had an internal reliability of 0.76.

Our Self-Control scale is based on an 8-item questionnaire developed to measure school-age students’ impulsivity (its obverse), which is defined as “inability to regulate behavior, attention, and emotions in the service of valued goals, impulsivity” (Tsukayama et al. 2013). This questionnaire asked students to indicate how often during the past school year they exhibited each of a set of behaviors indicative of a lack of self-control, with 5 response options ranging from “almost never” to “at least once a day.” Importantly, the use of response categories specifying objective, discrete time periods was motivated by a desire to “avoid reference bias” in students’ responses (Tsukayama et al. 2013, p. 881). The questionnaire included 4 items measuring inter-personal self-control (e.g., “I interrupted other students while they were talking”) and 4 items measuring intra-personal self-control (e.g., “I forgot something I needed for class”). We calculated an overall Self-Control score for each student as the average of their (reverse-coded) responses to all 8 items. This scale had an internal reliability of 0.83.

Students also completed the 8-item Short Grit Scale (Grit-S) developed by Duckworth and Quinn (2009) to measure trait-level persistence toward long-term goals.2 Students endorsed

---

2 Duckworth and Quinn (2009) demonstrate that adolescents’ Grit-S scores predict future GPA independently of IQ and are inversely related to the number of hours of television watched.
a series of items (e.g., “New ideas and projects sometimes distract me from old ones” and “I finish whatever I begin”) using a 5-category Likert Scale, where 1 = not like me at all and 5 = very much like me. Student’s Grit score was then calculated as their mean response across all 8 items. This scale had a somewhat lower internal reliability among the students in our sample than our Conscientiousness and Self-Control scales (0.64).

Finally, to probe students’ Implicit Theory of Intelligence, we administered a set of 3 items measuring the extent to which students view intelligence as a fixed trait rather than one which can improved with effort (Dweck, 1999). For example, students were asked to rate their agreement with the claim that “You have a certain amount of intelligence, and you really can’t do much to change it.” Following Blackwell et al. (2007), we used a 6-category Likert scale, where 1 = strongly disagree and 6 = strongly agree. After reverse coding, we calculated each student’s mean response across these three items to create a scale with an internal reliability of 0.86. Following recent work (see, e.g., Yeager et al. 2013), we refer to this scale as measuring the extent to which students have a Growth Mindset (as opposed to a Fixed Mindset).³

Student Perceptions of School Climate

The battery of surveys students completed also included a set of 10 items probing the disciplinary climate at their school. These items were drawn from a survey developed over the past decade by Ron Ferguson and the Tripod Project for School Improvement to measure various aspects of school or classroom climate. Each item asked students to respond to a descriptive statement about their school using a 5-category Likert scale, where 1 = totally untrue and 5 = totally true. The overall set included two items on each of five dimensions of school climate: High Expectations (e.g., “Teachers at this school demand that students work hard”); Teacher

³ Blackwell et al. (2007) show that adolescent students who exhibit a Growth Mindset have significantly higher rates of math test score growth than students who view intelligence as fixed.
Strictness (e.g., “Teachers are very strict here”); Clear Rules (e.g., “Students understand what will happen to them if they break a rule”); Negative Peer Effects (e.g., “In this school, some students try to keep others from working hard”); and Student Input (e.g., “Students in this school have a say in how things work”). We use the average of each student’s responses to the two items within each pair as a measure of his or her perception of the relevant aspect of the school’s climate.

Achievement Gains

We also used our administrative data to estimate measures of each student’s test score growth in math and English language arts between 4th- and 8th-grade. Specifically, we regressed students’ 8th-grade MCAS test scores in the relevant subject on a cubic polynomial of their 4th-grade test scores in both subjects and calculated the residual from that regression for each student. We use these residualized gain scores, which capture the extent to which a student’s 8th-grade performance in math and English language arts exceeded expectations based on their performance four years earlier, to examine the relationship between non-cognitive traits and improvements in test score performance over time.

Results

Correlations of Non-cognitive Traits and Academic Indicators

Table 2 reports student-level Pearson product-moment correlations among the full set of non-cognitive traits and academic indicators included in our analysis. Given that Conscientiousness, Self-Control, and Grit are closely related constructs, it is unsurprising that they are highly inter-related, with correlations ranging from .43 to .66. Growth Mindset is also positively and significantly correlated with each of these measures, but at lower levels ranging from .08 (Conscientiousness) to .18 (Grit).
Among these four non-cognitive measures, Growth Mindset is most strongly related to test score levels in 8th grade ($r = 0.32$ in math; $r = .36$ in ELA). Self-Control is also significantly related to test scores, but the correlations are $.13$ in math and $.10$ in ELA. The correlations between both Conscientiousness and Grit and test score levels are positive but small and statistically insignificant.

Of greater interest are the relationships between the non-cognitive measures and residualized test score gains, which measure students’ academic progress relative to expectations based on their performance in 4th grade. Each of the four non-cognitive measures is positively correlated with test score gains in both math and ELA; all of these correlations except that between Self-Control and ELA gains are statistically significant. The relationships are strongest for Growth Mindset, which has correlations with test score gains of $.21$ and $.17$ in math and ELA, respectively.

There is also some evidence that these non-cognitive measures are related to suspensions and absences, the two behavioral indicators available in our administrative data. All four non-cognitive measures are negatively correlated with the total number of suspensions or absences a student accumulated in 8th grade. Those correlations that are statistically significant include Self-Control with both suspensions ($-.14$) and absences ($-.12$), Grit with suspensions ($-.12$), and Growth Mindset with absences ($-.10$).

Mean Non-cognitive Traits and Academic Indicators by School Type

Table 3 compares the mean test-score gains and non-cognitive traits for students attending the 22 open-enrollment district and 5 over-subscribed charter schools included in our sample. Consistent with the descriptive statistics in Table 1, mean residualized test score gains between 4th and 8th grade are higher among students attending charter schools. The differences
are substantial, at 0.72 standard deviations in mathematics and 0.42 standard deviations in ELA, though it is important to note that they could reflect the selection of students into the application process for over-subscribed charter schools rather than differences in school quality.

Despite the fact that sampled students attending charter schools experienced larger test score gains than sampled students in district schools, the same students exhibit markedly lower levels of Self-Control as measured by student self-reports. This statistically significant difference of -0.23 standard deviations is in the opposite direction of that expected based on the positive student-level correlations between Self-Control and achievement gains evident in Table 2. The differences between the charter and district students in Conscientiousness (-0.09) and Grit (-0.13), although statistically insignificant, run in the same counter-intuitive direction.

Interestingly, the difference in mean Growth Mindset between charter school and district school students follows a different pattern than the other three non-cognitive traits. Charter school students score 0.38 standard deviations higher, on average, which is consistent with the student-level correlation between Growth Mindset and test score gains within the sample as a whole.

**Lottery-Based Estimates of the Effect of Charter School Attendance**

Simple comparisons of the outcomes of students attending charter and traditional public schools, such as those presented in Table 3, capture both any effects of attending a charter school on those outcomes and selection into charter schools based on characteristics correlated with the outcome. Although over-subscribed charter schools must admit students via lottery, applicants who seek to enroll in an academically demanding charter school are likely to differ from those who do not along both observed and unobserved dimensions. To better isolate the causal effect of attendance at the five over-subscribed charter schools represented in our sample, we exploit the lottery admissions process to these schools to restrict our analysis to students who entered the
admissions lottery of one or more of the over-subscribed charter schools and compare those students who were randomly offered admission to those who were not.

We acquired records from the lotteries used to admit the students in our sample directly from the charter schools and matched these records to state administrative data on all public school students using names, year, and grade of application. Of 702 verified lottery participants (481 of whom were offered a seat in one or more of the five schools), 497 appeared in the administrative data and had valid demographic data and test scores for both subjects in 2007 and 2011. A smaller subset of 200 students met those requirements and participated in the non-cognitive data collection. Although we can produce lottery-based estimates of charter school impacts on non-cognitive traits only in the latter sample, we present estimated impacts on test scores for both groups in order to be able to compare the results. To the extent that our estimated impacts on test scores are similar across the two groups, this should reduce concerns that our results are biased due to non-random sampling of successful and unsuccessful applicants.

Table 4 examines whether the demographic characteristics and 4th-grade test scores of students offered and not offered a seat in a charter school were balanced within these two subsamples of lottery participants. We first note that within both subsamples the share of applicants who were offered a seat (32 percent in the administrative data sample and 29 percent in the non-cognitive sample) is very similar to the share among all lottery applicants (31 percent). F-tests nonetheless reveal that students’ observed characteristics are jointly significant predictors of whether they were offered a seat in a charter school in both subsamples.

In the larger sample of students matched to the administrative data, students receiving an offer are significantly more likely to be male and eligible for a free or reduced price lunch. The 4th-grade test scores of students offered a seat are also modestly lower in both math and ELA,
though these differences are not statistically significant. Within the smaller non-cognitive sample, the differences in 4th-grade test scores are substantial in both subjects and marginally statistically significant in math (p<0.09). These differences, which favor students not offered a charter school seat, may reflect the fact that positive selection into the non-cognitive sample was less pronounced in over-subscribed charter schools (see Table 1). Given the imbalances between students offered and not offered a seat in a charter in both samples, we control for students’ observed characteristics throughout our lottery-based analysis of the effects of charter school attendance.

Following (Abdulkadiroglu et al. 2011), we implement the lottery-based analysis via a Two Stage Least Squares (2SLS) regression model in which we first predict charter attendance for each student based on whether they were offered admission and use those predictions to generate an estimate of the effect of charter attendance on our outcomes of interest. The first-stage model is:

\[
YEARS_i = \gamma_{OFFER_{ic}} + \theta A_{i,t-4} + \tau X_i + \sum_j \rho_{ij} d_{ij} + \xi_{is}
\]

where \(YEARS_i\) measures the number of years between 5th and 8th grade student \(i\) attended an over-subscribed charter school and \(OFFER_{ic}\) represents a vector of dummy variables indicating that the student was or was not offered a seat at over-subscribed charter school \(c\). We include as controls a cubic polynomial of lagged 4th-grade scores in math and ELA \((A_{i,t-4})\) and a vector of student demographic characteristics \((X_i,)\) including gender, race, age, free and reduced-priced lunch status, limited English proficiency, and special education status. The set of indicator variables \(d_{ij}\) controls for lottery “risk sets,” or the unique combination of lotteries to which each student applied, indexed by \(j\).
Within the administrative data sample, students offered a seat in at least one of the five over-subscribed charter schools spent 2.1 years between 5th and 8th grade in one of those schools, as compared with 0.39 years of charter attendance among students not initially offered a seat. Among students in the non-cognitive sample, students offered and not offered a seat spent 2.4 and 0.6 years in charter schools, respectively. Appendix Table A2, which presents the first-stage regression results, confirms that the $OFFER_{ic}$ indicators are relevant instruments for predicting variation in years of attendance, with joint F-test statistics of 35.5 and 16.0 for the administrative data and non-cognitive samples, respectively.

We implement the second stage of our 2SLS model as follows:

$$Y_{is} = \beta \text{YEARS}_{is} + \alpha A_{i,t-4} + \lambda X_i + \sum_j \delta_j d_{ij} + \epsilon_{is}$$

where $Y_{is}$ represents a given test score or non-cognitive outcome for student $i$ in school $s$, $\text{YEARS}_{is}$ captures each student’s predicted years of charter attendance based on the first-stage regression, and all other variables are as above. Parameter $\beta$ represents the quantity of interest: the effect of one year’s attendance at one of the five over-subscribed charter schools.

Table 5 presents the results. Consistent with Abdulkadiroglu et al. (2011), the first two columns show that, among the students in the administrative data sample, each additional year of charter attendance is estimated to increase 8th-grade math scores by 0.14 standard deviations. The estimated effect for ELA scores is positive and of non-negligible magnitude, but too imprecisely estimated to achieve statistical significance. We replicate these analyses in our non-cognitive sample and find very similar point estimates in both math and ELA; the math effect is statistically significant at the $p<0.1$ confidence level. The similarity of results across the two samples suggests that our lottery-based effect estimates on non-cognitive skills are unlikely to
suffer from substantial bias due to non-random sampling of students who participated in our non-cognitive data collection.

Within that sample, we estimate that one year’s attendance at an over-subscribed charter school had a statistically significant negative effect on students’ self-reported Conscientiousness, Self-Control, and Grit. The estimated effect sizes are in the opposite direction of the achievement effects and of similar or even larger magnitude, ranging between -0.117 (Grit) and -0.212 (Self-Control) standard deviations. These results are consistent with the descriptive patterns in Table 3, which show students in over-subscribed charter schools making larger achievement gains despite lower scores on these non-cognitive measures, and suggest that those patterns are not due merely to selection. Rather, it would appear that attending one of these charter schools adversely affects students’ non-cognitive abilities along these dimensions as assessed by self-reports. We discuss our interpretation of this unexpected finding in detail below.

Intriguingly, we estimate a near zero effect of attending an over-subscribed charter schools on the degree to which a student in our non-cognitive sample has a Growth Mindset, despite the fact that Table 3 showed students in over-subscribed charter schools scoring notably higher on this measure. The null result for this outcome in the lottery analysis suggests that the descriptive difference favoring charter schools may be an artifact of selection. In other words, it may be that students who believe that their intelligence can be improved with effort are more likely to seek out a school with a demanding environment, but that attendance at such a school has no causal effect on their implicit theory of intelligence.

*Longitudinal Trends in Non-Cognitive Traits*

We supplement our lottery analysis with longitudinal measures of students’ non-cognitive traits among a cohort of students who attended two over-subscribed charter schools and one
open-enrollment district school. Starting in fall 2009, we administered a similar battery of non-
cognitive measures to the entering student cohorts at three middle schools. We then re-
administered these batteries at the end of that school year and the two that followed. One of the
charter schools and the district school are both included in our larger study sample, while the
second charter school is located in another school district but shares a similar academic and
disciplinary orientation to the over-subscribed charters in our sample.

In Table 6, we present average scores for the 2009 entering cohort of students in each
school for whom we have complete data across four time points. Consistent with our lottery-
based estimates of the negative effects of attending a charter school on non-cognitive skills as
assessed by self-reports, we observe a steady decline in students’ Conscientiousness, Self-
Control, and Grit. Among students attending the district school, scores on these scales decline at
a more moderate pace; in two cases, the changes between the first and final time points are not
statistically significant. Also consistent with our lottery-based estimates, we observe a gradual
increase in Growth Mindset among oversubscribed charter school students which is mirrored by
a similarly large increase among students at the district school.

Evidence of Reference Bias

The results presented thus far contain divergent evidence concerning the relationship
between non-cognitive traits and achievement gains among the 8th-grade students in our sample,
particularly with respect to the related constructs of Conscientiousness, Self-Control, and Grit.
Student-level correlations indicate significant positive correlations between these non-cognitive
traits and test score growth – a pattern consistent with a large body of research using the same or
similar measures. However, we find that students in over-subscribed charter schools with large
positive impacts on student test scores rate themselves more critically along each of these

4 Appendix Table A2 reports the same averages using all available data; the patterns are substantively identical.
dimensions. Our lottery-based analysis indicates that these same charter schools have large negative “effects” on students’ self-reported levels of Conscientiousness, Self-Control, and Grit, suggesting the descriptive findings are not explained by the selection into charter schools of less conscientious or more impulsive students. It is of course possible that the over-subscribed charter schools in our sample improve student test scores at the expense of cultivating students’ non-cognitive capacities in these areas. Yet it is seems more likely that the apparent negative effects of charter school attendance on these non-cognitive traits are an artifact of reference bias.

We present two additional analyses intended to establish the plausibility of reference bias as an explanation for these counter-intuitive findings. First, we compare students’ perceptions of the academic and disciplinary climate in open-enrollment district and over-subscribed charter schools to see whether those perceptions differ in a way that could influence students’ self-ratings of their non-cognitive capacities. It could be that students are more likely to use a higher bar when assessing their own Conscientiousness, Self-Control, and Grit when they attend schools that establish high expectations for student effort and a “no-excuses” disciplinary culture. Second, we examine how the strength of the relationship between our non-cognitive measures and achievement growth changes if we use school fixed effects to limit the analysis to variation occurring within specific schools. If the environment of the school students attended and in which they completed our survey influenced their responses so as to produce the counter-intuitive findings concerning the effects of charter school attendance, the within-school correlations between non-cognitive measures and achievement growth should be stronger than the analogous correlations that capture both between- and within-school variation.

Table 7 confirms that students attending over-subscribed charter schools perceive their schools as having very different academic and disciplinary climates than students attending
open-enrollment district schools. Students in over-subscribed charter schools rate the work ethic expected of students, teacher strictness, and the clarity of rules in their school substantially higher do students in district schools. For example, charter students’ ratings of High Expectations exceed those of their district counterparts by 0.57 on the 5-point scale used for these items, or 63 percent of a standard deviation of district students’ responses. The analogous differences observed for Teacher Strictness and Clear Rules are of comparable magnitude. Students in the over-subscribed charter schools also reported substantially lower levels of Negative Peer Effects and modestly lower levels of Student Input in their schools.

Table 8 in turn confirms that the relationships between Conscientiousness, Self-Control, and Grit and test score growth among students attending the same school are consistently stronger than the same relationships across the sample as a whole. For example, the magnitude of the relationship between Self-Control and test score growth increases by 61 percent in math and 93 percent in ELA. Importantly, the same pattern is not evident for Growth Mindset, the one non-cognitive measure in which we did not find counter-intuitive evidence of adverse effects of attending an over-subscribed charter school. The relationship between Growth Mindset and test score growth is essentially unchanged in ELA and becomes weaker in math when between-school variation is excluded.

Discussion

Generations of parents have sought to instill in their children the virtues of self-discipline, diligence, and perseverance, and self-discipline. These qualities are at the core of legends and fables that societies around the world have developed to cultivate the traits most essential for human flourishing. In recent decades, scholars have begun to confirm this common wisdom by
developing measures for these non-cognitive traits and examining their relationship with children’s success in school and in life. We add to this literature evidence that four prominent and widely used measures of non-cognitive traits are positively correlated with achievement gains on standardized tests among a large and diverse sample of 8th-grade students attending distinctly different types of schools. Measures of Conscientiousness, Self-Control, Grit, and Growth Mindset were all correlated with math and ELA test score gains from 4th to 8th grade. Higher non-cognitive ability along the dimensions captured by these measures therefore may help explain why 8th-grade students score higher or lower than predicted by their 4th-grade achievement levels.

However, a paradox emerges when we juxtapose these results with two additional findings: 1) that students in a set of over-subscribed charter schools, where students are experience large test score gains, report lower average levels of Conscientiousness, Self-Control, Grit than students in open-enrollment district schools; and 2) that lottery-based analyses of the causal impact of attending these charter schools indicate negative “effects” on these non-cognitive traits. How can non-cognitive skills that are positively correlated with test score gains within our sample have deteriorated among students in schools where they were simultaneously achieving large test score improvements?

Two competing hypotheses could explain the paradox. One is that these measures are accurate and that the charter schools’ actions actually reduce students’ non-cognitive abilities along crucial dimensions such as conscientiousness and self-control. In that case, the academic gains posted by these schools occur in spite of their negative effects on these non-cognitive measures.
An alternative hypothesis is that these measures, all based on student self-reports, are misleading because they are prone to reference bias—the tendency for individuals responding to questionnaires to rate themselves based on the varying localized sample of people with whom they are familiar, not the broad sample of all respondents to whom their responses are being compared. Put specifically for these circumstances, students attending academically intense charter schools may redefine upward their notion of what it means to demonstrate strong self-control or a conscientious work ethic and thus rate themselves more critically, even if they are in fact improving on the underlying behavior. In theory, such reference bias could be so severe as to distort the magnitude of any changes in the underlying traits and even to invert their sign.

We find support for this alternative hypothesis not only in our data set, but also in other recent evaluations of high-performing charter middle schools. In their recent evaluation of KIPP Academy middle schools, Tuttle et al. (2013) find large positive effects of attending a KIPP school on student test scores and time spent on homework, but find no effects on student-reported measures of self-control and effort or persistence in school and negative effects on student-reported measures of undesirable behavior. Similarly, Dobbie and Fryer’s (2013) find that attending the Harlem Promise Academy reduced student-reported Grit despite having positive effects on test scores college enrollment and negative effects on teenage pregnancy (for females) and incarceration (for males). These parallel findings from research in similar settings strengthen our conclusion that reference bias provides the most likely explanation for the unexpected patterns we document.

Reference bias also may help explain seemingly paradoxical results in other areas of education research. For example, Robbins et al. (2006), report that first-year students at two-year colleges rate themselves higher in an ACT-developed measure of academic self-discipline that is predictive of college GPA and persistence than do seniors bound for four-year colleges, despite having lower ACT scores, high school GPAs, and subsequent rates of persistence.
Conclusion

Our results suggest three things: (1) that existing measures of non-cognitive traits based on self-reports help to explain important, proximate academic outcomes – test score gains over the middle school years – beyond what previous test scores predict and therefore show promise as traits for schools to learn to measure and influence; (2) that schools can have significant, even if initially surprising, impacts on these measures; and (3) that conclusions about the nature of those impacts may be misleading due to reference bias. The challenges posed by reference bias may grow more severe to the extent that schools explicitly work to change students’ traits and thereby change their reference frames.

Our study has important limitations. First, our evidence of the importance of reference bias is circumstantial rather than direct. Second, we have documented the potential problem posed by reference bias without providing a solution to overcome it. In particular, we have not examined whether teacher or parent ratings of students’ non-cognitive traits may be less prone to reference bias. Alternative solutions could include the use of anchoring vignettes within surveys to establish consistent reference points (King et al. 2004) or the development of behavioral indicators of non-cognitive traits that render reference points irrelevant (Jackson et al. 2010). Additional research that documents the conditions under which reference bias exists and validates strategies to address it is critical.

The current policy environment demands accountability, and accountability requires assessment. In the rush to embrace non-cognitive skills as the missing piece in American education, policymakers may overlook limitations of extant measures of non-cognitive skills. Our results raise important questions about the practice of assessing students’ non-cognitive traits based on the existing instruments that rely on student self-reported data. In particular,
studies of the effects of teacher, school, and family influences on non-cognitive skills could lead to false conclusions if the assessments used are biased by distinct frames of reference. Biased measures could similarly misguide scientific investigation of non-cognitive skills.

If we are correct that these measures show both promise and peril, it is imperative that the nascent field of measuring and seeking to improve non-cognitive traits through schooling develop new, better measures that are less susceptible to reference bias and therefore more likely to be robust enough to play a constructive role in managing and evaluating students, programs, and schools. We can and should measure students’ non-cognitive traits, but we should do so in full recognition of the flaws in our measures.
References


### Table 1: Mean student demographic characteristics and academic indicators by school type among all public schools in Boston, sampled schools, and sampled students

<table>
<thead>
<tr>
<th></th>
<th>Public Schools in Boston</th>
<th>Open-enrollment District Schools</th>
<th>Over-subscribed Charter Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Students</td>
<td>All Students in Sampled Schools</td>
<td>All Students in Sampled Schools</td>
</tr>
<tr>
<td></td>
<td>All Students</td>
<td>Sampled Schools</td>
<td>Sampled Students</td>
</tr>
<tr>
<td>Male</td>
<td>0.52</td>
<td>0.51</td>
<td>0.47</td>
</tr>
<tr>
<td>African-American</td>
<td>0.41</td>
<td>0.40</td>
<td>0.38</td>
</tr>
<tr>
<td>White</td>
<td>0.12</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td>Asian</td>
<td>0.09</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.37</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>Free/Reduced Price Lunch</td>
<td>0.79</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>Limited English Proficient</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>Special Education</td>
<td>0.22</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>8th-Grade Math</td>
<td>-0.52</td>
<td>-0.42</td>
<td>-0.26</td>
</tr>
<tr>
<td>8th-Grade ELA</td>
<td>-0.23</td>
<td>-0.15</td>
<td>-0.04</td>
</tr>
<tr>
<td>4th-Grade Math</td>
<td>-0.45</td>
<td>-0.45</td>
<td>-0.35</td>
</tr>
<tr>
<td>4th-Grade ELA</td>
<td>-0.82</td>
<td>-0.80</td>
<td>-0.71</td>
</tr>
<tr>
<td>Suspensions</td>
<td>0.25</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>Absences</td>
<td>11.85</td>
<td>11.03</td>
<td>9.00</td>
</tr>
<tr>
<td>Number of schools</td>
<td>49</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Number of students</td>
<td>3151</td>
<td>2121</td>
<td>1368</td>
</tr>
</tbody>
</table>

Note: All samples are restricted to students with valid 2011 and 2007 MCAS scores. Sampled schools are schools participating in non-cognitive trait data collection; sampled students are those with valid data on at least one non-cognitive trait. Math and ELA test scores are standardized to have a mean zero and unit variance statewide.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscientiousness</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Control</td>
<td>0.47</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grit</td>
<td>0.66</td>
<td>0.43</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth Mindset</td>
<td>0.08</td>
<td>0.10</td>
<td>0.18</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th-Grade MCAS Math</td>
<td>0.05</td>
<td>0.13</td>
<td>0.03</td>
<td>0.32</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8th-Grade MCAS ELA</td>
<td>0.05</td>
<td>0.10</td>
<td>0.04</td>
<td>0.36</td>
<td>0.69</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th-8th Grade Math Gain</td>
<td>0.10</td>
<td>0.08</td>
<td>0.12</td>
<td>0.21</td>
<td>0.63</td>
<td>0.38</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th-8th Grade ELA Gain</td>
<td>0.09</td>
<td>0.04</td>
<td>0.08</td>
<td>0.17</td>
<td>0.26</td>
<td>0.62</td>
<td>0.44</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspensions</td>
<td>-0.06</td>
<td>-0.14</td>
<td>-0.12</td>
<td>-0.04</td>
<td>-0.10</td>
<td>-0.11</td>
<td>-0.01</td>
<td>-0.03</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Absences</td>
<td>-0.06</td>
<td>-0.12</td>
<td>-0.03</td>
<td>-0.10</td>
<td>-0.29</td>
<td>-0.25</td>
<td>-0.17</td>
<td>-0.11</td>
<td>0.13</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: Sample restricted to students with complete data on each indicator (N=1,340); bolded entries are statistically significant at p<0.1.
<table>
<thead>
<tr>
<th></th>
<th>Open-enrollment District School</th>
<th>Over-subscribed Charter School</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MCAS Math Gain (4th-8th)</strong></td>
<td>-0.015</td>
<td>0.708</td>
<td>0.723**</td>
</tr>
<tr>
<td></td>
<td>(0.712)</td>
<td>(0.699)</td>
<td>(0.122)</td>
</tr>
<tr>
<td></td>
<td>[906]</td>
<td>[148]</td>
<td></td>
</tr>
<tr>
<td><strong>MCAS ELA Gain (4th-8th)</strong></td>
<td>-0.017</td>
<td>0.407</td>
<td>0.424**</td>
</tr>
<tr>
<td></td>
<td>(0.684)</td>
<td>(0.697)</td>
<td>(0.120)</td>
</tr>
<tr>
<td></td>
<td>[906]</td>
<td>[148]</td>
<td></td>
</tr>
<tr>
<td><strong>Conscientiousness</strong></td>
<td>0.022</td>
<td>-0.069</td>
<td>-0.091</td>
</tr>
<tr>
<td></td>
<td>(0.994)</td>
<td>(0.981)</td>
<td>(0.078)</td>
</tr>
<tr>
<td></td>
<td>[890]</td>
<td>[145]</td>
<td></td>
</tr>
<tr>
<td><strong>Self-control</strong></td>
<td>0.001</td>
<td>-0.225</td>
<td>-0.226*</td>
</tr>
<tr>
<td></td>
<td>(1.017)</td>
<td>(0.969)</td>
<td>(0.116)</td>
</tr>
<tr>
<td></td>
<td>[891]</td>
<td>[145]</td>
<td></td>
</tr>
<tr>
<td><strong>Grit</strong></td>
<td>0.034</td>
<td>-0.099</td>
<td>-0.133</td>
</tr>
<tr>
<td></td>
<td>(0.986)</td>
<td>(1.009)</td>
<td>(0.093)</td>
</tr>
<tr>
<td></td>
<td>[888]</td>
<td>[145]</td>
<td></td>
</tr>
<tr>
<td><strong>Growth Mindset</strong></td>
<td>-0.121</td>
<td>0.260</td>
<td>0.381**</td>
</tr>
<tr>
<td></td>
<td>(1.016)</td>
<td>(0.95)</td>
<td>(0.104)</td>
</tr>
<tr>
<td></td>
<td>[887]</td>
<td>[144]</td>
<td></td>
</tr>
</tbody>
</table>

Note: +p<0.1, *p<0.05, ** p<0.01; statistical significance is for difference in mean for over-subscribed charter schools and traditional public schools. In the first two columns, standard deviations are reported in parentheses and sample sizes in brackets. Standard errors reported in parentheses in the third column are adjusted for clustering by school. MCAS scores are standardized to have mean zero and unit variance statewide; non-cognitive traits are standardized to have mean zero and unit variance in the study sample.
Table 4: Balance of observed characteristics in admissions lotteries for over-subscribed charter schools

<table>
<thead>
<tr>
<th></th>
<th>Administrative Data Sample</th>
<th>Non-cognitive Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Offer</td>
<td>No Offer</td>
</tr>
<tr>
<td>Male</td>
<td>0.501</td>
<td>0.437</td>
</tr>
<tr>
<td>African-American</td>
<td>0.590</td>
<td>0.513</td>
</tr>
<tr>
<td>White</td>
<td>0.100</td>
<td>0.171</td>
</tr>
<tr>
<td>Asian</td>
<td>0.009</td>
<td>0.025</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.277</td>
<td>0.278</td>
</tr>
<tr>
<td>Free/Reduced Price Lunch</td>
<td>0.687</td>
<td>0.563</td>
</tr>
<tr>
<td>Limited English Proficient</td>
<td>0.032</td>
<td>0.070</td>
</tr>
<tr>
<td>Special Education</td>
<td>0.195</td>
<td>0.209</td>
</tr>
<tr>
<td>4th Grade MCAS Math</td>
<td>-0.498</td>
<td>-0.395</td>
</tr>
<tr>
<td>4th grade MCAS ELA</td>
<td>-0.784</td>
<td>-0.615</td>
</tr>
<tr>
<td>Joint F-test statistic</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>339</td>
<td>158</td>
</tr>
</tbody>
</table>

Note: Administrative data sample includes all applicants to over-subscribed charter schools matched to valid 2011 and 2007 test scores in the Massachusetts Department of Elementary and Secondary Education state database. Study sample includes students in the administrative data sample with valid data on at least one non-cognitive outcome. The first two columns for each sample provide the mean of each variable for students receiving at least one and no offers of admission to an over-subscribed charter school. Regression-adjusted differences control for fixed effects for lottery applicant risk sets used to estimate charter attendance effects. P-values are for the regression-adjusted difference.
Table 5: Instrumental variables estimates of the effects of a year's attendance at an over-subscribed charter school

<table>
<thead>
<tr>
<th></th>
<th>Administrative Data Sample</th>
<th>Non-cognitive Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>8th Grade MCAS Math</td>
<td>0.138**</td>
<td>0.118+</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>8th Grade MCAS ELA</td>
<td>0.039</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>-0.153+</td>
<td>(0.075)</td>
</tr>
<tr>
<td></td>
<td>[196]</td>
<td></td>
</tr>
<tr>
<td>Self-Control</td>
<td>-0.212*</td>
<td>(0.095)</td>
</tr>
<tr>
<td></td>
<td>[196]</td>
<td></td>
</tr>
<tr>
<td>Grit</td>
<td>-0.117*</td>
<td>(0.053)</td>
</tr>
<tr>
<td></td>
<td>[195]</td>
<td></td>
</tr>
<tr>
<td>Growth Mindset</td>
<td>-0.03</td>
<td>(0.092)</td>
</tr>
<tr>
<td></td>
<td>[195]</td>
<td></td>
</tr>
</tbody>
</table>

Notes: +p<0.1, *p<0.05, ** p<0.01. Standard errors reported in parentheses are clustered by 8th-grade school. Sample sizes for each outcome are in brackets. Each cell presents results from a separate regression. Administrative data sample includes all applicants to over-subscribed charter schools matched to valid 2011 test scores in the Massachusetts Department of Elementary and Secondary Education state database. Study sample includes students in the administrative data sample with valid data on at least one non-cognitive outcome. All regressions control for student gender, race, age, LEP, SPED, and free/reduced price lunch, cubic polynomials in 4th-grade MCAS ELA and math scores, and fixed effects for lottery applicant risk sets.
Table 6: Mean student non-cognitive traits in three middle schools over time, stable sample

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Fall Y1</th>
<th>Spring Y1</th>
<th>Spring Y2</th>
<th>Spring Y3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conscientiousness</td>
<td>Self-Control</td>
<td>Grit</td>
<td>Growth Mindset</td>
</tr>
<tr>
<td>Charter School 1</td>
<td>3.83</td>
<td>3.69</td>
<td>3.55</td>
<td>3.40</td>
</tr>
<tr>
<td>Charter School 2</td>
<td>3.70</td>
<td>3.70</td>
<td>3.49</td>
<td>3.34</td>
</tr>
<tr>
<td>District School</td>
<td>3.79</td>
<td>3.67</td>
<td>3.64</td>
<td>3.68</td>
</tr>
<tr>
<td>Charter School 1</td>
<td>3.48</td>
<td>3.22</td>
<td>3.08</td>
<td>2.91</td>
</tr>
<tr>
<td>Charter School 2</td>
<td>3.42</td>
<td>3.31</td>
<td>3.12</td>
<td>2.96</td>
</tr>
<tr>
<td>District School</td>
<td>3.46</td>
<td>3.31</td>
<td>3.22</td>
<td>3.26</td>
</tr>
<tr>
<td>Charter School 1</td>
<td>3.62</td>
<td>3.41</td>
<td>3.31</td>
<td>3.23</td>
</tr>
<tr>
<td>Charter School 2</td>
<td>3.58</td>
<td>3.44</td>
<td>3.20</td>
<td>3.24</td>
</tr>
<tr>
<td>District School</td>
<td>3.42</td>
<td>3.42</td>
<td>3.32</td>
<td>3.36</td>
</tr>
<tr>
<td>Charter School 1</td>
<td>3.74</td>
<td>4.25</td>
<td>4.50</td>
<td>4.40</td>
</tr>
<tr>
<td>Charter School 2</td>
<td>3.71</td>
<td>4.67</td>
<td>4.42</td>
<td>4.42</td>
</tr>
<tr>
<td>District School</td>
<td>3.74</td>
<td>4.30</td>
<td>4.26</td>
<td>4.44</td>
</tr>
</tbody>
</table>

Note: Sample restricted to students with valid data in each year. N=61 for Charter School 1; N=43 for Charter School 2; N=65 for District School. See Appendix Table A2 for data on all sampled students. Conscientiousness, Self-Control, and Grit are measured on a 5-point scale; Growth Mindset is measured on a 6-point scale. Years 1-3 correspond to grades 5-7 in Charter School 1 and grades 6-8 in Charter School 2 and District School.
Table 7: Student perceptions of school climate by school type

<table>
<thead>
<tr>
<th></th>
<th>Open-enrollment District School</th>
<th>Over-subscribed Charter School</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Expectations</td>
<td>3.929</td>
<td>4.496</td>
<td>0.567**</td>
</tr>
<tr>
<td></td>
<td>(0.900)</td>
<td>(0.669)</td>
<td>(0.149)</td>
</tr>
<tr>
<td></td>
<td>[885]</td>
<td>[112]</td>
<td></td>
</tr>
<tr>
<td>Teacher Strictness</td>
<td>3.526</td>
<td>4.107</td>
<td>0.581*</td>
</tr>
<tr>
<td></td>
<td>(0.888)</td>
<td>(0.904)</td>
<td>(0.211)</td>
</tr>
<tr>
<td></td>
<td>[878]</td>
<td>[112]</td>
<td></td>
</tr>
<tr>
<td>Clear Rules</td>
<td>3.789</td>
<td>4.186</td>
<td>0.397*</td>
</tr>
<tr>
<td></td>
<td>(0.938)</td>
<td>(0.854)</td>
<td>(0.154)</td>
</tr>
<tr>
<td></td>
<td>[881]</td>
<td>[110]</td>
<td></td>
</tr>
<tr>
<td>Negative Peer Effects</td>
<td>2.738</td>
<td>2.252</td>
<td>-0.486**</td>
</tr>
<tr>
<td></td>
<td>(0.962)</td>
<td>(0.796)</td>
<td>(0.122)</td>
</tr>
<tr>
<td></td>
<td>[878]</td>
<td>[112]</td>
<td></td>
</tr>
<tr>
<td>Student Input</td>
<td>2.514</td>
<td>2.264</td>
<td>-0.250*</td>
</tr>
<tr>
<td></td>
<td>(0.924)</td>
<td>(0.851)</td>
<td>(0.093)</td>
</tr>
<tr>
<td></td>
<td>[882]</td>
<td>[111]</td>
<td></td>
</tr>
</tbody>
</table>

Note: *p<0.05, ** p<0.01; statistical significance is of difference in mean for over-subscribed charter and traditional public schools. In the first two columns, standard deviations are reported in parentheses and sample sizes in brackets. Standard errors reported in parentheses in the third column are adjusted for clustering by school.
### Table 8: Relations of non-cognitive traits and test-score gains, overall and within schools

<table>
<thead>
<tr>
<th></th>
<th>4th-8th MCAS Math Gains</th>
<th>4th-8th MCAS ELA Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Within Schools</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.100**</td>
<td>.144**</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Self-Control</td>
<td>0.076*</td>
<td>0.122**</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Grit</td>
<td>.110**</td>
<td>.155**</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Growth Mindset</td>
<td>0.198**</td>
<td>.155**</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.040)</td>
</tr>
</tbody>
</table>

Note: *p<0.05, ** p<0.01. Each cell presents results from a separate regression of math or ELA MCAS gains on the relevant non-cognitive trait. Within-school regressions include fixed effects for schools. Standard errors reported in parentheses are clustered by 8th-grade school. Non-cognitive traits are standardized to have mean zero and unit variance in the study sample.
<table>
<thead>
<tr>
<th>Offer at Charter School A</th>
<th>Administrative Data Sample</th>
<th>Non-cognitive Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.576**</td>
<td>0.952*</td>
</tr>
<tr>
<td></td>
<td>(0.370)</td>
<td>(0.441)</td>
</tr>
<tr>
<td>Offer at Charter School B</td>
<td>1.710**</td>
<td>2.039**</td>
</tr>
<tr>
<td></td>
<td>(0.515)</td>
<td>(0.461)</td>
</tr>
<tr>
<td>Offer at Charter School C</td>
<td>1.183</td>
<td>1.467</td>
</tr>
<tr>
<td></td>
<td>(0.874)</td>
<td>(1.561)</td>
</tr>
<tr>
<td>Offer at Charter School D</td>
<td>1.319*</td>
<td>1.682*</td>
</tr>
<tr>
<td></td>
<td>(0.651)</td>
<td>(0.740)</td>
</tr>
<tr>
<td>Offer at Charter School E</td>
<td>1.600**</td>
<td>1.248*</td>
</tr>
<tr>
<td></td>
<td>(0.374)</td>
<td>(0.524)</td>
</tr>
<tr>
<td>Joint F-test statistic</td>
<td>35.53</td>
<td>16.04</td>
</tr>
<tr>
<td>Observations</td>
<td>497</td>
<td>200</td>
</tr>
</tbody>
</table>

Note: *p<0.05, ** p<0.01. Standard errors reported in parentheses are clustered by 8th-grade school. The administrative data sample includes all applicants to over-subscribed charter schools matched to valid 2011 test scores in the Massachusetts Department of Elementary and Secondary Education state database. The study sample includes all students in the administrative data sample with valid data on at least one non-cognitive outcome. All regressions include controls for student gender, race, age, LEP, SPED, and free/reduced price lunch, cubic polynomials in 4th-grade MCAS ELA and math scores, and fixed effects for lottery applicant risk sets.
Table A2: Mean student non-cognitive traits in three middle schools over time, all sampled students

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Charter School 1</th>
<th>Charter School 2</th>
<th>District School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Y1</td>
<td>3.83</td>
<td>3.76</td>
<td>3.70</td>
</tr>
<tr>
<td>Spring Y1</td>
<td>3.58</td>
<td>3.64</td>
<td>3.58</td>
</tr>
<tr>
<td>Spring Y2</td>
<td>3.52</td>
<td>3.50</td>
<td>3.55</td>
</tr>
<tr>
<td>Spring Y3</td>
<td>3.43</td>
<td>3.35</td>
<td>3.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Fall Y1</th>
<th>Spring Y1</th>
<th>Spring Y2</th>
<th>Spring Y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Y1</td>
<td>3.46</td>
<td>3.20</td>
<td>3.06</td>
<td>2.92</td>
</tr>
<tr>
<td>Spring Y1</td>
<td>[100]</td>
<td>[95]</td>
<td>[86]</td>
<td>[71]</td>
</tr>
<tr>
<td>Spring Y2</td>
<td>[95]</td>
<td>[90]</td>
<td>[61]</td>
<td>[49]</td>
</tr>
<tr>
<td>Spring Y3</td>
<td>[95]</td>
<td>[90]</td>
<td>[61]</td>
<td>[49]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grit</th>
<th>Charter School 1</th>
<th>Charter School 2</th>
<th>District School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Y1</td>
<td>3.53</td>
<td>3.58</td>
<td>3.40</td>
</tr>
<tr>
<td>Spring Y1</td>
<td>3.32</td>
<td>3.44</td>
<td>3.29</td>
</tr>
<tr>
<td>Spring Y2</td>
<td>3.25</td>
<td>3.24</td>
<td>3.20</td>
</tr>
<tr>
<td>Spring Y3</td>
<td>3.20</td>
<td>3.21</td>
<td>3.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Fall Y1</th>
<th>Spring Y1</th>
<th>Spring Y2</th>
<th>Spring Y3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Y1</td>
<td>3.73</td>
<td>4.20</td>
<td>4.47</td>
<td>4.38</td>
</tr>
<tr>
<td>Spring Y1</td>
<td>[100]</td>
<td>[95]</td>
<td>[86]</td>
<td>[71]</td>
</tr>
<tr>
<td>Spring Y2</td>
<td>[95]</td>
<td>[90]</td>
<td>[61]</td>
<td>[49]</td>
</tr>
<tr>
<td>Spring Y3</td>
<td>[95]</td>
<td>[90]</td>
<td>[61]</td>
<td>[49]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Growth Mindset</th>
<th>Charter School 1</th>
<th>Charter School 2</th>
<th>District School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Y1</td>
<td>3.73</td>
<td>4.20</td>
<td>4.16</td>
</tr>
<tr>
<td>Spring Y1</td>
<td>[100]</td>
<td>[95]</td>
<td>[170]</td>
</tr>
<tr>
<td>Spring Y2</td>
<td>[95]</td>
<td>[90]</td>
<td>[138]</td>
</tr>
<tr>
<td>Spring Y3</td>
<td>[95]</td>
<td>[90]</td>
<td>[117]</td>
</tr>
</tbody>
</table>

Note: Brackets report sample size by outcome, year, and school. Conscientiousness, Self-Control, and Grit are measured on a 5-point scale; Growth Mindset is measured on a 6-point scale. Years 1-3 correspond to grades 5-7 in Charter School 1 and grades 6-8 in Charter School 2 and District School.
Dr. Deirdre J. Knapp is Vice President of HumRRO’s Research and Consulting Operations Division. She earned her doctorate in industrial-organizational psychology from Bowling Green State University in 1984 and has been at HumRRO since 1987. Dr. Knapp’s primary area of expertise is in performance measurement. She has developed performance measures for use in professional certification programs and as outcome measures in large-scale criterion-related validation research studies in employment settings. Dr. Knapp has experience with a wide range of measurement methods, including selected response, oral interviews, live and computer-based simulations, experience records, and observer ratings.

Dr. Knapp served on the National Research Council (NRC) committee that evaluated the impact of the National Board of Professional Teaching Standards and helped organize and presented at an NRC workshop on the assessment of 21st century skills in K-12 educational settings. Dr. Knapp has authored multiple book chapters, technical reports, and articles on job analysis, performance measurement, and various professional practice topics.
Dr. James W. Pellegrino is Liberal Arts and Sciences Distinguished Professor and Distinguished Professor of Education at the University of Illinois (UIC) at Chicago. He also serves as Co-director of UIC’s interdisciplinary Learning Sciences Research Institute.

His research and development interests focus on children's and adult's thinking and learning and the implications of cognitive research and theory for assessment and instructional practice. He has published over 275 books, chapters and articles in the areas of cognition, instruction and assessment. His current research on learning, instruction, and assessment is funded by the U.S. National Science Foundation and the U.S. Institute of Education Sciences. He has served as the head of several U.S. National Academy of Sciences study committees, including co-chair of the Committee on Learning Research and Educational Practice, and co-chair of the Committee on the Foundations of Assessment which issued the report Knowing What Students Know: The Science and Design of Educational Assessment.

Most recently he served as a member of the Committee on Science Learning: Games, Simulations and Education, as chair of the Committee on Defining Deeper Learning and 21st Century Skills, and co-chair of the Committee on Developing Assessments of Science Proficiency in K-12. He is a past member of the Board on Testing and Assessment of the National Research Council and a lifetime member of the U.S. National Academy of Education. He currently serves on the Technical Advisory Committees (TAC) of several states as well as the TACs of multiple state consortia funded under the U.S. Department of Education’s Race to the Top assessment initiative.

Business, political, and educational leaders are increasingly asking schools to integrate development of skills such as problem solving, critical thinking, and collaboration into the teaching and learning of academic subjects. These skills are often referred to as “21st century skills” or “deeper learning.”

At the request of several foundations, the National Research Council appointed a committee of experts in education, psychology, and economics to more clearly define “deeper learning” and “21st century skills,” consider these skills’ importance for positive outcomes in education, work, and other areas of life, address how to teach them, and examine related issues.

The committee’s findings and recommendations are detailed in its report Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century.

**First Steps Toward Clarifying Terms**

*Deeper learning* is the process through which a person becomes capable of taking what was learned in one situation and applying it to new situations – in other words, learning for “transfer.” Through deeper learning, students develop expertise in a particular discipline or subject area.

Suppose a student learns about means, medians and modes in mathematics. Deeper learning would mean that the student would learn not only how to calculate these values, but also understand how and when each is best used. For example, if the student later worked at a store that tracked average daily sales each month, he or she would recognize that a special sale on the first day of a particular month could skew the mean and that an alternative measure like the median might be more representative of daily sales for that month.

Through the process of deeper learning, students develop **21st century competencies** – transferable knowledge and skills. In contrast to a view of “21st century skills” as general
The available research is limited and primarily educational attainment – the total number of cognitive competencies, which have been more intertwined with – knowledge of a particular discipline or subject area. The committee uses the broader term “competencies” rather than “skills” to include both knowledge and skills.

Precise definitions of the many terms used for “21st century skills” are not possible at this time, in part because there is little research to support such definitions. However, as a preliminary way to organize the skills, the committee first identified three broad domains of competence:

- **the cognitive domain**, which includes thinking, reasoning, and related skills;
- **the intrapersonal domain**, which involves self-management, including the ability to regulate one’s behavior and emotions to reach goals; and
- **the interpersonal domain**, which involves expressing information to others, as well as interpreting others’ messages and responding appropriately.

The committee then took several existing lists of “21st century skills” and, based on a content analysis, grouped them within these three domains.

The figure above links similar competencies together, groupings that provide a starting point for further research on the competencies’ meaning and value.

**THE IMPORTANCE OF 21ST CENTURY COMPETENCIES**

The committee examined evidence on the importance of 21st century competencies within the three domains for positive outcomes in education, work, health, and other areas. They reached the following conclusions:

- The available research is limited and primarily correlational in nature. To date, only a few studies have demonstrated a causal relationship between one or more 21st century competencies and adult outcomes.
- Cognitive competencies, which have been more extensively studied than interpersonal and intrapersonal competencies, show consistent, positive correlations of modest size with desirable outcomes in education, the workplace, and health.
- Among intrapersonal and interpersonal competencies, conscientiousness – being organized, responsible, and hardworking – shows the strongest correlation with desirable work and educational outcomes. Anti-social behavior, which has both intrapersonal and interpersonal aspects, is negatively correlated with these outcomes.
- Educational attainment – the total number of years a person spends in school – strongly predicts adult earnings, as well as health and civic engagement. It may be that schooling builds some mix of cognitive, interpersonal, and intrapersonal skills that are valued by the labor market.
If so, making it possible for students to get more education may itself be a useful complementary way to develop 21st century competencies.

More research is needed to increase our understanding of the relationships between particular twenty-first century competencies and desired adult outcomes.

Over a century of research on transfer has yielded little evidence that teaching can develop general cognitive competencies that are transferable to any new discipline, problem or context, in or out of school. Much of the research has been carried out in the cognitive domain and it shows that transfer does occur but is limited in scope. Studies of interventions to teach social and emotional skills suggest that these also support transfer beyond the immediate context in which they were acquired, affecting students’ behavior throughout the school day. More research is needed to illuminate whether, and to what extent, competencies learned in one discipline or context of application can generalize and transfer to other disciplines or contexts.

The committee found not only that deeper learning develops 21st century competencies, but also that the relationship flows both ways: 21st century competencies can aid the process of deeper learning in a discipline or subject area. For example, deeper learning to develop expertise in a discipline or subject area requires months of sustained, deliberate practice – a process supported by the intrapersonal competency of conscientiousness.

TEACHING FOR DEEPER LEARNING

Emerging evidence indicates that cognitive, intrapersonal, and interpersonal competencies can be taught and learned in ways that support transfer. Research in the cognitive domain has also identified features of instruction that are likely to support transfer within a given subject area. For example, transfer is supported when instruction helps learners understand the general principles underlying the specific examples included in their original learning. Teaching that emphasizes not only content knowledge, but also how, when, and why to apply this knowledge is essential to transfer. Instruction should follow these research-based teaching methods:

- **Encourage elaboration, questioning, and explanation** – for example, by prompting students who are reading a history text to explain the material aloud to themselves or others as they read.
- **Engage learners in challenging tasks**, while also supporting them with guidance, feedback, and encouragement to reflect on their own learning processes.
- **Teach with examples and cases**, such as modeling step-by-step how students can carry out a procedure to solve a problem while explaining the reason for each step.
- **Prime student motivation** by connecting topics to students’ personal lives and interests, engaging students in problem solving, and drawing attention to the knowledge and skills students are developing and their relevance, rather than grades or scores.
- **Use “formative” assessments**, which continuously monitor students’ progress and provide feedback to teachers and students for use in adjusting their teaching and learning strategies.

DEEPER LEARNING IN STANDARDS DOCUMENTS

The committee found important areas where goals for deeper learning and 21st century competencies overlap with the new Common Core State Standards in English language arts and mathematics and the NRC Framework for K-12 Science Education. All three documents highlight the importance of helping students understand the general principles underlying specific content, a hallmark of deeper learning. A cluster of cognitive competencies—including critical thinking, nonroutine problem solving, and constructing and evaluating evidence-based arguments—is included in all three disciplines. Coverage of other competencies—especially those in the intrapersonal and interpersonal domains—is uneven. Developing the full range of 21st century competencies within the disciplines will require systematic instruction and sustained practice, a change from current practice that will require additional instructional time and resources.

MOVING FORWARD

Because 21st century competencies support deeper learning of school subjects, their widespread acquisition could potentially reduce disparities in educational attainment, preparing a broader swathe of
young people for successful adult outcomes in work and other areas of life.

However, important challenges remain in two major areas. First, research and development is needed to create and evaluate new curricula based on the research-based instructional methods described above, and to find valid ways to assess cognitive, intrapersonal, and interpersonal skills. Second, at the level of education systems and policies, new approaches to teacher preparation and professional development will be needed to help instructors acquire a deep understanding of the role of 21st century competencies in learning core academic content and create environments that support students’ learning of these competencies.

To help address these systemic issues, the states and the federal government should establish policies and programs—in the areas of assessment, accountability, curriculum and materials, and teacher education—to support students’ acquisition of transferable knowledge and skills.

RESEARCH NEEDS

Far more research is needed to fill gaps in the evidence base on deeper learning and 21st century competencies. Foundations and federal agencies should support research aimed at:

• Establishing agreed-upon definitions of 21st century competencies and ways to measure and assess them.

• Better illuminating the relationships—particularly any causal relationships—between 21st century competencies and desired outcomes.

• Gaining a better understanding of whether, and to what extent, teaching for transfer within an academic discipline (such as mathematics) can facilitate the transfer of competencies across disciplines (for example, from mathematics to history).

COMMITTEE ON DEFINING DEEPER LEARNING AND 21ST CENTURY SKILLS

JAMES W. PELLEGRINO (Chair), Learning Sciences Research Institute, University of Illinois–Chicago; GREG J. DUNCAN, University of California, Irvine; JOAN L. HERMAN, National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles; MARGARET A. HONEY, New York Hall of Science, Queens; PATRICK C. KYLLONEN, Center for New Constructs, Educational Testing Service; HENRY M. LEVIN, Teachers College, Columbia University; CHRISTINE MASSEY, Institute for Research in Cognitive Science, University of Pennsylvania; RICHARD E. MAYER, University of California, Santa Barbara; KENT MCGUIRE, Southern Education Foundation, Atlanta, Georgia; P. DAVID PEARSON, Graduate School of Education, University of California, Berkeley; EDWARD A. SILVER, University of Michigan; MARGARET L. HILTON, Study Director

FOR MORE INFORMATION . . . This brief was prepared by the Board on Testing and Assessment based on the report Education for Life and Work: Developing Transferrable Knowledge and Skills in the 21st Century. The study was sponsored by the Carnegie Corporation of New York, the William & Flora Hewlett Foundation, the John D. & Catherine T. MacArthur Foundation, the National Science Foundation, the Nellie Mae Education Foundation, the Pearson Foundation, SCE, the Stupski Foundation and the Raikes Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not reflect those of the sponsoring organizations. Copies of the report are available from the National Academies Press, 500 Fifth Street, N.W., Washington, DC 20001; (800) 624-6242; http://www.nap.edu.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The nation turns to the National Academies—National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council—for independent, objective advice on issues that affect people’s lives worldwide.

www.national-academies.org
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 ................................................... 10
   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? ? ................................................................................ 12
   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? .............................. 13
   I. What restrictions apply to my interaction with foreign entities? ................................................................. 13
   J. What do I do if I am called to be an expert witness? .................................................................................... 14
   K. May I keep and use frequent flyer miles that I earn when I am on official NAGB travel? ......................... 14

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 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.

Keep in mind that when you are disqualified from a matter such as a contract competition, the particular matter that you must recuse yourself from is the entire competition for this contract. You are prohibited from doing anything at all with respect to this competition. This means, for example, that you may not review other proposals that are in competition with that of the organization in which you have a direct or imputed financial interest.

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’s 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
National Assessment Governing Board
Nominations Committee

December 7, 2013
7:30 – 8:15 am

AGENDA

Closed Session 7:30 – 8:15 am

<table>
<thead>
<tr>
<th>Time</th>
<th>Event Description</th>
</tr>
</thead>
</table>
| 7:30 – 7:35 am| Welcome, Introductions, and Agenda Overview
               | Tonya Miles, Chair                                     |
| 7:35 – 8:15 am| Discussion of 2014 Nominees
               | Committee Members                                     |